TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 4 (2012) 232–248

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2012.01183.x

The Interplay Between Gesture and Speech in the Production of Referring Expressions: Investigating the Tradeoff Hypothesis

Jan P. de Ruiter,[a,b] Adrian Bangerter,[c] Paula Dings[b]

[a]*Faculty for Linguistics and Literary Studies, University of Bielefeld*
[b]*Max Planck Institute for Psycholinguistics*
[c]*Institute of Work and Organizational Psychology, University of Neuchâtel*

Received 28 February 2010; received in revised form 30 August 2010; accepted 5 November 2010

Abstract

The tradeoff hypothesis in the speech–gesture relationship claims that (a) when gesturing gets harder, speakers will rely relatively more on speech, and (b) when speaking gets harder, speakers will rely relatively more on gestures. We tested the second part of this hypothesis in an experimental collaborative referring paradigm where pairs of participants (directors and matchers) identified targets to each other from an array visible to both of them. We manipulated two factors known to affect the difficulty of speaking to assess their effects on the gesture rate per 100 words. The first factor, codability, is the ease with which targets can be described. The second factor, repetition, is whether the targets are old or new (having been already described once or twice). We also manipulated a third factor, mutual visibility, because it is known to affect the rate and type of gesture produced. None of the manipulations systematically affected the gesture rate. Our data are thus mostly inconsistent with the tradeoff hypothesis. However, the gesture rate was sensitive to concurrent features of referring expressions, suggesting that gesture parallels aspects of speech. We argue that the redundancy between speech and gesture is communicatively motivated.

Keywords: Gesture; Pointing; Iconic gestures; Referring expressions; Speech production; Gesture–speech tradeoff; Gesture–speech redundancy

1. Introduction

Speaking is an activity often accompanied by meaningful movements of the hands, called gesture or gesticulation. Kendon (2004, p. 7) called these hand movements ‘‘visible actions as utterances.’’ It has long been established that gestures are intrinsically related to the

Correspondence should be sent to Jan P. de Ruiter, Faculty for Linguistics and Literary Studies, University of Bielefeld, P.O. Box 10 01 31, 33501 Bielefeld, Germany. E-mail: jan.deruiter@uni-bielefeld.de

process of (spontaneous) speaking (Kendon, 1972; McNeill, 1985, 1992). An intriguing question that has generated much research is: What exactly is the relationship between gesture and speech? Many answers to this question have been suggested.

One possibility that has been especially influential (see, e.g., Bangerter, 2004; De Ruiter, 2006; Melinger & Levelt, 2004; Van der Sluis & Krahmer, 2004, 2007) is that there is a *tradeoff* relation between gesture and speech in terms of their communicative load. That is, the tradeoff hypothesis assumes that if speech becomes more difficult (i.e., requires more effort to verbally encode intended meaning), the likelihood of a gesture occurring, to ''take over'' some of the communicative load, is higher. Alternatively, when gesturing becomes harder, the tradeoff hypothesis predicts that speakers will rely relatively more on speech. There is empirical evidence for the latter conjecture. For example, Bangerter (2004) found that pointing gestures decreased with increasing distance to the target. Melinger and Levelt (2004) found that speech produced with concurrent gestures was less explicit than speech without gestures. And Van der Sluis and Krahmer (2007) proposed a computational model for generating multimodal references adapted from Fitt's (1954) law that specifies the costs of pointing (and thus the relative reliance on words and gestures) depending on distance to the target and its relative size. They conducted two production experiments where they manipulated the costs of pointing to test their model.

Although the tradeoff hypothesis is plausible, an alternative account also exists. So, Kita, and Goldin-Meadow (2009) found that speakers' gestures paralleled rather than compensated for underspecifications in speech when describing scenes to an experimenter. They concluded that gesture is redundant with, or goes hand in hand with, speech. We will call this alternative hypothesis the *hand-in-hand* hypothesis. So et al. also suggested (pp. 116–117) that people may gesture ''for their own cognitive benefit,'' for instance, by facilitating speech planning processes (see also Kita, 2000; Krauss, 1998; Krauss, Apple, Morency, Wenzel, & Winton, 1981; Krauss, Chen, & Chawla, 1995; Krauss, Chen, & Gottesmann, 2000; Krauss, Morrel-Samuels, & Colasante, 1991). We will treat this issue as separate from the hand-in-hand hypothesis itself and address it in Section 5.

Resolving the tradeoff versus hand-in-hand issue may ultimately lead to a better understanding of the relationship between gesture and speech, especially the complex question of the communicative function of gesture (Kendon, 1994). It may also help decide between alternative theoretical models of gesture production (De Ruiter, 2000, 2007; Hostetter & Alibali, 2008) or constrain computational models (Van der Sluis & Krahmer, 2007).

The tradeoff versus hand-in-hand issue may depend on what types of gesture are being considered, as well as the communicative setting (Bavelas, 1994; Bavelas, Gerwing, Sutton, & Prevost, 2008). In this study, we explore the relationship between gesture and speech in situations where people collaboratively refer to something in the shared visual environment. We focus on two types of gesture that are often used in referring and closely synchronized with affiliated speech: (a) pointing (or *deictic*) gestures and (b) what McNeill (1992) termed *iconic* gestures. We focus on referential communication tasks because they constitute stringent tests of communicative intent (Melinger & Levelt, 2004). In other situations, for example, reciting a story or a movie fragment, factors irrelevant to the tradeoff and hand-in-hand hypotheses (difficulties in recalling content, the

conversational imperative of maintaining speech rate) may compete for speakers' resources, leading to suboptimal message design.

We focus on one part of the tradeoff hypothesis: When verbal referring is harder, there will be more gestures. This hypothesis is a central component of computational models like Van der Sluis and Krahmer's (2007). To date, there are few studies that directly test this hypothesis. De Ruiter (1998) found no differences in gesture rate in descriptions of stimuli that were either hard or easy to describe. Morsella and Krauss (2004) performed a similar experiment but found that less ''describable'' pictures resulted in a higher gesture rate. However, they did not report gesture-per-word rates, but rather the proportion of time that participants gestured during a description. This measure does not enable a direct comparison between the amount of gesture and the amount of speech.

So in order to test this hypothesis more thoroughly, we manipulated the difficulty of verbal referring, to explore the effect this has on the relationship between speaking on the one hand and iconic and pointing gestures on the other. We adapted Bangerter's (2004) variation of the classical matching task procedure (Clark & Wilkes-Gibbs, 1986) that allows collaborative referential communication, but with the possibility of using gestures. A collaborative setting where participants can freely engage in dialog is important for testing hypotheses about the communicative function of gestures (Bavelas et al., 2008). In our study, directors identified targets to matchers from an array of targets (tangram figures) visible to both of them. We manipulated two factors known to affect the difficulty of speaking: codability (within-subjects) and common ground (within-subjects). We also manipulated a third factor, mutual visibility (between-subjects), that is known to affect the gesture rate and the type of gesture produced (Bavelas et al., 2008). We analyzed their isolated and cumulative effects on the tradeoff between gesture and speech in referring. We review research on these factors before describing our procedures.

2. Factors affecting the difficulty of verbal referring and gesture production

2.1. Target set features: Codability

Features of the target set include properties of the referents themselves, like how complex they are or how easily they can be distinguished from other potential referents. These two aspects, *codability* and *discriminability*, affect directors' strategies, as well as the amount of verbal effort required for reference completion. Easily codable targets are described relatively more often with holistic expressions than piecemeal (Hupet, Seron, & Chantraine, 1991). The cognitive processes involved in producing referring expressions for easily codable figures resemble those involved in picture-naming tasks: object identification, lemma retrieval, and pronunciation (e.g., Roelofs, 1992). But for less codable targets, the director has to construct descriptive phrases, which increases cognitive load in informational, grammatical, and intonational planning (Levelt, 1989). Moreover, for less codable words, the director has to produce referring expressions that enable the matcher to single out the target. This often involves the cognitively demanding process of *recipient design* (Schegloff, 1972)

or *audience design* (Clark & Murphy, 1982; Keysar, Barr, & Balin, 1998; Keysar, Barr, Balin, & Brauner, 2000).

We created a codability factor with three decreasing levels of codability. The first level (*simple* tangrams) were figures with monomorphemic names like *star* or *circle*. The second level (*humanoid* tangrams) are a subset of the tangrams from Clark and Wilkes-Gibbs (1986) that are abstract but can be described as humanoid figures, like *the ice dancer*. The third level (*abstract* tangrams) consists of figures with complex shapes that do not resemble anything with a simple name. We expected simple tangrams to be easy to name, humanoids harder, and the abstract tangrams the hardest. By the tradeoff hypothesis, decreasing codability of targets should lead to an increase in the gesture rate.

2.2. Common ground

Common ground between partners makes verbal referring easier (Clark & Wilkes-Gibbs, 1986) because partners develop *conceptual pacts* (Brennan & Clark, 1996) when they repeatedly refer to the same object. Common ground also affects gestural referring. For example, pointing gestures may be used to focus addressees' gaze on referents that are outside of the joint focus of attention (Bangerter, 2004). Also, when partners converse about mutually known referents, gestures are reduced in complexity and precision, and are less informative (Gerwing & Bavelas, 2004). Similarly, in a study of how participants describe the size of referents in visual scenes, Holler and Stevens (2007) found that, when the referent was new, size was represented either in gesture only or both verbally and gesturally. But when the referent was known to both participants, size information was mainly represented verbally. Also, Jacobs and Garnham (2007) found that repeated narration of the same story to the same listener led to a decrease in the rate of gesturing (however, Holler & Wilkin, 2009, found that common ground led to an increased gesture rate). In light of these conflicting findings, we manipulated *repetition* of targets to study what happens to the gesture rate when conceptual pacts can be used. The tradeoff hypothesis predicts that repeated referring should lead to the elaboration of conceptual pacts about how to refer to targets, thus facilitating generation of verbal expressions and ultimately decreasing the gesture rate.

2.3. Mutual visibility

Mutual visibility does not seem to directly affect the ease of verbal referring. Bavelas et al. (2008) found no differences in speech production as a function of visibility. However, several studies have investigated the impact of mutual visibility on the gesture rate (for a summary, see Bavelas et al., 2008). A consistent finding is that the gesture rate decreases when partners are not mutually visible, without being completely reduced to zero, although this result varies according to the type of gesture, with the rate of beat gestures being independent of mutual visibility (Alibali, Heath, & Myers, 2001). Thus, mutual visibility seems to affect the relationship between gesture and speech as well as the possible function of gestures produced. Gestures produced when partners are mutually visible may be more communicative in nature, whereas those produced when partners are not may serve cognitive needs

of the speaker (Kita, 2000; Melinger & Kita, 2007). Thus, the hypothesized tradeoff relationship in communicative load between gesture and speech may only hold when gestures are produced with communicative intent, that is, when there is mutual visibility (Bavelas et al., 2008). We therefore manipulated *mutual visibility*, that is, whether the director and matcher could see each other.

3. Method

3.1. Participants and procedure

Ninety-six participants, all native speakers of Dutch and students at Radboud University Nijmegen, worked in 48 pairs on a collaborative matching task. Pairs consisted of one director and one matcher, seated side by side at a table facing a poster on a wall with 24 different tangram figures printed on it in a cloud-like shape. The poster was fully visible to both of them. The distance between the participants and the poster on the wall was approximately arm's length plus 25 cm. Director and matcher each had a touch screen beside them. Both could see their own screen, but not their partner's. Pairs were randomly assigned to a visibility condition and to one of two presentation order lists. In the visible condition, partners could see each other. In the nonvisible condition, a screen was placed between them, obscuring their view of their partner and their partner's gestures but not of the poster.

For every trial, one figure was circled on the director's screen. Directors described the circled figure to the matchers. Matchers identified the figure and marked it on their screen. Then, both partners pressed a button on their screen to move to the next trial. Both participants were allowed to speak freely and were told that they could use gestures if they wanted to, but were not explicitly encouraged to do so. The setup is shown in Fig. 1.

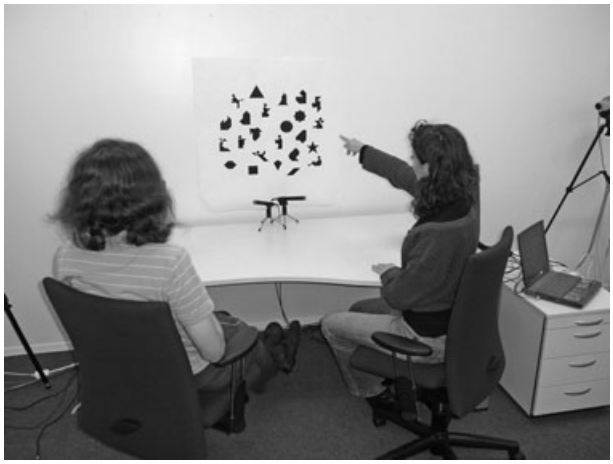

Fig. 1. Setup of experiment (mutually visible condition).

The director's screen depicted an exact copy of the array shown on the poster. The matcher's touch screen displayed the same figures, but arranged in a grid of six columns and four rows. This grid was kept constant throughout the experiment. Each codability level comprised eight figures.

Pairs identified all 24 figures on the poster once. We manipulated repetition as in the matching task paradigm (Clark & Wilkes-Gibbs, 1986; Krauss & Weinheimer, 1966) by presenting the three last-presented figures from each presentation order list for each of the three codability categories a second and third time. Thus, there were 42 trials in all (24 first presentations followed by 18 repetitions). We only repeated the three last-presented figures to minimize the time lag between the original presentation and the repetitions, and thus reduce the risk that conceptual pacts might decay. We used two different presentation orders for the first 24 trials (randomized with respect to codability); the second order was the inverse of the first one. The presentation order for the 18 repetition trials was also randomized.

3.2. Data acquisition and preparation

Interactions were recorded on video and audio. In the visible condition, synchronized recordings from three cameras were made. One camera was positioned in front of the director, one in front of the matcher, and a third was placed behind them. The front-view cameras recorded facial expressions and gestures. In the nonvisible condition, synchronized recordings from four cameras were made. Both participants had one side-view camera for recording facial expression, and one ceiling-mounted camera above them for recording gestures. Audio recordings were made with two high-quality direction-sensitive table microphones, one per participant, placed at the end of the table and pointed toward the participant. Recordings were digitized into MPEG format. Speech and gestures were transcribed and analyzed with the multimodal analysis program ELAN (Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, http://www.lat-mpi.eu/tools/elan/; see also Brugman & Russel, 2004).

3.3. Speech coding

For every trial, the first uninterrupted referring expression was identified. This usually corresponded to the first referring expression by directors, who would stop talking after completing the expression and wait for a response from matchers. In 80% of all trials, matchers responded immediately to this initial utterance by selecting a picture (which was correct in 84% of cases). This shows that in the majority of trials, the first uninterrupted utterance was deemed complete by the matchers and was sufficiently detailed to enable the matcher to select the correct picture. In the remaining cases, matchers either interrupted the initial utterance because something was unclear or because they already thought they knew which picture the director was describing. In those cases, we analyzed the initial part of the referring expression up to the interruption. Although referring is a collaborative process (Clark & Wilkes-Gibbs, 1986), we focused on the initial expressions of directors for a more stringent test of the tradeoff hypothesis. The mutually interactive *grounding* process (Clark,

1996) obviously plays a central role in collaborative referring, but it makes an accurate assessment of the tradeoff between gesture and speech information in the later utterances by the director highly dependent on the content of the preceding utterances by the matcher. The very real possibility of directors replying to (possibly implicit) clarification questions posed by matchers (Clark & Krych, 2004) might well obscure a potential tradeoff effect.

Referring expressions were coded for several variables. First, we coded the number of feature descriptions, or specifications of the target or some part thereof. For example, *the big pointy triangle* constitutes three feature descriptions. Second, we coded the number of locative descriptions, that is, specifications of the absolute location of a target (e.g., *the upper left corner*) or its location relative to a salient landmark (e.g., *below the big triangle*). Third, we coded the use of conceptual pacts (e.g., *the ice dancer*).

Interrater agreement was assessed by having a second coder doublecode 42 referring expressions from one pair for each of these variables, and computing Cohen's kappa. All kappas were between .87 and 1.0 (all ps < .0001).

Finally, we used Praat (Boersma & Weenink, 2007) to measure the lag between the presentation of each target on the director's screen and the onset of the director's referring expression (hereafter: *speech initiation time*). This is a measure of cognitive load in speech production (Levelt, 1989).

3.4. Gesture coding

We coded two types of gestures: pointing and iconic gestures. Pointing gestures were characterized by partial or full extension of the pointing arm with the elbow lifted from the table. Iconic gestures illustrated a particular feature of the target (e.g., shape). We further distinguished between obligatory and nonobligatory iconic gestures. Obligatory iconic gestures contain disambiguating information that is not represented in speech but is nevertheless essential for understanding it, for example, when a director says *the one with a shape like this* and traces a curve in the air.[1] In contrast, with nonobligatory iconic gestures, the affiliated speech can still be understood without access to the gestural information, as when a director says *the big triangle* while tracing a triangle in the air.

Interrater agreement was assessed by having a second coder doublecode 42 referring expressions from one pair for each gesture type, and computing Cohen's kappa. All kappas were between .90 and 1.0 (all ps < .0001).

3.5. Manipulation check

We checked the extent to which our manipulations were effective. According to the literature reviewed above, we expected codability and repetition to decrease verbal effort (measured by the number of words in the initial description) and cognitive load (measured by speech initiation time).

We analyzed the data using linear mixed-model analysis with visibility, codability, and repetition as fixed effects and items and pairs as random effects to predict verbal effort and speech onset latency. The levels of the codability variable were entered as dummy variables.

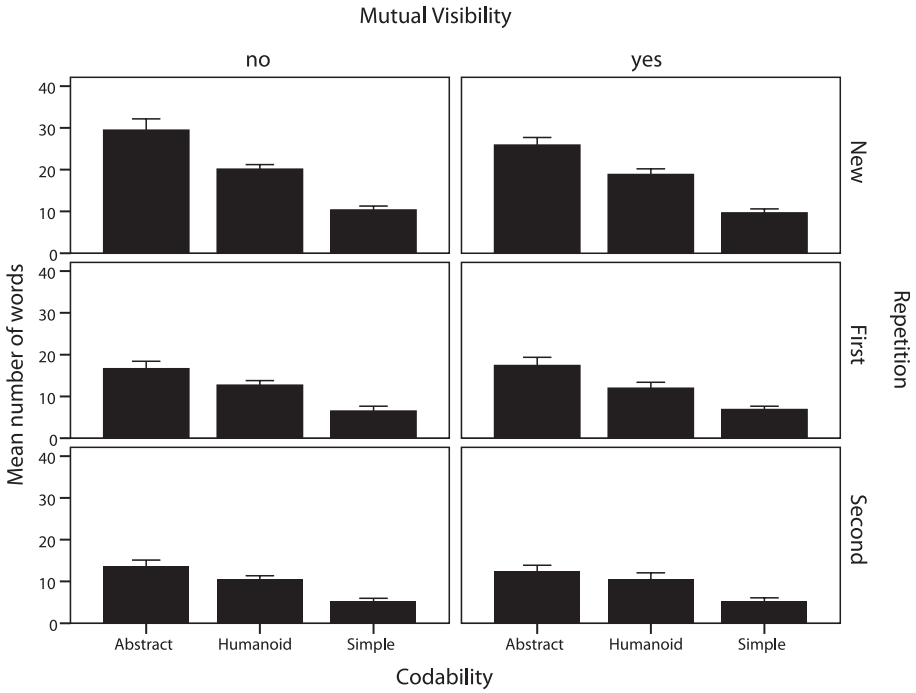

Fig. 2. Mean number of words in initial referring expressions as a function of codability, repetition, and visibility. *Note.* Error bars indicate 1.5 *SE*s.

The use of linear mixed-model analyses allows modeling items and pairs as random variables in the same analysis and thus eliminates the need for conducting separate analyses by subjects (F_1) and by items (F_2) or *minF′* analyses (Baayen, Davidson, & Bates, 2008; Locker, Hoffman, & Bovaird, 2007). All the *p*-values related to the significance of the *b* coefficients were estimated using the Markov Chain Monte Carlo method in the statistical package R (see Baayen et al., 2008, and references therein for details).

The manipulations affected verbal effort. First, repetition decreased the number of words used ($b_{\text{Repetition}} = -8.7$, $p = .0001$). Second, codability also decreased the number of words used; humanoid and simple tangrams required fewer words to identify than abstract tangrams ($b_{\text{Humanoid}} = -12.4$, $p < .0001$; $b_{\text{Simple}} = -24.5$, $p < .0001$). Codability and repetition also interacted, suggesting that effects of codability were stronger for initial references than for repeated references ($b_{\text{Simple} \times \text{Repetition}} = 5.9$, $p = .0001$; $b_{\text{Humanoid} \times \text{Repetition}} = 3.3$, $p = .0012$). Unexpectedly, mutually visible pairs also used fewer words to identify targets than hidden pairs ($b_{\text{Visibility}} = -4.7$, $p = .044$). Means and standard errors are shown in Fig. 2.

The manipulations also affected speech initiation time, even though we took the number of words into account by including it as a covariate in the statistical analysis. Repetition decreased speech initiation time ($b_{\text{Repetition}} = -0.35$, $p = .006$). Codability also decreased speech initiation time: Descriptions of simple tangrams were initiated faster than those of abstract tangrams ($b_{\text{Simple}} = -0.67$, $p = .04$). Means and standard errors are shown in Fig. 3.

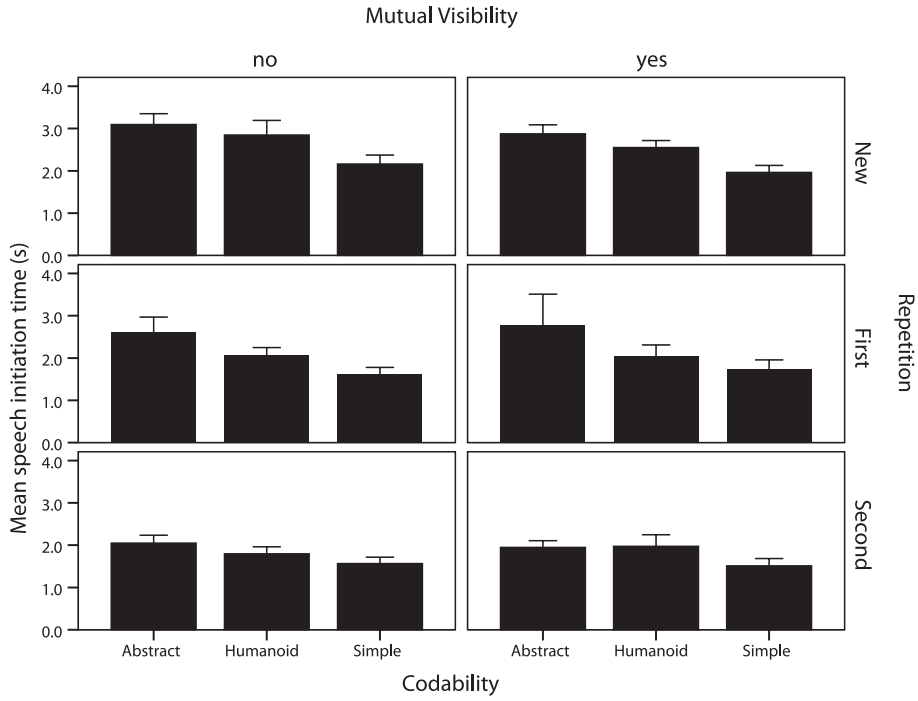

Fig. 3. Mean speech initiation time in initial referring expressions as a function of codability, repetition, and visibility. *Note*. Error bars indicate 1.5 *SE*s.

Note that our using speech initiation time as an estimate of cognitive load does not mean that this cognitive load only affects the initial planning phase of a referring expression. It is known from earlier research in psycholinguistics that speech production is incremental (Levelt, 1989), and that speakers do not plan and memorize entire utterances ahead of time before initiating articulation (Kempen & Huijbers, 1983; Schriefers, De Ruiter, & Steigerwald, 1999).

4. Results

To operationalize the notion of a tradeoff between language and gesture, we computed the frequency of gestures per 100 words (hereafter: *gesture rate*) for pointing gestures, obligatory iconic gestures, and nonobligatory iconic gestures. The higher the gesture rate, the more a given description relies on gestures relative to words. The part of the tradeoff hypothesis we focused on predicts that when the difficulty of describing increases, directors will rely relatively more on gestures. Thus, we expected our manipulations of codability and repetition to increase the gesture rate. Means and standard errors of the three dependent variables (number of pointing gestures, number of nonobligatory iconics, and number of obligatory iconics) are depicted in Fig. 4.

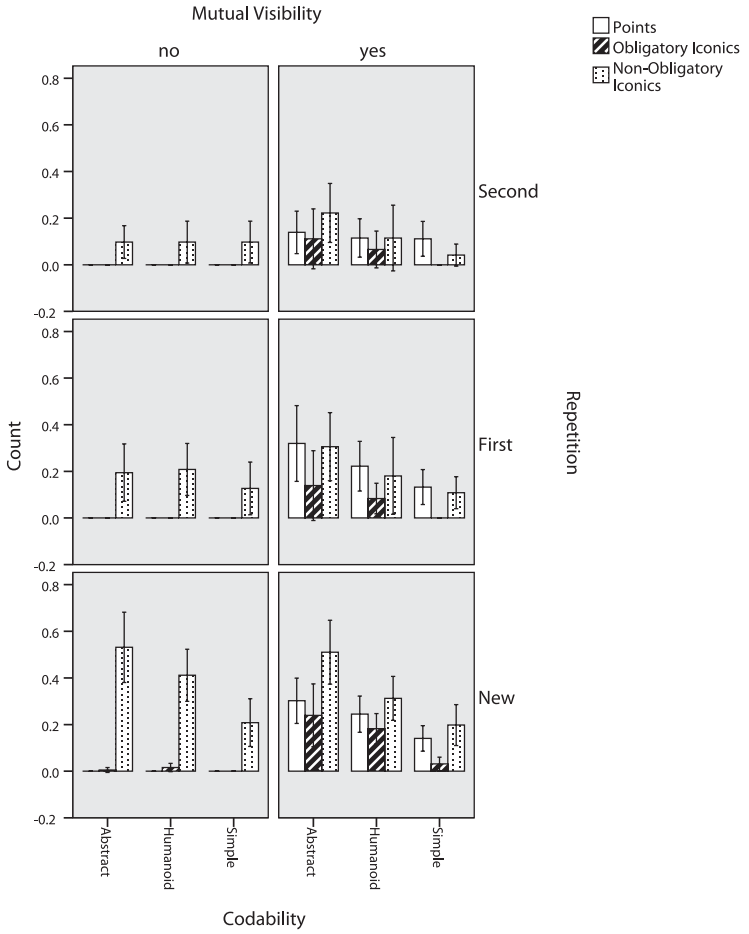

Fig. 4. Mean rates per 100 words of pointing gestures, obligatory iconic gestures, and nonobligatory iconic gestures as a function of codability, repetition, and visibility. *Note*. Error bars indicate 1.5 *SE*s.

We analyzed the data using linear mixed-model analysis to predict the dependent variables with visibility, picture repetition, codability, and their interactions as fixed effects and items and pairs as random effects. The levels of the codability variable were entered as separate dummy variables. We also entered the number of feature descriptions, number of locative descriptions, and the number of references to conceptual pacts into the model as covariates. This allowed us to test whether and how concurrent verbal features of referring expressions are related to the gestural dependent variables independently of the manipulations.

4.1. Gesture rate: Pointing

Directors produced no pointing gestures at all when they were not mutually visible; this naturally corresponds to a strong main effect of visibility ($b_{\text{Visibility}} = 1.8$, $p = .0066$). This

suggests that pointing gestures are designed for communicative purposes, because their use by directors is sensitive to shared visual context. Otherwise, there were no effects of the manipulated variables of codability and repetition on gesture rate ($b_{\text{Repetition}} = 0.11$, $p = .5396$; $b_{\text{Humanoid}} = 0.07$, $p = .8624$; $b_{\text{Simple}} = 0.07$, $p = .8728$). Thus, manipulating the difficulty of describing the targets did not affect the directors' relative reliance on gestures. This is evidence against the tradeoff hypothesis.

There were, however, effects of the covariates: The pointing gesture rate increased with the presence of locative descriptions ($b_{\text{Locative}} = 0.23$, $p = .0066$) in the referring expression and decreased with the use of conceptual pacts ($b_{\text{Conceptual pacts}} = -0.29$, $p = .0084$). Thus, directors' relative reliance on pointing gestures increased when they produced locative descriptions. This is evidence consistent with the hand-in-hand hypothesis: Pointing parallels a specification of the target location produced in speech. The fact that the use of conceptual pacts (i.e., referring to targets in common ground) decreases the relative reliance on pointing is consistent with the tradeoff hypothesis.

Taken together, then, there was little support for the tradeoff hypothesis for pointing gestures, and some support for the hand-in-hand hypothesis.

4.2. Gesture rate: Obligatory iconics

Directors produced almost no obligatory iconic gestures when they were not mutually visible; this naturally corresponds to a strong main effect of visibility ($b_{\text{Visibility}} = 0.67$, $p = .0092$). This suggests that, like pointing gestures, obligatory iconics are designed for communicative purposes, because their use by directors is sensitive to the shared visual context.

Oddly, even though the obligatory iconic gestures are (by definition) not accompanied by redundant speech, their rate nevertheless increased with the presence of feature descriptions in the referring expression ($b_{\text{Feature}} = 0.08$, $p = .0008$). This positive relationship between feature descriptions and iconic gestures also holds, more strongly, in the nonobligatory gestures (see below).

Importantly, there was no effect of repetition on gesture rate ($b_{\text{Repetition}} = 0.06$, $p = .5274$), nor of codability ($b_{\text{Humanoid}} = 0.03$, $p = .9198$; $b_{\text{Simple}} = 0.02$, $p = .9514$). Thus, manipulating the difficulty of describing the targets did not affect the directors' relative reliance on gestures. This is evidence against the tradeoff hypothesis.

Taken together, then, there was no support for the tradeoff hypothesis for obligatory iconic gestures.

4.3. Gesture rate: Nonobligatory iconics

The rate for nonobligatory iconic gestures was not affected by visibility ($b_{\text{Visibility}} = -0.50$, $p = .4488$). This suggests that nonobligatory iconic gestures are not produced for communicative purposes.

Again, there was no effect of repetition on gesture rate ($b_{\text{Repetition}} = -0.26$, $p = .2638$), nor of codability ($b_{\text{Humanoid}} = 0.19$, $p = .7388$; $b_{\text{Simple}} = -0.03$, $p = .9668$).

Thus, manipulating the difficulty of describing the targets did not affect the directors' relative reliance on iconic gestures. This is evidence against the tradeoff hypothesis. The only factor that correlated with the gesture rate for nonobligatory iconics was the number of feature descriptions ($b_{\text{Feature}} = 0.36$, $p = .0001$) in the referring expression. Thus, directors' relative reliance on nonobligatory iconic gestures increased when they described target features. This is evidence consistent with the hand-in-hand hypothesis: Gesturing is consistent with a specification of the target produced in speech. Thus, there was no support for the tradeoff hypothesis for nonobligatory iconic gestures, and some support for the hand-in-hand hypothesis.

5. Discussion

This study investigated the tradeoff hypothesis, which entails that if speaking gets harder, gesture will take over the communicative load, and vice versa. The alternative hypothesis, which we referred to as the hand-in-hand hypothesis (So et al., 2009), represents the opposite assumption: More speech goes with more gesture, less speech with less gesture. In order to operationalize the notion of ''difficulty in speaking'' as comprehensively as possible, we systematically varied three central aspects of the referring context: codability of the stimulus, mutual visibility, and repetition of reference.

Codability and repetition affected both the number of words and the speech initiation times (corrected for number of words) of referring expressions, which is consistent with previous research (e.g., Morsella & Krauss, 2004; Jacobs & Garnham 2007). Although we knew from previous work (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986) that referring expressions tend to get shorter when they are repeated with the same interlocutor, our controlled experimental approach enabled us to establish that conceptual pacts also reduce the *cognitive load* of formulating appropriately designed referring expressions.

Mutual visibility had a clear effect: Mutually hidden directors and matchers did not point at all, and hardly produced any obligatory iconics. However, the rate of nonobligatory iconics was unaffected. Pointing gestures and other gestures that are generally necessary to understand the whole utterance (such as obligatory iconics) are communicatively motivated and designed for a recipient. This is why directors refrain from using them if there is no mutual visibility. Nonobligatory iconics, on the other hand, are not necessary for interpreting speech, as the frequent use of these gestures in telephone conversations suggests. This finding extends results of Alibali et al. (2001) that the gesture rate is unaffected by mutual visibility for so-called *beat* gestures but decreases for so-called *representational* gestures. However, we have shown that not all representational gestures are equal: Some are meant to be seen, and others not necessarily so. This supports the finding by Bavelas et al. (2008) that gestures that have a *demonstrative* function are produced predominantly when there is mutual visibility. Combining our findings with those of Alibali et al. (2001) and Bavelas et al. (2008), a more complete picture emerges: Beats and nonobligatory iconics are not influenced by mutual visibility, whereas pointing gestures and obligatory iconic gestures are.

Although it is tempting to jump to the conclusion that nonobligatory iconic gestures are therefore produced for speaker-internal reasons (cf., Kita, 2000; Krauss, 1998; Krauss et al., 1981, 1991, 1995, 2000; Melinger & Kita, 2007; So et al., 2009), we believe that our results provide evidence against that conclusion. If producing nonobligatory iconic gestures reduces cognitive load, we would expect higher gesture rates in conditions with higher cognitive load. The cognitive load as measured by the speech initiation time was both *high* (average speech initiation time was 2,321 ms, much higher than for standard naming tasks, e.g., Jescheniak & Levelt, 1994) and *sensitive* to the different levels of codability (1,865, 2,392, and 2,713 ms for simple, humanoid, and abstract targets, respectively). Nevertheless, iconic gesture rates were not affected at all by codability. If gesture indeed facilitated speech planning, it is hard to explain why codability affected the length and planning times of referring expressions but had no effect at all on the gesture rate. Note, however, that we only manipulated cognitive load due to speaking (formulation) processes, and not cognitive load due to memory processing. There is evidence that the rate of iconic gestures is affected by memory load (De Ruiter, 1998; Morsella & Krauss, 2004; Wesp, Hesse, Keutmann, & Wheaton, 2001). However, given that memory load was both low and the same for the different conditions in our experiment, it is unlikely that memory load had any differential influence on our results.

This brings us to the main point of this study, namely the evaluation of the tradeoff and hand-in-hand hypotheses. Only one result supported the tradeoff hypothesis: The rate of pointing decreased when directors repeated a referring expression. This result underlines the central role of conceptual pacts in facilitating conversational referring, although it remains unclear how exactly this happens (e.g., by facilitating lexical access or audience design). However, the iconic gesture rate did not change, and all the other manipulations that made speaking more difficult had a strong effect on speech (as they were designed to) but no effect on the rate of any of the three gesture types. This is inconsistent with the tradeoff hypothesis. In addition, we found evidence supporting the hand-in-hand hypothesis: The rate of pointing gestures was positively related to the amount of locative descriptions in speech, and the rate of iconic gestures with the amount of feature descriptions in speech. It appears that when people gesture during initial referring, gesture and speech tend to express similar types of information.

What are the implications of these findings for the generation of referring expressions in natural language generation, for instance, in artificial agents? We suggest that a computational model in which gesture and speech go hand in hand is more natural and possibly also more effective than one based on a tradeoff of the communicative load over the two modalities, which has been suggested several times (e.g., Bangerter, 2004; De Ruiter, 2006; Van der Sluis & Krahmer, 2007). It is also easier to implement: Locative expressions and semantic features in natural language generation could be extended by gestural counterparts, for instance, by employing multimodal representation systems such as MURML (Kranstedt, Kopp, & Wachsmuth, 2002). A problem, however, is that people do not always produce gestures for every feature or locative they express in speech, and there are also individual differences in gesture rate. An intriguing and remarkably successful solution for this problem has been proposed by Bergmann and Kopp (2009), who used large data sets of speech

and gesture to create Bayesian nets that can reproduce gesture and speech behavior of real humans. They were able to reproduce in an artificial agent not only the production of gestures in referring expressions but also interpersonal variation in the type and rate of gestures.

Finally, we want to address the issue of why speakers use gesture and speech redundantly in their generation of referring expressions. We suggest two reasons. First, our data are obtained from directors' initial referring expressions, produced before there was any feedback from matchers. So the director and matcher have not had the opportunity to engage in *grounding* (Clark, 1996). This also applies to other studies with confederates that give minimal feedback (e.g., Melinger & Kita, 2007) or none at all. Such studies often find that speakers overspecify their referring expressions (Engelhardt, Bailey, & Ferreira, 2006; for a gesture-related study, see Van der Sluis & Krahmer, 2007). Following the principle of *least collaborative effort* (Bard et al., 2007; Clark & Wilkes-Gibbs, 1986), our directors may also have decided to use gesture and speech redundantly to increase the communicative effectiveness of their initial utterance (Van der Sluis, 2005; Van der Sluis & Krahmer, 2007). Second, as So et al. also suggest in one of their candidate explanations, the reason may be that gesture and speech originate from a single underlying cognitive representation, as suggested by McNeill's (1992; McNeill & Duncan, 2000) *growth point* theory. This underlying representation could be a meta-modal analog of Levelt's (1989) *preverbal message*, which gets expanded into speech and gesture during utterance production. Neither of these explanations needs to invoke a speaker-internal facilitatory function of gesture.

Only when the referring process becomes a truly collaborative enterprise does the need for redundancy decrease, as many studies on common ground have shown (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Gerwing & Bavelas, 2004; Holler & Stevens, 2007). We therefore argue that the observed redundancy between gesture and speech during initial referring is communicatively motivated: Redundant signals are more likely to be decoded correctly and are, thus, an efficient strategy to use in initial situations where uncertainty about what the matcher has understood is high. Our only result that was consistent with the tradeoff hypothesis, the finding that use of conceptual pacts decreases the rate of pointing gestures, supports this argument.

Note

1. One could also call these gestures *demonstrative* iconic gestures.

Acknowledgments

The work of J.P. de Ruiter was supported by the EU Integrated Project JAST (Joint Action Science and Technology) Grant FP6-IST2-003747. The authors wish to thank Ielka van der Sluis for helpful discussions, Annett Jorschick for her assistance with the statistical analyses, and Claudia Wild for her work on doublecoding the data.

References

Alibali, M. W., 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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

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., Chen, Y., Nicholson, H., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language*, *57*, 616–641.

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

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.

Bergmann, K., & Kopp, S. (2009). *GNetic—Using Bayesian decision networks for iconic gesture generation*. Paper presented at the IVA09, Amsterdam, The Netherlands.

Boersma, P., & Weenink, D. (2007). *Praat: Doing phonetics by computer (version 4.2)*. Available at: http://www.praat.org.

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

Brugman, H., & Russel, A. (2004). Annotating multimedia/multi-modal resources with ELAN. In *Proceedings of LREC 2004, fourth international conference on language resources and evaluation*, Lisbon, Portugal.

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

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

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. Leny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam, The Netherlands: North Holland.

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

De Ruiter, J. P. (1998). *Gesture and speech production*. Unpublished Doctoral Dissertation, Radboud University, Nijmegen.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge, UK: 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. (2007). Postcards from the mind: The relationship between thought, imagistic gesture, and speech. *Gesture*, *7*, 21–38.

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*, 554–573.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391.

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

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

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*, 267–289.

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

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effect of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, *82*, 449–462.

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

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*, 185–209.

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

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

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

Keysar, B., Barr, D. J., & Balin, J. A. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, *39*, 1–20.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation. The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32–38.

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

Kranstedt, A., Kopp, S., & Wachsmuth, I. (2002). MURML: A multimodal utterance representation markup language for conversational agents. AAMAS'02 Workshop on Embodied Conversational Agents, Bologna, Italy.

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

Krauss, R. M., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal, and visible factors in judgments of another's affect. *Journal of Personality and Social Psychology*, *40*, 312–319.

Krauss, R. M., Chen, Y., & Chawla, P. (1995). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 389–450). Tampa, FL: Academic Press.

Krauss, R. M., Chen, Y., & Gottesmann, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge, UK: Cambridge University Press.

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

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.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*, *39*, 723–730.

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

McNeill, D. (1992). *Hand and mind*. Chicago: The Chicago University Press.

McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge, UK: Cambridge University Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*, 473–500.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*, 119–141.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology*, *117*, 411–424.

Roelofs, A. (1992). A spreading activation theory of lemma retrieval in speaking. *Cognition*, *4*, 2.

Schegloff, E. A. (1972). Notes on a conversational practice: Formulating place. In D. N. Sudnow (Ed.), *Studies in social interaction* (pp. 75–119). New York: McMillan, The Free Press.

Schriefers, H., De Ruiter, J. P., & Steigerwald, M. (1999). Parallelism in the production of noun phrases: Experiments and reaction time models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 702–720.

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 der Sluis, I. (2005). *Multimodal reference*. Unpublished Doctoral Dissertation, Katholieke Universiteit Brabant, Tilburg.

Van der Sluis, I., & Krahmer, E. (2004). *The influence of target size and distance on the production of speech and gesture in multimodal referring expressions*. Paper presented at the The 8th International Conference on Spoken Language Processing (ICSLP), Jeju Island, Korea.

Van der Sluis, I., & Krahmer, E. (2007). Generating multimodal references. *Discourse Processes*, *44*, 145–174.

Wesp, R. K., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, *114*, 591–600.