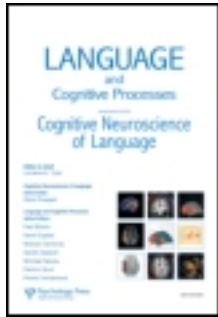


This article was downloaded by: [Stony Brook University]

On: 18 July 2013, At: 00:16

Publisher: Routledge

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



## Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/plcp20>

### Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture

Alexia Galati<sup>a b</sup> & Susan E. Brennan<sup>a</sup>

<sup>a</sup> Department of Psychology , Stony Brook University , SUNY, Stony Brook , NY , 11794-2500 , USA

<sup>b</sup> Department of Psychology , University of Cyprus , PO Box 205371678, Nicosia , Cyprus

Published online: 29 May 2013.

To cite this article: Language and Cognitive Processes (2013): Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture, Language and Cognitive Processes, DOI: 10.1080/01690965.2013.796397

To link to this article: <http://dx.doi.org/10.1080/01690965.2013.796397>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture

Alexia Galati<sup>a,b\*</sup> and Susan E. Brennan<sup>a</sup>

<sup>a</sup>Department of Psychology, Stony Brook University, SUNY, Stony Brook, NY 11794-2500, USA; <sup>b</sup>Department of Psychology, University of Cyprus, PO Box 205371678, Nicosia, Cyprus

(Received 30 May 2012; final version received 4 April 2013)

Are gesturing and speaking shaped by similar communicative constraints? In an experiment, we teased apart communicative from cognitive constraints upon multiple dimensions of speech-accompanying gestures in spontaneous dialogue. Typically, speakers attenuate old, repeated or predictable information but not new information. Our study distinguished what was new or old for speakers from what was new or old for (and shared with) addressees. In 20 groups of 3 naive participants, speakers retold the same Road Runner cartoon story twice to one addressee and once to another. We compared the distribution of gesture types, and the gestures' size and iconic precision across retellings. Speakers gestured less frequently in stories retold to Old Addressees than New Addressees. Moreover, the gestures they produced in stories retold to Old Addressees were smaller and less precise than those retold to New Addressees, although these were attenuated over time as well. Consistent with our previous findings about speaking, gesturing is guided by both speaker-based (cognitive) and addressee-based (communicative) constraints that affect both planning and motoric execution. We discuss the implications for models of co-speech gesture production.

**Keywords:** audience design; gesture; dialogue; common ground; spontaneous speech; storytelling; partner-specific adaptation; communication

Gesturing is ubiquitous during communication. Like speech, gestures can be expressive in multidimensional ways, encoding both propositional and imagistic information (McNeill & Duncan, 2000), representing abstract information (e.g., Müller, 1998), adding emphasis (McClave, 1994), and conveying arousal and emotion (de Meijer, 1989; Wallbott, 1998). Gestures and speech have been proposed by some to emerge from the same planning process and to form an integrated message in communication (e.g., Clark, 1996; Goodwin, 2000; Kendon, 1994; McNeill, 1985, 1992). Gestures are semantically co-expressive with speech (e.g., McNeill, 1985, 1992), pragmatically integrated with speech (Kelly, Barr, Church, & Lynch, 1999), and temporally coordinated with speech (Mayberry & Jaques, 2000; McClave, 1994).

Thus, speech and gesture are coordinated in complex ways in spontaneous communication; in this paper we consider how speech and co-speech gestures are shaped by the same speaker- and addressee-related factors all the way through formulation and articulation. Current models of gesture production are usually based on research that explores how gestures facilitate the speakers' own cognitive processes, such as lexical retrieval (Krauss, Chen, & Gottesman, 2000) or

packaging information for speaking (Kita & Özyürek, 2003). Fewer models address how gestures may be produced with the addressee's needs in mind (de Ruiter, 2000; McNeill & Duncan, 2000), and such models do not specify how contextual or addressee-specific information is represented such that it can affect planning processes. Although cognitive (*for-the-speaker*) and communicative (*for-the-addressee*) functions of gestures have long been acknowledged (Bavelas & Chovil, 2000; Kendon, 1994), current frameworks do not account for how cognitive and communicative constraints shape gesture planning jointly.

In the current study, we extend our prior work on effects of cognitive and communicative constraints on speaking (Galati & Brennan, 2010) to examine effects of cognitive and communicative constraints on gesturing at different grains of production (encoding semantic content vs. executing surface form). In particular, we ask whether for-the-speaker and for-the-addressee constraints affect the planning and motoric execution of gestures just as they affect the planning and articulation of speech (Galati & Brennan, 2010). By considering how cognitive and communicative effects emerge at different junctures in planning, we aim to elucidate how these two expressive modalities interact.

\*Corresponding author. Email: galati@ucy.ac.cy

We begin by reviewing research on both cognitive and communicative factors that have been shown individually to constrain gesture production, along with the processing stages these factors have been shown to affect. Then we discuss how current theories of gesture production can accommodate communicative and cognitive factors and identify the processes for which these theories posit (explicitly or implicitly) that speech and gesture planning interact. Finally, we present our study and its implications for the planning and production of co-speech gesture.

### Factors shown to constrain gesture production

Gestures have been shown to be guided by speakers' cognitive needs, including facilitating lexical retrieval, managing cognitive load and organising information into constituents appropriate for speaking; in other words, gesturing helps people speak. Speakers prevented from gesturing become more dysfluent, producing more filled pauses and slower speech (Morsella & Krauss, 2004; Rauscher, Krauss, & Chen, 1996) and their descriptions become less vivid (Rimé, Schiatura, Hupet, & Ghysseleux, 1984). Moreover, when speakers have rehearsed their speech, they produce fewer gestures than when they speak spontaneously (Chawla & Krauss, 1994).

Gesture production also helps speakers manage cognitive load. When performing dual tasks, such as explaining a math problem while remembering a string of letters, speakers recall more letters when they are permitted to gesture than when they are not (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). Pointing gestures also help manage cognitive load, facilitating performance on counting tasks for both young children (Alibali & DiRusso, 1999) and adults (Carlson, Avraamides, Cary, & Strasberg, 2007). Speakers gesture more frequently when describing objects that are difficult to encode verbally or have to be described from memory (Morsella & Krauss, 2004). They also gesture more when having to figure out what to describe first or when resuming a task after performing another task using the same cognitive resources (Melinger & Kita, 2007).

The way in which information is packaged in speech also influences gesturing, presumably because gesturing helps with organising information into syntactic constituents. Explanations elicit different patterns of gestures than descriptions, even when controlling for the semantic content of speech; explanations, which are more constrained and complex than descriptions, elicit more non-redundant, representational gestures – gestures that depict semantic content by virtue of handshape, placement and motion (Alibali, Kita, & Young, 2000). Also, differences in gesture mirror cross-

linguistic differences in how the manner and path of motion events are encoded in the sentence, either within verb forms or as adverbials (Kita & Özyürek, 2003).

However, co-speech gesture is guided not only by these cognitive factors, but also by communicative factors, which involve taking into account the informational needs or perspectives of addressees. For instance, visual co-presence (whether conversational partners share the same environment and can see one another) constrains gesturing. Speakers generally gesture less frequently when addressees cannot see them (Cohen, 1977; Cohen & Harrison, 1973). More recently, there have been several demonstrations that the visibility between conversational partners affects some types of gestures more than others. In a study by Bavelas, Chovil, Lawrie, and Wade (1992), there were increased “interactive” gestures (associated with managing the dialogue) when partners were visually co-present; in another study by Alibali, Heath, and Myers (2001) there were increased representational gestures. When speakers can see their partner they also adapt their pointing gestures, being more likely to point as their distance from the target display increases (Bangertter, 2004). Speakers also produce larger gestures, more gestures that were non-redundant with speech, more interactive gestures and more verbal references to their gestures when talking face-to-face than when talking over the telephone or into a tape recorder (Bavelas, Gerwing, Sutton, & Prevost, 2008). Finally, speakers adapt their gestures according to the space they share with their addressees, as determined by their relative locations; during expressions of motion events, the directionality of gestures accompanying the words *in* and *out* depends on the directionality of these words in the shared space (Özyürek, 2000, 2002).

Yet another factor that constrains gesture production is common ground (or what interlocutors mutually know). Reference to information already in common ground is often either attenuated or omitted. For instance, speakers describing the location of large, salient targets were more likely to express size information in gestures to addressees with whom they *had not* previously studied displays of the targets than to those with whom they *had* (Holler & Stevens, 2007). Also, speakers encoded more semantic features in their gestures when they had not watched the narrated clip together with an addressee than when they had (Holler & Wilkin, 2009). Similarly, in a study where speakers retold the same story repeatedly either to the same confederate addressee or to new confederate addressees, gesture rates decreased with each retelling to the same addressee but not across retellings to different addressees (Jacobs & Garnham, 2007). Common ground can also affect qualitative aspects of

gesture; speakers describing toys to addressees who had played with the same toys produced gestures that were less complex, less precise, and less informative compared to speakers describing the toys to addressees who had played with a different set of toys (Gerwing & Bavelas, 2004).

Linguistic context can also constrain gesture production. Levy and McNeill (1992) observed that utterances encoding information that is new or thematic for future discourse (e.g., introductory mentions of a character) involves more complex referring expressions in speech and more gesturing than does already-mentioned or presupposed information. Indeed, gestures are less likely to occur with presupposed information, such as zero and unstressed pronoun references, and more likely to occur with more complex noun phrases and predications (McNeill & Levy, 1993). In addition to informational factors, less-studied factors such as arousal, emotion and individual differences could affect gesture planning as well.

In sum, quite a variety of constraints have been shown, at least individually, to affect the *formulating* of preverbal messages into co-speech gestures, whereas much less is known about constraints on gestures' *motoric execution* (although common ground has been shown to be one of these constraints; Gerwing & Bavelas, 2004). A model of co-speech gesture production should ultimately account for how cognitive, social and contextual factors affect both the planning and execution of gestures, as well as for how gesture and speech planning may be co-constrained.

### Theories of gesture production

Current theories of gesture planning can be distinguished by the assumptions they make about: (1) modularity or architectural barriers in planning, that is, whether computations performed during a given planning process are independent from and therefore unaffected by those performed elsewhere in the system, (2) the points in planning at which speech and gesture interface and (3) the extent to which communicative factors can affect gesture planning.

Some theories that can be classified as modular take as a starting point Levelt's (1989) information-processing model of speech production, which involves cascading stages for conceptualising, formulating and articulating; a gesture module is linked at some stage to a speech module (e.g., de Ruiter, 2000; Krauss et al., 2000). One such proposal by Krauss et al. (2000), focusing on how gestures facilitate lexical retrieval, links the gesture module to the speech module at the level of working memory and at Levelt's formulator stage (where words are retrieved and filled into

syntactic forms). According to this *Lexical Gestures* proposal (Krauss et al. 2000), spatio-motoric features selected in working memory cross-modally prime semantic features, facilitating the access of lemmas and word forms. This modular theory focuses solely on speaker-internal constraints (namely, lexical access) and does not make predictions about how communicative or contextual factors could affect gesture planning (see de Ruiter, 2000, for a related criticism).

Another modular proposal by McNeill 2000, the *Sketch Model*, differs from *Lexical Gestures* (Krauss et al., 2000) both in its assumptions about the function of gesture and the stages at which the speech and gesture modules interface. Rather than viewing gestures as artefacts of lexical retrieval, the *Sketch Model* posits that gesture serves as "a communicative device from the speaker's point of view" (de Ruiter, 2000, p. 292). Gestures are initiated and linked to speech production at Levelt's conceptualiser stage (before filling in lexical items in syntactic constituents at the formulator stage), such that a preverbal message and a representation with imagistic and spatio-temporal information (the "sketch") emerge simultaneously. As the preverbal message is sent to the speech formulator, the gesture planner constructs a motor programme based on the "sketch" to send to the motor execution unit. At the articulation/execution stage, speech and gesture proceed ballistically since there is no interaction between the two modalities at lower levels of processing. Synchrony is instead achieved only at the onset of the message, when the gesture planner sends a message to the conceptualiser that speech formulation can be initiated. In the *Sketch Model*, communicative and contextual factors can affect gesture production only if represented during higher-level planning, for instance by keeping track of anaphoric references in the form of a discourse record during the conceptualisation of the message (McNeill 2000, p. 289).

Another proposal that presupposes Levelt's (1989) information-processing model but that extends the coordination of speech and gesture through the formulation phase is Kita and Özyürek's (2003) *Interface Hypothesis*. This proposal aims to explain how gestures are constrained not only by the semantic properties of their referents, but also by the possibilities of linguistic encoding. On this view, what corresponds to Levelt's Conceptualiser is divided into a Communication Planner and a Message Planner. The Communication Planner takes into account the "communicative intention" (i.e., performs the "macro-planning" of Levelt's Conceptualiser), determining what information should be expressed, in roughly what order, and by which modality. The Communication Planner's resulting output is then sent to the Message Generator and an Action Generator – a general mechanism for generat-

ing both gestures and instrumental actions. The Message Generator selects propositions for formulating in speech based on the communicative goals and context (i.e., performs the “micro-planning” of Levelt’s Conceptualiser); links between the Message Generator and the Action Generator enable coordinating what information to encode in speech and in gesture. Bidirectional feedback between the Speech Formulator and Message Generator can account for gestures being shaped by constraints for linguistic encoding (via the Message Generator’s link to the Action Generator). The Interface Hypothesis focuses mainly on cognitive and linguistic constraints – namely, on how the demands for packaging information in speech affect the distribution of semantic (particularly, spatiomotoric) features in gesture. Communicative factors are not explicitly addressed, although presumably they can be handled by the Communication Planner. As with the Sketch Model, no coordination is posited between the motoric articulation of speech and gesture.

Although it is intuitively appealing to consider gesture planning from the perspective of distinct stages of speech planning, there is evidence of interaction and fine-grained coordination between speech and gesture during virtually every point in their production. For instance, gestures are synchronous with tone group nuclei (McClave, 1994), and are suspended during stuttering (Mayberry & Jaques, 2000) and before speech repairs (Seyfeddinipur, 2006). To reconcile the architecture of co-speech gesture production with these findings would require additional links between articulatory and monitoring processes across the speech and gesture that are not currently posited by the modular accounts.

Non-modular accounts tend to be less explicit about how speech and gesture planning is coordinated, and instead focus on accounting for effects of the discourse record, common ground and other speaker-external constraints. In one such theory, gestures are thought to arise from representations – known as “growth points” – encoding imagery and linguistic content that mark significant contrasts in the immediate context (McNeill & Duncan, 2000). Although Growth Point Theory highlights the contribution of speaker-external constraints, it does not clarify what constitutes a significant contrast in the unfolding discourse context and does not make specific predictions about how multiple constraints affect gesture production.

The framework closest to ours, Hostetter and Alibali’s (2008, 2010) *Gesture as Simulated Action* framework, is agnostic about any architectural barriers in gesture production and permits multiple factors to influence production. According to this framework, a representational gesture is produced when activation exceeds a certain threshold during neural simulation of

the action event that is to be described. So a speaker’s likelihood of expressing a simulated action as a gesture depends not only upon the neural activation arising from simulating the action, but also upon the current threshold for a gesture. This threshold may be affected by additional neural factors (e.g., connections between motor and premotor areas), cognitive factors (e.g., working memory constraints) or aspects of the communicative situation (e.g., whether what is being described is difficult for the audience). Nonetheless, the Gesture as Simulated Action framework does not make explicit predictions about the gesture’s motoric execution. The motoric execution of a gesture presumably could be affected by analogue features of the mental simulation underlying it, but in their current formulation Hostetter and Alibali do not specify whether adjustments in the gesture threshold can shape not only whether a gesture is produced (adapting the gesture rate, a quantitative adaptation), but also the gesture’s shape or size (qualitative adaptations).

Beginning with Hostetter and Alibali’s (2008, 2010) view that several factors can constrain gesture production simultaneously, we aim to extend their framework by examining how the effects of both cognitive and communicative factors may be distributed across different grains of speech and gesture planning.

### The current study

Spontaneous speaking is flexible enough to be influenced by addressees’ needs, not only during conceptualisation and formulation but also during articulation, as long as such needs are simple and known to the speaker (Galati & Brennan, 2010). In the current study, we aim to discover whether co-speech gesture is shaped at the same grains and by the same constraints as is speech. Previously we examined measures of spoken content (number of narrative events realised, number of words, amount of detail, and lexical perspective) and articulation (the duration and intelligibility of lexically identical expressions) for the same speaker across three retellings of the same story (with the second and third retellings being to either a new addressee or to the same addressee as the first retelling) (Galati & Brennan, 2010). This design offered the advantage of teasing apart the speaker’s perspective from the addressee’s, enabling direct comparisons of stories under  $\text{Speaker}_{\text{new}}\text{-Addressee}_{\text{new}}$ ,  $\text{Speaker}_{\text{old}}\text{-Addressee}_{\text{old}}$  and  $\text{Speaker}_{\text{old}}\text{-Addressee}_{\text{new}}$  conditions. For all these speech measures except lexical perspective and duration, we found that stories retold to Old Addressees were attenuated compared to those retold to New Addressees. In the current study we coded and analysed the co-speech gestures from the

same speech corpus to examine the distribution of for-the-speaker and for-the-addressee effects on gestures for both planning (as reflected by the distribution of different types of gestures) and motoric execution (as reflected by adaptation in the size and iconic precision of gestures).

A study by Jacobs and Garnham (2007) used a design similar to ours, but focused only on quantitative aspects (measuring gesture *rates*) rather than their forms. In that study, speakers retold comic strip stories in four conditions, two of which are of interest here: speakers repeated one story three times to the same addressee, and they repeated another story three times, but each time to one of three different addressees. Speakers gestured less frequently with each retelling to the same addressee but not across retellings to different addressees, demonstrating that addressee's knowledge affected gesture formulation. These findings are relevant to our study in that they isolate for-the-addressee effects; however, our study goes beyond Jacobs and Garnham's by (a) measuring adaptation not only in gesture frequency but also in motoric execution, (b) considering adaptation in gesture alongside adaptation in speech and (c) not only dissociating the perspective of the speaker from the perspective of the addressee, but also allowing direct comparisons between the same narrative retold to the same addressee (Speaker<sub>old</sub>-Addressee<sub>old</sub>) versus to a new addressee (Speaker<sub>old</sub>-Addressee<sub>new</sub>).

### Design

Speakers told a story to two addressees a total of three times: after telling the story for the first time to a naive addressee, they retold the same story to the same addressee and then to a new naive one, or vice-versa. Therefore, the status of information (old vs. new) varied across narrations (Speaker<sub>new</sub>-Addressee<sub>new</sub>, Speaker<sub>old</sub>-Addressee<sub>old</sub> and Speaker<sub>old</sub>-Addressee<sub>new</sub>). By dissociating the perspective of the speaker from the perspective of the addressee through retellings to the same or to a new addressee, we could tease apart the effects of cognitive and communicative factors on gesturing. Adaptations in gesture made while retelling the story to a new addressee could be clearly attributed to be for-the-addressee and were not confounded by the speakers' own perspective (see Keysar, 1997, for a related discussion). The counterbalancing of the addressee's identity in the second and third retellings was done to account for potential hypermnesia (increased recall with repeated attempts; see Payne, 1987, for a review) and other order effects (such as fatigue or forgetting) that might wash out any effect of partner-specific adaptation.

As a measure of content-related adaptation in gesture production, we focused on the distribution of different gesture types. We reasoned that encoding the intended message into a co-speech gesture should involve selecting whether to depict certain semantic features in analogue fashion (with a representational gesture), or else to not depict any such features and instead mark emphasis (with a beat gesture) or aspects of the dialogue process (with a metanarrative gesture). In other words, selecting a gesture type should be part of gesture formulating or planning. To examine surface-form-related adaptation, we considered the size of a gesture and its iconic precision. Adaptation in size and iconic precision that cannot be attributed to changes in the narrative content (which we kept constant) should reflect adjustments in motoric execution.

### Predictions

Alongside our previous findings for speech (Galati & Brennan, 2010), we were able to test predictions about how the two modalities are coordinated – specifically, whether they are subject to the same patterns of adaptation all the way from planning to articulation. If adaptation in gesturing parallels that in speaking, this would support speech and gesture production being closely coordinated during both planning and articulation (as opposed to proceeding ballistically, or without coordination once each modality is launched). If on the other hand adaptation in gesturing does not parallel that in speaking, this would suggest that, even if speech and gesture share representations during conceptualising (e.g., de Ruiter, 2000), their formulating and articulating may unfold more independently.

Consistent with our previous findings for speech, we predicted that since representational gestures encode semantic content, they would particularly be sensitive to addressees' informational needs, and thus their rate would be attenuated more in retellings to old addressees than to new addressees. We also predicted that the rate of metanarrative and beat gestures would increase more in retellings to old addressees than to new addressees, consistent with the findings that the rate of interactive gestures increases as common ground increases (e.g., when narrating in dialogue rather than in monologue, Bavelas, Chovil, Coates, & Roe, 1995). That addressee knowledge can affect gesture frequency is also supported by Jacobs and Garnham's (2007) findings, although their task did not reveal reliable differences in the distributions of representational and other types of gestures. If speakers are more likely to produce fewer gestures when retelling stories to Old Addressees than to New Addressees, this would parallel the adaptation we found for utterance planning

and would suggest that speech and gesture formulation are closely coordinated.

Seeing that few studies have examined qualitative adjustments in gestures (e.g., Bavelas et al., 2008; Gerwing & Bavelas, 2004; Kuhlen, Galati, & Brennan, 2012), and none have considered them in conjunction with articulatory adjustments in speech, our predictions about the motoric execution of gestures were more exploratory. Motoric adjustments in gestures could parallel those in speech, such that speakers are more likely to attenuate their gestures' size and precision when retelling stories to Old Addressees than to New Addressees. This would suggest that speech and gesture articulation interface in a way that permits them both to be shaped by partner-specific information.

Alternatively, adaptation in gestures' motoric execution may pattern differently than spoken articulation. This would suggest that the two modalities involve more independently planned or encapsulated processes. Such a pattern would be expected with the "Dual Process Model" (proposed for speaking by Bard & Aylett, 2001; Bard et al., 2000), in which automated processes like articulation are encapsulated from communicative constraints and default to being egocentric, whereas other more inferential processes like the planning of referring expressions can be guided by the partner's needs. This modular "Dual Process Model", extended to gesture production, would predict that the distribution of gesture types, reflecting a more inferential process, may be influenced by for-the-addressee factors (e.g., the addressee's identity), but the motoric execution of gestures should be influenced only by for-the-speaker factors, with gestures attenuating their size and iconic precision with each retelling as the stories became more accessible to speakers. Although our findings on spoken articulation do not support such encapsulation within speech planning (Galati & Brennan, 2010), it is possible that gesture processes are structured differently than in speech and involve more independent computations.

## Method

In triads of naive participants, one person acted as *speaker* and the other two as *addressees*. Speakers watched a cartoon and narrated it three times: twice to the same addressee (A1) and once to the new addressee (A2); the order of addressees (same vs. new addressee) in the second and third narrations was counterbalanced (A1-A1-A2 or A1-A2-A1). We examined adaptation in terms of the number of gestures, the distribution of gesture types and the relative size and iconic precision of gestures produced by speakers across retellings.

## Participants

Sixty-nine students from Stony Brook University were grouped into 23 triads. Speakers were all native English speakers; addressees were all fluent in English. Three of the triads were excluded due to idiosyncrasies of the speaker or addressee.<sup>1</sup> Of the remaining 60 participants, 20 served in the role of the speaker and 40 in the role of the addressee. Forty-five of these participants were female and 15 were male. In none of the 20 triads did participants know each other in advance. Speakers were recruited separately from addressees to ensure their native English speaker status. Addressees were randomly assigned to the roles of A1 (the first addressee to hear the speaker tell the story) and A2. Participants were compensated with research credit that could be used to fulfil a requirement in a psychology course.

## Materials

A Looney Tunes animated cartoon without dialogue (entitled "Beep Beep") starring Road Runner and Wile E. Coyote was used to elicit narratives.<sup>2</sup> The cartoon was edited for length such that it had four distinct episodes, corresponding to four attempts of Coyote to capture Road Runner, 3 min and 10 sec long.

## Procedure

All participants were told that the study investigated storytelling and memory and that the addressees, upon hearing the stories narrated by the speakers, would be tested for their memory of the stories at the end of the session. The expectation of a memory test for addressees provided a reason to pay attention to the story and a rationale for telling or being told the same story twice. Speakers watched the cartoon alone on a computer. After watching it twice, they moved to another room to narrate the story three times. They were informed in advance that they would be narrating the story twice to one of the addressees and once to the other (but not informed who would hear the story twice and in what order). Speakers were asked to narrate in as much detail as possible each time. Addressees were told that they would hear the story once or twice and told to remember as much of the story as possible for the upcoming memory test. They were also told that they could freely comment or ask questions for clarification during the narration. Only one addressee was present in the room during each storytelling session, which was videotaped (with both partners visible) with a digital camcorder. The experimenter was present off to the side during the sessions to supervise the recording; speakers and addressees were told to face each other and ignore the experimenter and camera. When the storytelling sessions were completed, participants were told that there would not in fact be a

memory test for the addressees, and they were all debriefed.

### **Transcribing**

All three narrations of the Road Runner cartoon for each triad were transcribed in detail by the first author. To be able to compare the narrations within and across speakers, we created a script for 85 *narrative elements* for the Road Runner cartoon. A narrative element referred to a proposition or set of propositions forming a sub-event that advances the plot of the story (see Galati & Brennan, 2010, for more details on creating the script and transcribing speech).

### **Coding**

The 20 narrative elements most frequently mentioned by speakers were selected for coding co-speech gestures. Appendix 1 provides a list of these elements. For a narrative element from a given speaker to be included in our coding, it had to be realised in all three retellings in the speech transcript. Out of the possible 400 triplets of narrative elements (20 for each of 20 speakers), 59 were excluded because in one or more narrations the narrative element was omitted. Thus, a total of 341 triplets of narrative elements were included in our coding; 162 of these triplets came from narration order *A1-A1-A2* and 179 came from narration order *A1-A2-A1*.

Video clips of these 341 triplets of narrative elements were excised from the digital recordings. The onsets and offsets of the videos were adjusted so that gestures associated with the narrative elements were included in their entirety, regardless of whether their beginning or end overlapped partially with spoken descriptions of other narrative elements. On each resulting clip, a video effect was applied in Final Cut Express HD to block out the addressee in the video, and each video file was given an uninformative label; these steps were taken to ensure that the coding of speakers' gestures was blind to addressee identity, behaviour or knowledge status.

Video clips of narrative elements were coded in triplets for their gestures' relative size and iconic precision, and were coded individually for gesture number and types. Before coding, the first author identified all irrelevant hand movements that were not gestures (e.g., *self-adaptors*, such as scratching nose, adjusting glasses) and made notations regarding their form and onset in the coders' rating sheets. This was necessary because the coding of relative gesture size and iconic precision was made without sound and it may thus not have been clear whether those movements were meant to represent a character's action.

### **Distribution of gesture types**

Two coders categorised each gesture in each video of the 341 triplets as belonging to one of these types: (1) Representational, (2) Metanarrative, (3) Beats or (4) Combination. To classify gestures, coders considered the semantic features that were encoded in gesture (or a lack of semantic features in the case of beat gestures), the accompanying speech (unlike for the size and precision coding) and the original events within the animated cartoon.

*Representational gestures* (also known as *iconics*, McNeill, 1992, or *illustrators*, Ekman & Friesen, 1969) depict semantic content by virtue of handshape, placement and motion and often represent the movement of characters or properties of objects. Representational gestures, for our purposes, also included pointing gestures that were used to set up or locate characters or objects in gestures space (referred to as *abstract deictic* gestures at the *narrative* level, Cassell & McNeill, 1991).

*Metanarrative gestures* included gestures that facilitate dialogue (referred to as *interactives* by Bavelas et al., 1992, e.g., presenting a single hand with cupped fingers directed towards the addressee while saying "you know") or support discourse cohesion (referred to as *cohesives* by McNeill, 1992). Metanarrative gestures included metaphoric pointing gestures (referred to as *abstract deictic* gestures at the *metanarrative* level, Cassell & McNeill, 1991), such as pointing towards the right while saying "in the Coyote's next attempt".

*Beats* were simple, rhythmic gestures that did not encode semantic content (e.g., Alibali et al., 2001; McNeill, 1992). In our coding we considered beat gestures to be distinct from other metanarrative gestures known as *interactive* gestures, because interactive gestures have been shown to be affected by common ground manipulations (Bavelas et al., 1992, Exp 2; Bavelas et al., 1995), whereas beat gestures on their own have been shown not to be (Alibali, Heath, & Myers, 2001). Coders distinguished metanarrative and beat gestures by considering both the accompanying speech and the gestures' kinetics and form. Gestures were generally classified as beats when they were biphasic with a small motion range (i.e., simple "up-and-down" movement of the wrist or fingers), while lacking a clear semantic relationship to speech and instead emphasising (usually stressed) accompanying words. On the other hand, gestures were classified as metanarrative when they were non-representational gestures but triphasic (involving a preparation, stroke<sup>3</sup> and retraction phase), with a more varied form and greater range, and could be tied to the accompanying



speech or the established discourse context (e.g., to referents established in gesture space).

Finally, we created the category of *Combination* gestures for those gestures that were either (a) representational gestures substantially attenuated to the point that most iconic features were lost or (b) representational gestures on which beat or other metanarrative gestures had been superimposed. An example of a combination gesture from our corpus is an interactive gesture involving “hand-flailing” (with spread fingers and loose wrists) to indicate problems with lexical retrieval or uncertainty when accessing the word *anvil*, superimposed on a representational gesture of Coyote’s hands holding an anvil.

For the 1023 videos (of the 341 triplets), the 2 coders agreed on 92% of the cases on the total number of gestures produced in the videos, 89% of the cases on the number of representational gestures, 94% of the cases on the number of metanarrative gestures, 97% of the cases on the number of beat gestures and 89% of the cases on the number of combination gestures. Most of the disagreements concerned the classification of representational versus combination gestures, which is not surprising since combination gestures by definition included representational features. The first author’s coding of gesture types was used in the data analyses.

#### *Measuring relative gesture size*

Two coders (the first author and a different undergraduate research assistant than the one who coded gesture types) watched the 341 triplets of videos without sound and rated the relative size of the gestures produced by the speaker (both did 100% of the coding). Gesture size was defined as the amount of space that the speakers’ hands spanned when gesturing. This involved both the displacement of the speakers’ hand during gesturing (e.g., the length of the upwards trajectory of a speaker’s hand when representing Coyote being propelled into the air by the tightrope) and also, in the case of two-handed gestures, the space between the speakers’ hands (e.g., the space between two hands when the speaker is representing Coyote holding an anvil). When judging gesture size, coders considered the largest displacement of the hand in one-handed gestures or the largest distance between the hands in two-handed gestures during the entire gesture rather than just during the gesture stroke. The coders recorded the relative size of gestures across the three triplets by entering their judgement (using the video labels) for each of the three on a 1–7 scale with 0.5 point increments. If the speaker did not produce a gesture while describing a particular event, gesture size was coded as 0. Since speakers’ self-adaptors (e.g., adjusting glasses) were annotated in the coders’ rating

sheets in advance, coders excluded them from judgements of the gestures’ size.

In determining the relative ordering, the coders could watch the three videos within a triplet in any order and as many times as they needed to. If they judged that the speaker’s gestures in two or all three of the videos used the same amount of space, they could assign the same rating to those videos. When a speaker clearly repeated a gesture (when the same motion was performed repeatedly without the hands returning to rest), coders did not consider the gesture space used in subsequent movements to increase the size of the gesture. However, if one of the repeated gestures was larger than the initial one, the size of the largest repetition determined the coders’ judgement.

To assess reliability between the two coders, we calculated the proportion of video triplets for which both coders ranked the three videos in the same order for gesture size; this occurred in 83% of the cases. In another 12% of the cases the coders agreed on which video included contained gestures of either the largest or smallest size, but disagreed on the relative ranking of the remaining two videos. Only in 5% of the cases did the coders make completely different judgements about the relative ranking of the three videos in terms of gesture size. Because the undergraduate coder was also blind to the experimental hypotheses, his judgements were used in the data analyses.

#### *Measuring relative gesture precision*

After coding the 341 triplets for gesture size, the two coders who coded gesture size coded the same triplets for iconic precision relative to the action in the original cartoon. Iconic precision was defined as the similarity between a gesture from a narration and the original cartoon event. For each triplet, the coders watched the part of the original cartoon that was described in that triplet and then judged the three associated narrative clips (without sound). As with coding gesture size, the coders rated iconic precision for each video on a 1–7 scale with 0.5 point increments. If the speaker did not produce any gestures in a video, the degree of iconic precision for that video was coded as 0. Coders could also assign a 0 rating when they judged that the speakers’ gestures did not convey any of the semantic information in the stimulus clip. This was the case, for example, when speakers produced beat gestures, which could not be mapped onto any representational aspect of the cartoon stimulus.

Coders could replay both the original cartoon segment and the narratives as many times they wanted in order to make these judgements. If they judged that the speaker in two or all three videos of a triplet produced gestures with the same degree of iconic

precision, they could assign the same rating. When a speaker clearly repeated a gesture, the coders considered the repetition matching the event in the stimulus most closely to determine their judgement of iconic precision for that video.

Reliability for iconic precision coding was assessed as it was for relative gesture size coding. In 85% of the cases the two coders ranked the three videos identically; in another 10%, the coders agreed on which video included the highest or lowest degree of iconic precision, but disagreed on the relative ranking of the remaining two videos. Only in 6% of the cases did they make rank the three videos completely differently in terms of iconic precision. As before, the undergraduate coder's judgements were used in the data analyses.

**Analyses**

Analyses were 3 × 2 analyses of variance (ANOVAs) with knowledge status (Speaker<sub>new</sub>-Addressee<sub>new</sub>, Speaker<sub>old</sub>-Addressee<sub>old</sub> or Speaker<sub>old</sub>-Addressee<sub>new</sub>) as a within-subjects factor and addressee order (A1-A1-A2 vs. A1-A2-A1) as a between-subjects factor. To examine the distribution of gesture types, the ANOVA also included gesture type as a factor (representational, beats, metanarrative and combination). Two planned contrasts were examined: the first compared the speaker's production for the first telling to A1 and the retelling to A2<sub>new</sub> (S<sub>new</sub> vs. S<sub>old</sub>), and the second compared the speaker's production for the retelling to A1 and the telling to A2 (A1<sub>old</sub> vs. A2<sub>new</sub>), both of which were old for the speaker. The first contrast was speaker-centred, focusing on whether speakers attenuated their gesture production over time regardless of the knowledge status of the addressee. The second contrast was addressee-centred, focusing on whether the speaker's production differed depending on the addressee's knowledge status. For each result, we report two analyses:  $F_1$  is the analysis by subjects (for which means are computed for triads of participants) and  $F_2$  is the analysis by items (for which means are computed for script elements).

**Results**

**Adaptation in the distribution of gesture types**

In this narrative task, speakers produced primarily representational gestures: overall, the corpus included 1553 representational gestures, 213 metanarrative gestures, 50 beat gestures and 362 combination gestures. Gesture frequency differed significantly across these gesture types,  $F_1(3, 54) = 164.54, p < .001$ ;  $F_2(3, 57) = 108.52, p < .001$ . Table 1 shows the distribution of gesture types according to the addressee's knowledge.

The distribution of gesture types depended on addressee knowledge, at least when generalising across speakers, as suggested by a reliable interaction of the two factors by subjects,  $F_1(6, 108) = 4.69, p < .001$ ;  $F_2(6, 114) = .72, ns$ .

Given our previous findings on partner-specific adaptation in speech planning (Galati & Brennan, 2010), and in so far as the two modalities are coordinated during planning, we expected that content-related adaptation in gesture planning would reflect sensitivity to addressees' knowledge. To test our prediction that representational gestures would be particularly sensitive to addressees' informational needs, we first examined whether there was a lower rate of representational gestures than other gestures in retellings to the same addressee relative to a new addressee. The *for-the-addressee* effect was greater for representational gestures than for metanarrative gestures ( $F_1(1, 18) = 13.82, p < .01$ ;  $F_2(1, 19) = 9.59, p < .01$ ), beats ( $F_1(1, 18) = 10.32, p < .01$ ;  $F_2(1, 19) = 5.89, p < .05$ ) or combination gestures ( $F_1(1, 18) = 11.13, p < .01$ ;  $F_2(1, 19) = 8.55, p < .01$ ).

Because representational gestures were the most common type in our corpus and the most likely to be semantically informative for addressees, we conducted focused ANOVAs on their rate alone to examine whether it depended on addressees' knowledge. If addressees' informational needs affect gesture planning, speakers should produce more representational gestures per narrative element in retellings to addressees for

Table 1. Distribution of gesture types: Means (and SDs) for number of representational, metanarrative, beat and combination gestures that speakers produced per narrative element.

|                         | Speaker <sub>new</sub> -<br>Addressee <sub>new</sub> | Speaker <sub>old</sub> -<br>Addressee <sub>old</sub> | Speaker <sub>old</sub> -<br>Addressee <sub>new</sub> |
|-------------------------|--|--|--|
| <b>Representational</b> |  |  |  |
| <i>A1-A1-A2</i>         | 1.65 (1.25)  | 1.49 (1.11)  | 1.63 (1.28)  |
| <i>A1-A2-A1</i>         | 1.53 (0.99)  | 1.30 (1.08)  | 1.53 (1.20)  |
| <i>Total</i>            | 1.59 (1.12)  | 1.39 (1.10)  | 1.58 (1.23)  |
| <b>Metanarrative</b>    |  |  |  |
| <i>A1-A1-A2</i>         | 0.15 (0.46)  | 0.19 (0.48)  | 0.17 (0.41)  |
| <i>A1-A2-A1</i>         | 0.24 (0.57)  | 0.26 (0.65)  | 0.22 (0.59)  |
| <i>Total</i>            | 0.20 (0.53)  | 0.23 (0.58)  | 0.20 (0.51)  |
| <b>Beat</b>             |  |  |  |
| <i>A1-A1-A2</i>         | 0.03 (0.17)  | 0.05 (0.24)  | 0.06 (0.28)  |
| <i>A1-A2-A1</i>         | 0.03 (0.21)  | 0.07 (0.35)  | 0.06 (0.28)  |
| <i>Total</i>            | 0.03 (0.19)  | 0.06 (0.30)  | 0.06 (0.28)  |
| <b>Combination</b>      |  |  |  |
| <i>A1-A1-A2</i>         | 0.35 (0.71)  | 0.37 (0.65)  | 0.31 (0.63)  |
| <i>A1-A2-A1</i>         | 0.38 (0.76)  | 0.36 (0.81)  | 0.35 (0.75)  |
| <i>Total</i>            | 0.37 (0.73)  | 0.36 (0.74)  | 0.33 (0.69)  |
| <b>Total</b>            |  |  |  |
| <i>A1-A1-A2</i>         | 2.19 (1.54)  | 2.10 (1.27)  | 2.17 (1.39)  |
| <i>A1-A2-A1</i>         | 2.18 (1.46)  | 1.99 (1.57)  | 2.15 (1.61)  |
| <i>Total</i>            | 2.18 (1.50)  | 2.04 (1.43)  | 2.16 (1.51)  |

whom the information is new compared to addressees for whom the information is old. As shown in Table 2, the difference in the number of representational gestures per narrative element between retellings to  $A1_{old}$  and  $A2_{new}$  was significant by subjects but not by items (a marginal *for-the-addressee* effect), whereas the number of representational gestures per narrative element did not differ significantly between the tellings to  $A1_{old}$  and  $A2_{new}$  (no *for-the-speaker* effect).

There were no *for-the-speaker* or *for-the-addressee* differences among the other types of gestures. We had hypothesised that the rate of metanarrative gestures might increase in retellings to old addressees than to new addressees, given evidence that interactive gestures become more frequent as common ground increases (Bavelas et al., 1995); however, metanarrative gestures showed only a numerical increase in the retelling to the same addressee relative to the retelling to a new addressee that was not statistically reliable. Representational gestures seem to drive the effects observed for the total number of gestures produced per narrative element. As shown in Table 1, speakers gestured less overall in the retelling to A1 than in the *first* telling to A1 and the retelling to A2.

#### Alternative measure of gesture rate

So far we have considered gesture rate in terms of the number of gestures per narrative element, without normalising for the number of words. In our previous data-set on partner-specific effects in speaking, adaptation for the subset of 20 narrative elements selected for the current study showed the same pattern as that for the entire narration (Galati & Brennan, 2010): for the 20 elements, speakers used fewer words in the retelling to the same addressee ( $A1_{old}$  vs.  $A2_{new}$ :  $F_1(1, 18) = 13.49, p < .01$ ;  $F_2(1, 19) = 7.14, p < .05$ ), but not in the retelling to a new addressee ( $A1_{new}$  vs.  $A2_{new}$ :  $F_1(1, 18) = 1.87, p = .19$ ,  $F_2(1, 19) = 2.19, p = .16$ ). However, for those 20 elements, dividing the number of

gestures per narrative element by the number of words produced for that element made the *for-the-addressee* effect for representational gestures no longer reliable ( $F_1(1, 18) = 1.76, p = .20$ ;  $F_2(1, 19) = 1.72, p = .20$ ). This suggests that adaptive processes attenuate both speech and gesture in parallel. Moreover, even though the number of representational gestures per narrative element did not reliably differ in the tellings to  $A1_{new}$  and  $A2_{new}$  (reported earlier), this comparison, when normalised for words, becomes marginally reliable by subjects ( $F_1(1, 18) = 3.94, p = .06$ ,  $F_2(1, 19) = 1.77, p = .20$ ). This confirms that it is more appropriate to compare adaptation in the number of gestures per unit of semantic content than per word, especially since here we ensured that these most frequently mentioned narrative elements were realised in all three retellings.

Dividing the number of gestures by the number of words can obfuscate any partner-specific (or other) attenuation in gesture<sup>4</sup> because it assumes that both speech and gesture encode meaning compositionally and sequentially, whereas gesture in fact encodes meaning globally, with different components (such as handshape and movement) giving rise to a single gesture (McNeill, 1992). For these salient elements, retellings to the same addressee were on average less than a word shorter than retellings to new addressees (11.80 words per narrative element to  $A1_{old}$  vs. 12.28 to  $A1_{new}$  and 12.77 to  $A2_{new}$ ). Although the reduction in the number of words in the retelling to A1 was a reliable, it is so small that it is unlikely that the parallel reduction in the number of representational gestures per narrative element is due to a significant loss in opportunities to gesture (given that representational gestures typically unfold over longer stretches of speech).

#### Adaptation in gesture size

Speakers' use of gesture space was constrained by partner-specific knowledge: speakers adapted the size

Table 2. Partner-specific contrasts for the frequencies of different types of gestures. Significant effects are highlighted.

|                  | <i>For-the-speaker</i> effect   |  | <i>For-the-addressee</i> effect   |  |
|------------------|---|--|---|--|
|                  | Speaker <sub>new</sub> -Addressee <sub>new</sub> vs. Speaker <sub>old</sub> -Addressee <sub>new</sub> |  | Speaker <sub>old</sub> -Addressee <sub>old</sub> vs. Speaker <sub>old</sub> -Addressee <sub>new</sub> |  |
| Representational | $F_1(1, 18) = 0.07, ns$   |  | <b><math>F_1(1, 18) = 16.76, p &lt; .01</math></b>  |  |
|                  | $F_2(1, 19) = 0.31, ns$   |  | $F_2(1, 19) = 0.93, ns$   |  |
| Metanarrative    | $F_1(1, 18) = 0.00, ns$   |  | $F_1(1, 18) = 1.53, p = .23$  |  |
|                  | $F_2(1, 19) = 0.05, ns$   |  | $F_2(1, 19) = 2.23, p = .15$  |  |
| Beat             | $F_1(1, 18) = 2.10, p = .17$  |  | $F_1(1, 18) = 0.02, ns$   |  |
|                  | $F_2(1, 19) = 2.31, p = .15$  |  | $F_2(1, 19) = 0.08, ns$   |  |
| Combination      | $F_1(1, 18) = 0.10, ns$   |  | $F_1(1, 18) = 0.27, ns$   |  |
|                  | $F_2(1, 19) = 1.39, ns$   |  | $F_2(1, 19) = 1.30, p = .27$  |  |
| Total            | $F_1(1, 18) = 0.16, ns$   |  | $F_1(1, 18) = 3.26, p = .09$  |  |
|                  | $F_2(1, 19) = 0.31, ns$   |  | $F_2(1, 19) = 2.03, p = .17$  |  |

of their gestures according to whether addressees had heard the story before or not. Gesture size was rated 3.20 ( $SD = 1.18$ ) for the first telling, 2.79 ( $SD = 1.28$ ) for the retelling to the same addressee and 3.06 ( $SD = 1.23$ ) for the retelling to a new addressee. Figure 1 illustrates the gesture size ratings for addressee orders *A1-A1-A2* and *A1-A2-A1*. For both addressee orders, speakers attenuated their gesture size significantly more when retelling the story to the same addressee than to a new addressee (see Table 3).

In addition to attenuating gestures for-the-addressee, there was also a clear for-the-speaker effect, with speakers attenuating gesture size based on their own experience with the story: gestures produced in the telling to A2 used less space compared to gestures produced in the first retelling to A1 (Table 3). In other words, after the first telling speakers attenuated the size of their gestures, but less so if the retelling was to a new addressee than to an old addressee.

The interaction between addressee’s knowledge status and addressee order was reliable by items but not by subjects,  $F_1(2, 36) = 1.94, p = .16$ ;  $F_2(2, 38) = 5.01, p < .05$ . When speakers retold the stories to a

knowledgeable addressee ( $A1_{old}$ ), they attenuated the size of their gestures more in narrative order *A1-A2-A1* than in order *A1-A1-A2*. As Figure 1 shows, gesture size for the retelling to the knowledgeable addressee A1, when this immediately followed the first telling, was only slightly smaller than the retelling to the naive addressee A2; this numerical difference was not reliable,  $F_1(1, 9) = 1.47, ns$ ;  $F_2(1, 19) = 1.10, ns$ . However, when the third retelling was to the knowledgeable A1, gesture size was significantly smaller compared to the retelling to the naive A2,  $F_1(1, 9) = 10.94, p < .01$ ;  $F_2(1, 19) = 42.72, p < .001$ . Gestures directed at A2 were of similar size whether they appeared the second or third time the speaker told the story,  $F_1(1, 18) = 0.004, ns$ ;  $F_2(1, 19) = 0.14, ns$ . And for retellings to A1, the difference in gesture size across orders *A1-A1-A2* and *A1-A2-A1* was reliable only by items,  $F_1(1, 18) = 0.49, ns$ ;  $F_2(1, 19) = 7.19, p < .05$ . In other words, the interaction of addressee’s knowledge and addressee order makes sense in light of speakers adapting gesture size both for their addressees and for themselves. In order *A1-A2-A1*, attenuation across the second and third retellings was reliable and consistent with both an effect of partner-specific adaptation and an effect of practice. But in order *A1-A1-A2*, the two factors worked against each other, resulting in less of a dip in the black bar than the grey bar for  $A1_{old}$  in Figure 1. Another possibility is that switching back and forth between addressees may have made the identity of the addressee more salient than switching addressees only once, leading to the numerically greater attenuation in gesture size for retellings to A1 in order *A1-A2-A1*.

**Adaptation in gesture precision**

Speakers adapted the iconic precision of their gestures in similar ways as they did gesture size, with both for-the-addressee<sup>5</sup> and for-the-speaker effects (see Table 3). The mean rating for the iconic precision of gestures was 3.24 ( $SD = 1.17$ ) for the first telling, 2.80 ( $SD = 1.26$ ) for the retelling to the same addressee and 3.05 ( $SD = 1.18$ ) for the retelling to a new addressee. Figure 2 illustrates the mean ratings for the amount of iconic precision with which speakers gestured in the addressee order *A1-A1-A2* and *A1-A2-A1*. For both addressee

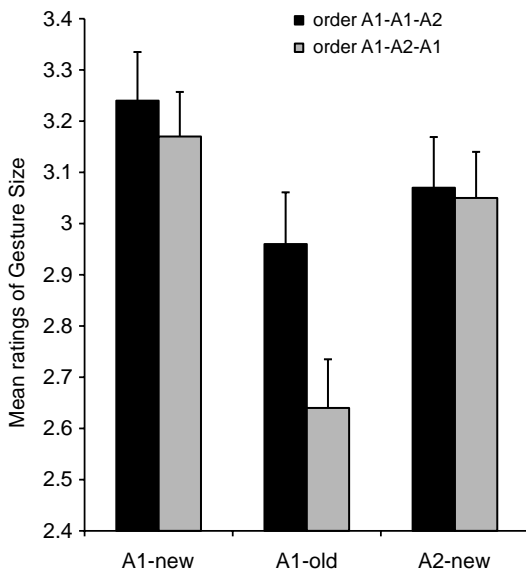


Figure 1. Mean rating for gesture size judgements (ranging from 1 to 7); bars represent standard errors.

Table 3. Partner-specific contrasts for ratings of gestures’ size and iconic precision.

| Contrasts  | Gesture size                   | Iconic precision               |
|--|--------------------------------|--------------------------------|
| <i>For-the-speaker effect:</i>                           |                                |                                |
| Speaker <sub>new</sub> -Addressee <sub>new</sub> vs. new | $F_1(1, 18) = 4.29, p = .05$   | $F_1(1, 18) = 8.98, p < .01$   |
| Speaker <sub>old</sub> -Addressee                        | $F_2(1, 19) = 21.54, p < .001$ | $F_2(1, 19) = 18.51, p < .001$ |
| <i>For-the-addressee effect:</i>                         |                                |                                |
| Speaker <sub>old</sub> -Addressee <sub>old</sub> vs.     | $F_1(1, 18) = 10.54, p < .01$  | $F_1(1, 18) = 10.90, p < .01$  |
| Speaker <sub>old</sub> -Addressee <sub>new</sub>         | $F_2(1, 19) = 33.44, p < .001$ | $F_2(1, 19) = 29.86, p < .001$ |

orders, gestures in retellings to the same addressee were less precise than to a new addressee. As with gesture size, there was also a for-the-speaker effect: speakers were less precise when gesturing in a retelling to a new addressee than in their first telling to a new addressee.

For gesture precision, there was a reliable interaction between the addressee's knowledge status and addressee order,  $F_1(2, 36) = 3.38, p < .05$ ;  $F_2(2, 38) = 4.49, p < .05$ . When speakers retold stories to the same addressee, they attenuated their gestures in terms of iconic precision more so in the narrative order *A1-A2-A1* (in the third retelling) than in the narrative order *A1-A1-A2* (in the second retelling). This is not surprising, as in order *A1-A2-A1*, the third telling may have been shaped in the same direction by *both* for-the-speaker and for-the-addressee attenuation (or possibly by the increased salience of addressee's identity in order *A1-A2-A1* due to the successive switching). As with gesture size, the for-the-addressee effect (the difference between iconic precision in the retellings to A1 and A2) was therefore more pronounced in this order than in order *A1-A1-A2*, where the old partner is in the second telling. As Figure 2 shows, relative to the retelling to A2, iconic precision in the retelling to A1 was more attenuated for order *A1-A2-A1* than in order *A1-A1-A2*. In other words, when the retelling to A1 was the third telling, the gestures were judged to be significantly less precise relative to the retelling of A2,  $F_1(1, 9) = 10.27, p < .05$ ;  $F_2(1, 19) = 26.80, p < .001$ . On the other hand, when the retelling to A1 immediately followed the first telling, the speakers' gestures to A1 were judged to be numerically slightly less precise than

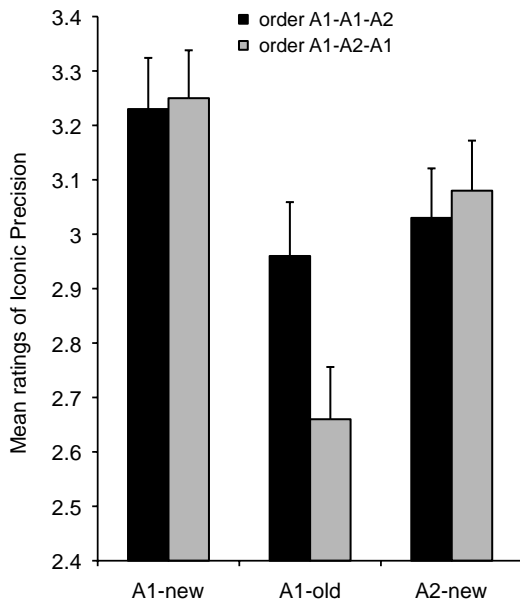


Figure 2. Mean rating for iconic precision judgements (ranging from 1 to 7); bars represent standard errors.

the retelling to A2, but for this order the for-the-addressee difference was not reliable,  $F_1(1, 9) = 1.42, ns$ ;  $F_2(1, 19) = 0.20, ns$ . Gestures directed at A2 were of comparable iconic precision whether they occurred in the second or third time the speaker told the story,  $F_1(1, 18) = 0.08, ns$ ;  $F_2(1, 19) = 0.18, ns$ . For retellings to A1, the difference in the gestures' iconic precision seen across orders *A1-A1-A2* and *A1-A2-A1* was reliable only by items,  $F_1(1, 18) = .53, ns$ ;  $F_2(1, 19) = 5.47, p < .05$ . That is, gestures to A2 did not differ in their iconic precision whether they were produced in the second or third telling, and neither did gestures to A1. The interaction of the addressee's knowledge and the addressee order is consistent with speakers adapting their gestures' precision both for their addressees and for themselves. Although gestures in retellings to A2 were more precise than to A1 in order *A1-A2-A1*, they were not reliably so in the other order, *A1-A1-A2*. The attenuation in precision observed across the three retellings in order *A1-A2-A1* appears to have been cancelled out by having a new partner on the third telling in order *A1-A1-A2*.

We also examined the relation between the iconic precision of gestures and their size. These qualitative dimensions of gestures were highly correlated: for the first telling, Pearson's  $r = .60, p < .001$ ; for the retelling to A1, Pearson's  $r = .68, p < .001$ ; and for the retelling to A2, Pearson's  $r = .65, p < .001$ . These correlations were not driven simply by narrative elements for which no gestures were produced (where both iconic precision and size had sizes of zero in the data-set); they remained reliable even when these instances were filtered out (first telling, Pearson's  $r = .55, p < .001$ ; retelling to A1, Pearson's  $r = .60, p < .001$ ; and retelling to A2, Pearson's  $r = .61, p < .001$ ).

### Addressees' feedback

Speakers accrue information about their addressee's informational needs not only from prior experience but also from verbal and nonverbal cues from the addressee as the dialogue unfolds (see Brennan, Galati, & Kuhlen, 2010, for discussion). Indeed, in a study in which we dissociated the speakers' prior expectations about the addressees' attentiveness from their addressees' feedback we found that speakers used both sources of information in a highly interactive manner to adapt their gestures (Kuhlen et al., 2012; see also Kuhlen & Brennan, 2010). To explore the possibility that the addressees' feedback shaped speakers' gesturing in the current study, we coded in the transcript of each narration the number of turns, or instances in which addressees made comments, asked clarification questions, provided audible feedback responses (e.g.,

*mm-hmm, uh-huh, yeah*), made expressive exclamations (*oh, ah*), or laughed.

Consistent with addressees' knowledge status, A1 provided overt feedback in an average of 7.15 turns ( $SD = 7.34$ ) the first time they heard the story ( $A1_{new}$ ), and 3.05 turns ( $SD = 5.00$ ) the second time they heard the story ( $A1_{old}$ ),  $F(1, 18) = 4.74, p < .05$ . Addressees who heard the story only once ( $A2_{new}$ ) provided feedback in an average of 3.95 turns ( $SD = 4.57$ ); although the numerically lower amount of feedback by  $A2_{new}$  compared to  $A1_{new}$  was unexpected, this difference was not reliable,  $F(1, 18) = 2.27, p = .15$ . It is not clear whether  $A2$ 's (numerically) attenuated feedback was related to the speaker's gestures (which were somewhat attenuated in size and precision but not in number relative to the first telling) or possibly to  $A2$ 's awareness that the speaker had told the story to A1 already. In any event, feedback from  $A1_{old}$  and  $A2_{new}$  did not differ ( $F(1, 18) = .39, ns$ ), and so feedback could not have been solely responsible for the speaker's reliably larger, more frequent and more precise gestures to  $A2_{new}$ .

## Discussion

Our findings show that both speech and gesture production are constrained by speakers' awareness of their addressees' informational needs, and that the two modalities are coordinated during production. The partner-specific adaptation we observed extends from early phases in gesture planning (influencing whether to gesture and what information to encode), all the way through motoric execution (determining qualitative aspects of a gesture). Speakers tended to produce marginally fewer representational gestures in retellings to old addressees than to new addressees. And in executing these gestures, they were significantly more likely to attenuate their relative size and their iconic precision to old addressees than to new addressees. These findings for gesture are consistent with our previous findings from the same corpus for speech: addressee knowledge shapes both utterance planning and articulation. In Galati and Brennan (2010), the number of events realised, words, amount of detail, and intelligibility of lexically identical expressions were all attenuated in speech directed to old addressees relative to speech directed to new addressees.

In addition to finding strong evidence of partner-specific adaptation in the current study, we also found attenuation for-the-speaker in gestures' motoric execution. Relative to the first telling, speakers attenuated the size and iconic precision of gestures in retellings to new addressees. In other words, after the first telling, speakers attenuated both gesture size and iconic

precision, but less so if the retelling was to a new addressee than to an old addressee. Attenuation in the two addressee orders further highlights that motoric execution of gestures appears to be constrained by both for-the-speaker and for-the-addressee information.

The interaction of these effects with partner order (illustrated by the differences in height between the middle pair of bars in both Figures 1 and 2) was likely due to the fact that for-the-speaker and for-the-addressee influences worked in concert for the retelling to A1 and in opposition for the retelling to A2. For the retelling to A1, both factors work in concert since attenuation is compatible both with the story being accessible to speakers and with the addressee's knowledge status. With both factors contributing to attenuating gestures in the retelling to A1, gesture size and iconic precision scores were lower in order  $A1-A2-A1$  (where the story is maximally accessible to speakers in the third retelling) than  $A1-A1-A2$ . On the other hand, for the retelling to A2, the two factors work in opposition: attenuation is driven by the story being accessible to the speaker but is inappropriate given the addressee's knowledge status. We found that  $A2$ 's knowledge status curbed further attenuation arising from the order of the retelling: although relative to the first telling gestures to A2 were attenuated, they were not attenuated more in a third retelling than in a second retelling (the difference between orders  $A1-A1-A2$  and  $A1-A2-A1$  was not reliable). It is also possible that switching partners twice during the experiment (order  $A1-A2-A1$ ) made the identity of the partner more salient than switching partners only once (order  $A1-A1-A2$ ), such that the attenuation to A1 in order  $A1-A2-A1$  was greater than in order  $A1-A1-A2$ .

The evidence from the current study, taken together with that from Galati and Brennan (2010) using the same experimental corpus, suggests that speech and gesture production are coordinated: in both modalities, speakers adapted their behaviour for-the-addressee, during both planning and articulation. In addition, the current study found adaptation in the motoric execution of gesture for-the-speaker, even though there was no reliable for-the-speaker adaptation in the intelligibility of lexically identical expressions in Galati and Brennan (2010). Why should this be so? One possibility is that gesturing is more sensitive than speaking to motor practice effects (another way of looking at for-the-speaker attenuation). For instance, the delay in between retellings of a particular narrative event may have been long enough to swamp any practice effects for repeated words but not for repeated gestures. Another possibility is that the presence of a for-the-speaker effect in the current study and absence of one in Galati and Brennan (2010) can be attributed to semiotic differences between speech and gesture.

Gestures, unlike spoken language, are not compositional, and the mappings they represent between form and meaning are not arbitrary, so their communicative potential may be less threatened by attenuation: speakers may be freer to attenuate their gestures without significant cost to addressees' comprehension. If that is the case, gestures may be more susceptible than speech to for-the-speaker effects. Despite the differences we found in for-the-speaker attenuation of speech vs. gestures, the two modalities must interface during their articulation to account for their fine-grained temporal coordination (e.g., Mayberry & Jaques, 2000; McClave, 1994; Seyfeddinipur, 2006).

Our findings go beyond those of Jacobs and Garnham (2007), who used a similar design but focused on partner-specific adjustments only in quantitative terms (gesture rate). Jacobs and Garnham's (2007) findings converge with ours in that they demonstrate that gesture frequency is constrained by addressees' knowledge. In their study, gesture rate across retellings to different addressees did not differ significantly, whereas it decreased with each retelling to the same addressee. In our study, measuring within-speaker adaptation in surface form afforded a more nuanced understanding of the speaker-specific and addressee-specific factors shaping not only gesture planning, but also the gestures' motoric execution.

Unconfounding for-the-speaker and for-the-addressee factors and examining them simultaneously in the same experimental design goes beyond previous attempts to model gesture production that focused on a single cognitive or communicative constraint. The adaptation we report here cannot be accounted for by proposals that gestures are driven egocentrically, for instance, only for facilitating the speakers' lexical retrieval (Krauss, Dushay, Chen, & Rauscher 1995; Krauss, Morrel-Samuels, & Colasante, 1991; Krauss et al., 2000). The Lexical Gestures framework of Krauss and colleagues would predict that the rate of gesturing would simply decrease across the three tellings (either gradually or all at once), since with each retelling words are more accessible to the speaker. Instead, we found enhanced gesture rates with addressees who were hearing stories for the first time, establishing that the effect is shaped by addressees' needs.

Our findings are consistent with mounting evidence that gestures are not produced exclusively for either speaker or addressee (see Bavelas & Chovil, 2000, for discussion). These findings are also consistent with Hostetter and Alibali's (2008, 2010) Gesture as Simulated Action framework, which posits that several factors can constrain gesture production simultaneously. The Gesture as Simulated Action framework enables multiple constraints, including communicative ones, and can adapt the threshold that needs to be

exceeded for a gesture to be generated. Nonetheless, it does not currently specify whether adjustments in the gesture threshold can affect qualitative aspects of a gesture, although such refinements could be made to account for the kind of motoric adaptation we report here. For example, to the extent that increased common ground between partners raises the threshold for producing a gesture, when speakers retell (upon mentally simulating) events to old addressees, they should be not only less likely to produce a gesture representing an event (relative to new addressees), but also less likely to motorically realise the gesture as vividly or as precisely.

Our findings also fit the assumptions of a constraint-based model of speech and co-speech gesture production in which different sources of information – including that from common ground, discourse context and within-sentence structural and lexical biases – are weighted depending on their salience and relevance to the task and are integrated probabilistically and in parallel to shape production (e.g., Jurafsky, 1996; MacDonald, 1994; Tanenhaus & Trueswell, 1995). Critically, in constraint-based models there is no need to posit information encapsulation or architectural barriers in the planning process. Instead, adaptation across different grains of processing reflects probabilistic constraints on information processing. This means that tailoring utterances to an addressee's needs can occur when information about those needs is available early enough (see Brennan & Hanna, 2009; Brown-Schmidt & Hanna, 2011; Hanna, Tanenhaus, & Trueswell, 2003; Hanna & Tanenhaus, 2004). That both “inferential” processes (like message planning) and “automatic” ones (like spoken articulation and the motoric execution of gestures) are adapted to addressees' knowledge runs contrary to modular proposals that consider “automatic” processes to be necessarily encapsulated from and unaffected by partner-specific information (Bard & Aylett, 2001; Bard et al., 2000).

To adapt to addressees, speakers need not elaborately represent their addressee's informational needs. Instead, they may represent addressees' needs as simple, pre-computed constraints (often captured by binary alternatives, e.g., *my addressee has heard this story before*, or not) that combine probabilistically (Brennan & Hanna, 2009; Galati & Brennan, 2010). Such a constraint-based model of partner-specific adaptation is supported by the many demonstrations for adjustments in gesture based on whether the addressee can see the speaker (e.g., Alibali et al., 2001; Bangerter, 2004; Bavelas et al., 1992, 2008; Cohen & Harrison, 1973), where their addressee is seated (Özyürek, 2000, 2002), and whether the addressee can see or has seen what the speaker is describing (Gerwing & Bavelas, 2004; Holler & Stevens, 2007; Holler & Wilkin, 2009; Jacobs & Garnham, 2007). As long as such distinctions about a partner's knowledge

are salient or previously computed – as global variables available continuously in processing – they could be integrated in parallel with other constraints to probabilistically impact even “automatic” processes like spoken articulation or the motoric execution of gestures. We consider such constraints to be flexible; over the course of a dialogue, local cues such as a partner’s feedback could also reinforce or update a constraint-based representation of the partner’s knowledge or needs (Brennan et al., 2010; Kuhlén et al., 2012).

Finally, we caution that making claims about functional differences between gesture types would be premature based solely on our findings. Like other studies that have used a cartoon elicitation task (e.g., Alibali et al., 2001), we found clear partner-specific effects for representational gestures (the most frequent type of gesture in narrations of motion events). Partner-specific constraints may impact other types of gestures as well, with different effects to the extent that those gestures serve different functions. The partner-specific pattern we had hypothesised for metanarrative gestures (i.e., an increase in the retellings to the same addressee) might emerge in a different elicitation task.

### Conclusion

This study distinguished speaker-specific from addressee-specific influences on gesturing in a narrative task, finding strong evidence that gesturing is shaped by both. There was no evidence that speakers’ perspectives take priority over addressees’ in the speakers’ narrations. This finding contrasts with other proposals that speakers default to egocentric behaviour, adjusting to any distinct needs of addressees only later, as a repair (e.g., Horton & Keysar, 1996; Pickering & Garrod, 2004). This joint influence of for-the-speaker and for-the-addressee factors is consistent with models that consider gesture to be shaped by multiple constraints (Hostetter & Alibali, 2008, 2010). In conjunction with previous findings on partner-specific adaptation in speaking (Galati & Brennan, 2010), the current findings clarify how multiple factors can guide spontaneous utterance planning and execution in both speech and gesture. The strong for-the-addressee effects we have found across the two modalities suggest that speech and gesture are closely coordinated from the earlier phases of planning, all the way through articulation. As long as communicative constraints are simple enough (e.g., *my addressee has heard this story before, or not*), salient and relevant to the task at hand, they can be integrated with other sources of information to affect the planning and execution of utterances with little or no discernable cost to the speaker.

### Acknowledgements

This material is based upon work supported by NSF under Grants IIS-0527585 and ITR-0325188. We thank our colleagues from the Adaptive Spoken Dialogue Project, the Shared Gaze Project, the Dialogue Matters Network (funded by Leverhulme Trust), and the Gesture Focus Group for many helpful discussions, especially Arthur Samuel and Anna Kuhlén. We are grateful to Randy Stein and Marwa Abdalla for their assistance with coding.

### Notes

1. One triad was excluded because the speaker had a broken arm, which impeded his gestural production. A second triad was excluded because the speaker misunderstood the task and provided a nearly frame-by-frame report of the cartoon’s progression instead of narrating a story; the speaker ran out of time and did not complete the task. The third triad was excluded because one of the addressees had his eyes closed throughout the speaker’s narrations (presumably trying to visualise the cartoon events and not falling asleep!). According to Bavelas et al. (1992) the rate of gesturing, especially for interactive gestures, is significantly affected by visual availability, and as such the speaker’s gestural production when narrating to this listener might have been affected.
2. Additional cartoons were also narrated, one with Tweety and Sylvester and one with Bugs Bunny and Yosemite Sam. The third cartoon was dropped from the task after the first six triads, because narrating a total of nine stories (three for each cartoon) was often too tiring for the speakers. The narrations of the second cartoon were not analysed.
3. The stroke is the expressive and dynamic part of the gesture, bearing its semantic content, and it is optionally preceded by a preparation phase, during which the hands move from rest towards the space where the stroke is executed, and a retraction phase, during which the hands return to rest (McNeill, 1992).
4. A study that normalised the number of gestures by words to assess partner-specific adaptation yielded the paradoxical result that gesture frequency increased when speakers and addressees had common ground (had both watched the story told by the speaker) compared to when they did not, even though speakers produced fewer words and fewer gestures when telling the story to an addressee who had seen the story than to one who had not (Holler & Wilkin, 2009).
5. Since coders assigned a single score to the size and iconic precision of all gestures for a given narrative element, the effects we observed could conceivably be due to speakers producing more beats and metanarrative gestures rather than attenuating the precision of their representational gestures. However, attenuation in these qualitative dimensions appears to be driven by representational gestures, since the frequency of beats and metanarrative gestures did not differ across retellings. Although the frequency of beats and of metanarrative gestures increased numerically after the first telling (see Table 1), this increase was not reliable. In particular, the for-the-addressee effect was not significant for either metanarra-



tive gestures ( $F_1(1, 18) = 1.53, p = .23$ ;  $F_2(1, 19) = 2.23, p = .15$ ) or beat gestures ( $F_1(1, 18) = 0.02, ns$ ;  $F_2(1, 19) = 0.08, ns$ ). Given the preponderance of representational gestures in our corpus and the lack of adaptation in beat or metanarrative gestures across retellings, the observed effects for size and iconic precision seem to be primarily due to adaptation in the motoric execution of representational gestures.

## References

- Alibali, M. W., & DiRusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development, 14*(1), 37–56. doi:10.1016/S0885-2014(99)80017-3
- 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*(2), 169–188. doi:10.1006/jmla.2000.2752
- Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*, 593–613. doi:10.1080/016909600750040571
- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*, 415–419. doi:10.1111/j.0956-7976.2004.00694.x
- 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 & Language, 42*(1), 1–22. doi:10.1006/jmla.1999.2667
- Bard, E. G., & Aylett, M. P. (2001). *Referential form, word duration, and modeling the listener in spoken dialogue*. Paper presented at the twenty-third annual conference of the Cognitive Science Society, Edinburgh, Scotland.
- Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology, 19*(2), 163–194. doi:10.1177/0261927X00019002001
- Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin, 21*, 394–405. doi:10.1177/0146167295214010
- Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469–489. doi:10.1080/01638539209544823
- Bavelas, J. B., 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. doi:10.1016/j.jml.2007.02.004
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 301–345). Elsevier: Academic Press.
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science (Special Issue on Joint Action), 1*, 274–291. doi:10.1111/j.1756-8765.2009.01019.x
- Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective taking in language processing. *Dialog and Discourse, 2*, 11–33.
- Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 747–756. doi:10.1037/0278-7393.33.4.747
- Cassell, J., & McNeill, D. (1991). Gesture and the poetics of prose. *Poetics Today, 12*, 375–404. doi:10.2307/1772644
- Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*, 580–601. doi:10.1006/jesp.1994.1027
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Cohen, A. (1977). The communicative function of hand gestures. *Journal of Communication, 27*, 54–63. doi:10.1111/j.1460-2466.1977.tb01856.x
- 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. doi:10.1037/h0035792
- Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49–80.
- de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior, 13*, 247–268. doi:10.1007/BF00990296
- de Ruiter, J. P. A. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 284–311). Cambridge: Cambridge University Press.
- Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*(1), 35–51. doi:10.1016/j.jml.2009.09.002
- Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture, 4*(2), 157–195. doi:10.1075/gest.4.2.04ger
- Goldin-Meadow, S., Nusbaum, H., Kelly, S., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516–522. doi:10.1111/1467-9280.00395
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489–1522. doi:10.1016/S0378-2166(99)00096-X
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43–61. doi:10.1016/S0749-596X(03)00022-6
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*(1), 105–115. doi:10.1207/s15516709cog2801\_5
- 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. doi:10.1177/0261927X06296428
- 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. doi:10.1080/01690960802095545
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91–117. doi:10.1016/0010-0277(96)81418-1
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review, 15*, 495–514. doi:10.3758/PBR.15.3.495
- Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language, 63*, 245–257. doi:10.1016/j.jml.2010.04.003
- Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291–303. doi:10.1016/j.jml.2006.07.011
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20*(2), 137–194. doi:10.1207/s15516709cog2002\_1
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language, 40*, 577–592. doi:10.1006/jmla.1999.2634
- Kendon, A. (1994). Do gestures communicate?: A review. *Research on Language and Social Interaction, 27*(3), 175–200. doi:10.1207/s15327973rlsi2703\_2
- Keysar, B. (1997). Unconfounding common ground. *Discourse Processes, 24*, 253–270. doi:10.1080/01638539709545015

- 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(1), 16–32. doi:10.1016/S0749-596X(02)00505-3
- Krauss, R. M., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.
- Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, 31, 533–552. doi:10.1006/jesp.1995.1024
- Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, 61, 743–754. doi:10.1037/0022-3514.61.5.743
- Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, 47, 567–587. doi:10.1080/01638530903441339
- Kuhlen, A. K., Galati, A., & Brennan, S. E. (2012). Gesturing integrates top-down and bottom-up information: Joint effects of speakers' expectations and addressees' feedback. *Language and Cognition*, 4, 17–41. doi:10.1515/langcog-2012-0002
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levy, E., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse Processes*, 15, 277–301. doi:10.1080/01638539209544813
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9(2), 157–201. doi:10.1080/01690969408402115
- Mayberry, R. I., & Jaques, J. (2000). Gesture production during stuttered speech: Insights into the nature of gesture-speech integration. In D. McNeill (Ed.), *Language and gesture* (pp. 199–213). Cambridge: Cambridge University Press.
- McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, 23(1), 45–66. doi:10.1007/BF02143175
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92, 350–371. doi:10.1037/0033-295X.92.3.350
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, D. (2000). Catchments and contexts: Non-modular factors in speech and gesture production. In D. McNeill (Ed.), *Language and gesture* (pp. 312–328). Cambridge: Cambridge University 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.
- McNeill, D., & Levy, E. (1993). Cohesion and gesture. *Discourse Processes*, 16, 363–386. doi:10.1080/01638539309544845
- Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, 22, 473–500. doi:10.1080/01690960600696916
- Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology*, 117, 411–424. doi:10.2307/4149008
- Müller, M. (1998). Iconicity and gesture. In S. Santi, et al. (Eds.), *Oralità et Gestualità: Comunicazione Multimodale, Interazione* (pp. 321–328). Montréal: L'Harmattan.
- Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (Ed.), *Language and gesture. Series in language, culture and cognition* (pp. 64–82). Cambridge: Cambridge University Press.
- Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, 46, 688–704. doi:10.1006/jmla.2001.2826
- Payne, D. G. (1987). Hypermnnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, 101(1), 5–27. doi:10.1037/0033-2909.101.1.5
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 167–226.
- Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7, 226–231. doi:10.1111/j.1467-9280.1996.tb00364.x
- Rimé, B., Schiatura, L., Hupet, M., & Ghysselelinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, 8, 311–325. doi:10.1007/BF00991870
- Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture. MPI Series in Psycholinguistics*, 39. Nijmegen: Max Planck Institute of Psycholinguistics.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception*, vol. 11 (pp. 217–262). San Diego, CA: Academic Press.
- Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28, 879–896. doi:10.1002/(SICI)1099-0992(199811)28:6<879::AID-EJSP901>3.0.CO;2-W

## Appendix 1

### The 20 narrative elements for the Road Runner cartoon included for gesture coding

- Road Runner is being chased by Coyote/Road Runner and Coyote are running (in the desert).
- Coyote is holding a fork and a knife and has a napkin around his neck.
- Coyote lunges forward with fork and knife, trying to stab Road Runner.
- Road Runner says “Beep Beep”, takes off; Coyote misses Road Runner. (RR is too fast.)
- Coyote attaches boxing glove (on spring), (on metal band), on rock.
- The boxing glove initially does not move. The rock is propelled backwards.
- Coyote is pushed backwards by the rock and is smashed against the side of the cliff.
- The boxing glove is propelled backwards and punches Coyote in the face.
- Coyote steps on tightrope, holding the anvil.
- The tightrope stretches all the way to the ground, bringing Coyote all the way down to the ground.
- Road Runner comes and stops in front of Coyote.
- Coyote drops anvil on the ground (to go run after Road Runner).
- Coyote is propelled up in the air by the tightrope.
- Coyote pulls the string on his backpack to release the parachute.
- Coyote grabs some aspirin pills and pops them into his mouth.
- Coyote falls and hits the ground.
- Coyote is seen chasing after Road Runner. Road Runner runs into the “Old Cactus Mine”.
- The zigzag path splits into an upper path and a straight path. The green light (Road Runner) takes the upper path while the red light (Coyote) takes the straight path below/Road Runner loses the Coyote.
- Coyote lights up a match.
- The cactuses fly up into the air and fall back down spelling “YIPE!”.