Gesturing integrates top-down and bottom-up information: Joint effects of speakers' expectations and addressees' feedback

ANNA K. KUHLEN[+], ALEXIA GALATI[#] AND SUSAN E. BRENNAN[o]*

[+]Humboldt University of Berlin, Germany
[#]University of Cyprus, Cyprus
[o]Stony Brook University, USA

Abstract

Speakers adapt their speech based on both prior expectations and incoming cues about their addressees' informational needs (Kuhlen and Brennan 2010). Here, we investigate whether top-down information, such as speakers' expectations about addressees' attentiveness, and bottom-up cues, such as addressees' feedback during conversation, also influence speakers' gestures. In 39 dyads, addressees were either attentive when speakers told a joke or else distracted by a second task, while speakers expected addressees to be either attentive or distracted. Independently of adjustments in speech, both speakers' expectations and addressees' feedback shaped quantitative and qualitative aspects of gesturing. Speakers gestured more frequently when their prior expectations matched addressees' actual behavior. Moreover, speakers with attentive addressees gestured more in the periphery of gesture space when they expected addressees to be attentive. These systematic adjustments in gesturing suggest that speakers flexibly adapt to their addressees by integrating bottom-up cues available during the interaction in light of attributions made from top-down expectations. That these sources of information lead to adjustments patterning similarly in speech and gesture informs theoretical frameworks of how different modalities are deployed and coordinated in dialogue.

Keywords
speech-accompanying gestures, communication, audience design, spontaneous spoken dialogue, feedback cues, distracted addressees

* Correspondence address: Anna Kuhlen, Berlin School of Mind and Brain, Humboldt University of Berlin, 10117 Berlin, Germany. E-mail: anna.kuhlen@bccn-berlin.de. We thank Elizabeth Cohen, Krystan Lenhard, Marina Khan and Sandy Weng for their assistance with data collection and coding, Stony Brook's Gesture Focus Group for discussions, and anonymous reviewers for helpful comments. This material is based upon work supported by NSF under Grants IIS-0527585 and ITR-0325188.

Language and Cognition 4–1 (2012), 17–41
DOI 10.1515/langcog-2012-0002
1866–9808/12/0004–0017
© Walter de Gruyter

1. Introduction

Conversation involves coordination, not only *intra*personally, but also *inter*personally. Intrapersonal cognitive processes such as planning, articulating, perceiving, and interpreting utterances are coordinated within the minds of individuals, and across multiple modalities. *Inter*personally, speakers and addressees adapt these processes to be contingent on one another (e.g. Brennan et al. 2010; Clark and Wilkes-Gibbs 1986; Kraut et al. 1982; Kuhlen and Brennan 2010; Shockley et al. 2009). We propose that such adaptation cannot be accounted for through a simple process of contingent responding during alternating turn-taking; even though only one speaker tends to speak at a time, all partners display their degree of engagement continuously: As a speaker presents an utterance, she monitors her addressee for evidence of understanding and uptake, and as her addressee listens, he displays such evidence both verbally and nonverbally (Brennan 2005; Clark and Brennan 1991; Yngve 1970). Broadly speaking, our investigation focuses on the interpersonal coupling of communicative processes in a spontaneous face-to-face narration task.

Both verbal and nonverbal signals are smoothly integrated into utterances, and yet the two kinds of signals seem quite different. Gesturing, a nonverbal behavior in which visible bodily actions accompany speech, is not subject to the same linguistic constraints as speaking, and the two modalities have very different semiotic properties; for instance, words are associated with conventional meanings that can be combined compositionally, whereas most gestures are not. Although the two modalities can co-refer, they need not be redundant; each may express information that the other does not (McNeill 2000; Alibali et al. 2000; Church and Goldin-Meadow 1986). Yet co-speech gestures are temporally synchronized and co-expressive with speech, and so have been proposed to emerge from the same underlying set of representations (McNeill and Duncan 2000; McNeill 1992). Parallel adjustments in both speech and gesture in response to social factors (e.g. Bavelas et al. 2008; Holler and Stevens 2007; Holler and Wilkin 2009) broadly support the proposal that speech and gesture production interact or share some processing resources, but they do not clarify whether adjustments in gesture are a consequence of adjustments in speech. The study that we report here builds on our previous work on speaking (Kuhlen and Brennan 2010) and investigates whether the same cues that influence speaking also influence gesturing. We control for both the content and the amount of accompanying speech, which allows us to determine whether partner-specific adaptations in gesture are driven by the same forces that drive adaptations in speech. This approach informs a theoretical framework about the coordination of speech and gesture and the extent to which the processes underlying these two modalities are yoked, or shaped by similar cues.

As speakers adapt the verbal and nonverbal aspects of their utterances to their addressees, they can draw from: (a) top-down information such as expectations about their addressees, and (b) bottom-up information that unfolds during the interaction, such as feedback that signals addressees' engagement, uptake, understanding, and informational needs as the conversation unfolds. These two sources of information often converge, to the extent that information that becomes available during the interaction commonly confirms expectations or coincides with other information available prior to the interaction. However, they need not always converge. For example, you might expect that your friend will be very interested in a particular bit of news (e.g. gossip), but as you launch into your story you notice (perhaps with some disappointment), that she seems absent-minded, maybe because she is tired, or she has other things on her mind. To what extent will your story be shaped by your expectation about your friend's interest, and to what extent by your friend's exhibited interest? Not teasing these two sources of information apart makes it difficult to tell on what grounds speakers make partner-specific adjustments.

In this paper, we examine how top-down and bottom-up sources of information each contribute to how speakers gesture during spontaneous narrative, and how such factors may influence one another. We begin by reviewing evidence for adaptation, which we find to be driven both top-down and bottom-up. We then present new data analyzing both quantitative and qualitative aspects of gestures, using an experimentally parameterized corpus of face-to-face communication that we have previously investigated only for speech (Kuhlen and Brennan 2010). We consider (1) whether gesture, like speech, is shaped by both bottom-up and top-down factors, and in particular whether gesture production is flexible enough that a bottom-up factor such as addressee feedback can be mediated by a top-down factor such as expectations; (2) whether expectations and feedback affect qualitative changes in gesturing any differently than quantitative changes; and (3) whether or not adaptations in gesturing emerge as artifacts, redundant with adjustments made concurrently in speaking. We then discuss the theoretical implications of our findings for understanding the processes involved in gesture production, as well as for understanding how speech and gesture are integrated into a multimodal utterance.

1.1. *How bottom-up and top-down information shape gesture*

Speakers have been shown to adapt what they say based on top-down information such as their prior knowledge, beliefs, or expectations about the addressee within the conversational situation, as well as based on more bottom-up cues that unfold moment-by-moment during conversational interaction (e.g. Clark and Wilkes-Gibbs 1986; Kraut et al. 1982; Kuhlen and Brennan 2010; Richardson et al. 2009). Here we review factors that have been shown to influence spontaneous gesturing in conversation and classify them as top-down or bottom-up. Most of these factors have been examined in isolation.

Bottom-up cues consist of information perceived moment-by-moment from the physical environment, especially from a conversational partner's behavior. These cues include a partner's verbal and nonverbal feedback responses, such as eye gaze. Speakers can track their addressees' eye gaze to gauge what or whom they are attending to at that moment (e.g. Argyle and Cook 1976; Brennan et al. 2008; Goodwin 1979; Hanna and Brennan 2007; Kendon 1967), and this can influence both subsequent speaking and gesturing. Even infants as young as 12 months adjust their gestures according to their conversational partners' attention, initiating fewer pointing gestures when their conversational partners do not display visual attention (Liszkowski et al. 2008). Eye gaze and other cues present in addressees' feedback are informative about whether they have attended to, understood, or agreed with what speakers have said (e.g. Bavelas et al. 2000).

Top-down cues include the knowledge and expectations that speakers bring into the conversational situation. Based on top-down information, speakers can adapt their gestures in terms of frequency, size, or complexity. In one study, when speakers were told to describe a physical action (playing with a toy) to addressees whom they knew were either familiar with the action or naïve (having either played with or not played with the toy), their gestures were judged to convey more information and be more complex and precise when directed to naïve addressees (Gerwing and Bavelas 2004). In another study, when speakers described pictures that they had previously viewed and discussed with their addressees, they adapted whether and how they gestured: they were less likely to gesture when describing target objects, and when they did gesture, their gestures were smaller (Holler and Stevens 2007). These studies demonstrate that top-down knowledge about addressees affects how speakers gesture.

These two kinds of information, bottom-up and top-down, often corroborate one another. In the studies reviewed thus far, information that became available to speakers during the interaction matched the information they had prior to the interaction. For example, the feedback that addressees provided when hearing a description of a familiar picture or a familiar toy likely corroborated speakers' prior expectations regarding their informational needs. Because top-down and bottom-up factors are so often confounded in interpersonal communication, it is unclear whether adaptation in gesturing is driven mainly by perceptual cues that unfold during the interaction, or by more stable information about a speaker's prior expectations, or by both. And it is especially unclear how bottom-up and top-down information is integrated in cases where incoming evidence does not match prior expectations. If such factors are shown to interact, then adaptation in gesturing can be considered to be flexible, with incoming evidence interpreted differently depending on expectations, or with expectations modified based on incoming evidence.

1.2. *Joint impact of top-down and bottom-up information*

Most of the evidence about how bottom-up cues and top-down beliefs work together in communication focuses on language use rather than gesture. For instance, one study showed that people in dialogue with a remote partner initially framed their utterances based on their expectations of their partner, but then adapted to their partners' actual behavior (Brennan 1991). In this study, people were led to believe that they were typing textual turns to either a human partner or a computer that could interpret natural language. Although in both cases a confederate was producing responses according to a set of rules, people who believed they were interacting with a computer began the dialogue with telegraphic utterances, while those who believed they were interacting with a human partner began the dialogue with longer grammatical sentences. By the last half of the interaction, however, people had adapted to the type of behavior they experienced in the interaction, producing telegraphic turns to a partner who used telegraphic turns, and more complete sentences to a partner who used complete sentences. Thus, top-down expectations about partners' identity guided initial behavior, but later on, interaction was guided more by incoming perceptual cues. This suggests that bottom-up cues can either update top-down expectations (perhaps people revised their expectations about what a language-interpreting computer was capable of) or else cause expectations to be suspended altogether (perhaps people became so engaged in the dialogue that initial models did not matter, or their behavior automatically converged with the partner's regardless of expectations).

Other studies have demonstrated that, in some situations, people ignore incoming evidence that conflicts with their expectations. People who erroneously expect that their partners have the same goals that they do may resist correcting this belief despite strong behavioral evidence to the contrary, attributing the inconsistency instead to deficiencies in their partners' personalities (Russell and Schober 1999; Wilkes-Gibbs 1986). Under some circumstances, a perceived difference between interlocutors' perspectives can lead to closer coordination. For example, in a study by Richardson et al. (2009), interlocutors discussing a previously watched movie clip synchronized their eye gaze more closely (interpreted as a proxy for joint activity) when they believed their partner did not share the same visual context (could not see the screen where the movie clip had played), perhaps compensating for differences in their perspectives.

In some situations top-down information may inform and guide the perception and interpretation of bottom-up cues by supporting attributions for the behavior (see Teufel et al. 2010, for an overview of how processing even very basic perceptual cues can be influenced by top-down knowledge, such as what perceivers believe and expect of their partner). For example, when people believed that their partner was able to see information relevant to a task (looking through transparent goggles), their perceptual system adapted to the dominant gaze direction of their partner, subsequently leading to a bias in judging people's gaze direction. However, when they believed that their partner was not able to see (looking through opaque goggles), they did not develop this bias (Teufel et al. 2009). And when speakers' idiosyncratic pronunciation was consistent with a dialectal difference or with a temporary cause, such as the speaker having a pen in her mouth, listeners did not show perceptual learning of the pronunciation difference, whereas they did show perceptual learning when the variation could be attributed to a stable idiosyncrasy of the speaker (Kraljic, Brennan et al. 2008; Kraljic, Samuel et al. 2008).

Previously, we teased apart effects on speech production of speakers' top-down expectations from addressees' bottom-up feedback cues (Kuhlen and Brennan 2010). Speakers told addressees jokes in form of narratives: half of the addressees listened attentively to the jokes, and the other half were distracted by a secondary task. This secondary task resulted in different bottom-up cues: attentive addressees gave more feedback than distracted addressees. The instructions to the speakers evoked two different top-down expectations: half of the speakers expected to interact with attentive addressees, and the other half expected to interact with distracted addressees. Thus, in two experimental conditions speakers held inaccurate expectations about their conversational partners: some speakers expected addressees to be attentive, but encountered a distracted addressee, and some speakers expected addressees to be distracted, but encountered an attentive addressee. Having speakers' expectations be either congruent or incongruent with the addressees' feedback enabled teasing apart the effects of expectations from evidence on speakers' narratives. The results showed that speech was shaped not only by addressees' feedback, but also by speakers' expectations: speakers spent more time interacting with attentive addressees, but only when they also *expected* attentive addressees, and they spent more time interacting with distracted addressees only if they also expected these addressees to be distracted. Thus, whenever speakers' prior expectations matched the behavior they encountered in the interaction, they put more effort into the interaction by spending more time with their addressees. But there were further nuances; speakers were more likely to add vivid or extra detail to the jokes when they had attentive addressees, but only when they also expected them to be attentive. This suggests that speakers construe addressee feedback on the basis of their expectations about addressees' behavior.

Finally, one study so far has considered speakers' gestural adaptations to both bottom-up information, in the form of addressee feedback, and top-down information, in the form of situational affordances (Jacobs and Garnham 2007). In this study, participants described comic strips to two different confederate addressees, one of whom was trained to behave in an attentive manner (using gaze, posture, verbal and nonverbal feedback), while the other was trained to behave inattentively. Speakers were told that confederates either could or could not view the comic strips during the interaction; we consider this to be top-down information, since it was available to speakers prior to speaking. Speakers gestured less frequently when confederates were inattentive, as well as when confederates had visual access to the comics. There was a marginal interaction between comics' visibility and confederates' attentiveness: when the comic strip was not visible to the confederates, speakers gestured more frequently to the attentive than to the inattentive confederate; however, when the comic strip was visible, speakers gestured equally often to attentive and inattentive confederates. Jacobs and Garnham (2007) suggested that, when faced with an inattentive addressee, speakers may have attributed the addressee's lack of feedback to a preoccupation with looking at the comic strip (as opposed to disinterest) and therefore did not adapt their rate of gesturing.

For the current experiment, we returned to the speech corpus collected in Kuhlen and Brennan (2010) and transcribed, coded, and analyzed the co-speech gestures in a subset of the corpus. Although Jacobs and Garnham's findings are consistent with studies suggesting that speakers interpret bottom-up cues in light of top-down information (Kuhlen and Brennan 2010; Teufel et al. 2009), they did not explicitly make claims about the respective contribution of these sources of information. The top-down and bottom-up information in Jacobs and Garnham's study concerned two very different characteristics of the addressee—their general attentiveness and whether they could see the object being described. The design of the Kuhlen and Brennan corpus crosses top-down and bottom-up sources of information about a single characteristic of addressees—their attentiveness. This enables us to examine more directly how people integrate top-down and bottom-up sources of information to draw inferences about their partner's needs and adapt their behavior accordingly.

Another difference between the present study and Jacobs and Garnham's is that the latter investigated adaptive behavior only quantitatively, in terms of how frequently speakers gestured. Although gesture frequency is a useful measure, other qualitative variations in terms of gesture size, location, complexity and precision allow for a more nuanced understanding of the gesture production process. Specifically, quantitative and qualitative adjustments in gesture may reflect different processes of gesture production, similar to encoding propositional content and selecting articulatory plans in speech production. In line with this distinction, some models of gesture production posit processes responsible for formulating a gesture and processes responsible for its execution and motor control (e.g. Krauss et al. 2000; de Ruiter 2000). While there is evidence that information about the communicative situation or the addressees' needs can affect gesture formulation, as reflected in adjustments in gesture frequency (e.g. Alibali et al. 2001; Bavelas et al. 1992), it is less clear whether the motoric execution of gestures is similarly affected. Moreover, there is some evidence that addressees can benefit from qualitative adjustments in gesturing, insofar as they pay more overt attention to gestures produced in the body's periphery (Gullberg and Holmqvist 1999; but see Gullberg and Kita 2009). Gestures whose adjustment involves exaggerated movement (e.g. increasing the size of the gestures, or how high or peripherally they are executed in gesture space) have the potential to capture attention or may instead reflect a vivid style of narrating, and could be particularly informative about how speakers adapt in response to the addressees' (expected or actual) needs. Whether such qualitative adjustments are primarily top-down (i.e. driven by speakers' expectations) or can incorporate bottom-up information regarding the addressees' needs has yet to be determined.

In this work, we examine whether quantitative and qualitative aspects of gesture are shaped jointly by incoming, bottom-up cues in light of top-down expectations about addressee's attentiveness. Considering both quantitative and qualitative adjustments in gesture can provide a more complete account of whether, in the course of gesture production, social factors can have similar effects on both the formulation and the motoric execution of gestures.

2. Method

2.1. *Corpus design*

The experimental design of the original Kuhlen and Brennan corpus (2010) consisted of four conditions from manipulating two factors: (a) the speakers' expectations—whether speakers expected to be interacting with distracted or attentive addressees, and (b) the addressees' actual attentiveness—whether they were distracted or attentive during speakers' narrations. Thus, in two of the conditions speakers held mistaken expectations about their addressees' behavior: they expected distracted addressees but the addressees were, in fact, attentive, and they expected attentive addressees but the addressees were, in fact, distracted.

2.1.1. *Participants.* Seventy-eight undergraduates from Stony Brook University who were native speakers of English (32 men and 46 women) participated in 39 dyads. The gender composition of each pair was balanced across experimental conditions and conversational role of being either speaker or addressee. With the exception of one dyad, all participants were unacquainted previous to the experiment.[1] Participants received research credit toward a course requirement or were paid $8.

2.1.2. *Materials.* The original corpus was elicited by having speakers retell two written jokes; only one joke was included in the sub-corpus used for this study (several participants did not find the other joke to be funny). In this joke, an atheist has an encounter with a bear, which leads him to enter a dialogue with God, ultimately resulting in the bear converting to Christianity (see Appendix for the original 313-word text of this joke).

2.1.3. *Procedure.* Participants were randomly assigned to the role of either speaker or addressee. Speakers were taken to a separate room to receive instructions and study the joke, which was given to them in written text form. Half of the speakers were informed that their addressees would be working on a second task, and they should therefore not be surprised if their addressees seemed distracted.[2] The exact nature of their addressees' distraction was not revealed. The other speakers were not informed about any second task. All speakers were told that addressees would later have to recall the jokes and that their performance as a team would be judged on the accuracy of their addressee's recall. Speakers were given time to study the joke until they felt ready to retell it from memory.

While speakers were studying the joke, half of the addressees were instructed to listen carefully to the joke that speakers were going to retell them, since they later would have to retell it themselves. The other half of the addressees was instructed to not only listen to the joke but also to secretly count the number of times speakers used the word *and*. This manipulation was inspired by a procedure developed by Bavelas et al. (2000), which was shown to cause a decrease in the amount of feedback addressees are able to give to their speakers.

The set-up during the joke tellings was as follows: speaker and addressee were seated across from each other at a slight angle, and two cameras, shooting 'over the shoulder' of the other participant, recorded the speaker and the addressee, respectively. The position of their chairs and of the video cameras was marked on the floor and therefore remained identical across all sessions.

[1] Excluding this dyad from our analysis did not affect any of the results.

[2] Given the human tendency to juggle multiple tasks, this task is not particularly unnatural. Consider speaking to an interlocutor who is also paying some degree of attention to a sporting event on television, or one who is viewing a scrolling screen on a hand-held electronic device; speaking on the phone with an interlocutor who is known to be driving a car supports different expectations about their attention as well.

2.1.4. *Transcription and speech coding.* Speakers' joke tellings were transcribed in detail (see Kuhlen and Brennan 2010). To segment for coding purposes, a script of narrative elements was developed from the text of the original joke. Narrative elements were considered to be idea units that advanced the plot of the joke. The speakers' joke re-tellings were then also segmented into narrative elements, compared to the original script of narrative elements, and classified as either matching a narrative element in the original joke ('original element'), as adding extra details that were not part of the original joke ('extra detail'), or as meta-narrative utterances not advancing the narrative (e.g. *I'm gonna tell you a joke* classified as 'other'; see Kuhlen and Brennan 2010, for more details). Once utterances were classified according to the script, the number of words speakers used to complete a narrative element was determined. Vocalizations such as *uh*, *um* or *mhm* did not count as words. Contracted forms such as *don't* and *it's* counted as one word. Words that were aborted before completion were not counted.

2.1.5. *Addressee feedback.* Feedback responses were also recorded in the transcript. Feedback was defined as verbal and/or nonverbal responses on the part of addressees that indicated that they were attending, following, appreciating, or reacting to what speakers were saying (Bavelas et al. 2000). Included were verbal contributions to the interaction, as well as head nods and vocalizations such as *yeah*, *mhm*, *huh*, and laughter. Excluded from the analysis was behavior such as eye blinking, raising an eyebrow, or faint smiling that was too ambiguous or subtle to detect in a reliable fashion. Addressee feedback that continued as the speaker told the joke (e.g. nodding continuously) was coded as one feedback response. Different types of addressee feedback that occurred at the same time and appeared to convey the same meaning (e.g. head nodding while saying *yeah*) were also coded as one feedback response. For the current analysis, the number of addressee feedback responses was counted for each narrative element that was realized (see below) and normalized by the number of words speakers used (number of addressee feedback responses per 100 speaker words).

2.2. *Corpus sampled for the current study*

For the joke used in the present study, 13 of the 44 narrative elements in the joke were selected for gesture analysis. These 13 elements were selected from consecutive sequences of narrative elements at the beginning, middle and end of the joke (see Appendix), because they were consistently mentioned across speakers (they concerned critical points in the narrative, including the joke's punchline) and according to preliminary observations, were often accompanied by gestures. In order to avoid confounding variation in gesture production with variation in speech production, only references to material that was part of the original written version of the joke were analyzed; any gestures that may have been produced when a speaker provided extra detail for that element were excluded. Because not all 39 speakers mentioned all 13 narrative elements, the corpus included a total of 402 narrative elements, amounting to 3758 words. Our previous analysis of the larger speech corpus showed that although speakers across conditions differed in the amount of extra detail they provided, they did not differ in the number of narrative elements they included from the original joke (Kuhlen and Brennan 2010).

2.3. *Gesture coding*

The video sequences containing the 13 narrative elements were excised from the larger corpus for gesture coding. The onsets and ends of the target narrative elements were adjusted so that gestures were maintained in their entirety, regardless of whether their beginning or end overlapped partially with descriptions of other elements (or pauses). Gestures co-occurring with speech for a particular narrative element were classified as belonging to that narrative element under the assumption that speech and gesture are temporally synchronized and co-expressive (McNeill and Duncan 2000; McNeill 1992). All irrelevant hand movements that were not gestures (e.g. self-adaptors, such as scratching nose or adjusting glasses) were excluded from coding. The remaining corpus was coded for gesture frequency (a quantitative measure), gesture location, gesture height, and gesture size (qualitative measures). In a separate analysis, gestures were further differentiated into different types. Because there was no evidence that our experimental manipulations affected gesture types differently[3], the remaining analyses combine all gesture types.

[3] Gestures were differentiated into three different types: representational, meta-narrative, and ambiguous. Representational gestures were defined as gestures depicting semantic content by virtue of handshape, placement, and motion (Alibali et al. 2001). Meta-narrative gestures were defined as gestures that departed from the chain of events of the plot line, and instead emphasized information or supported the process of interacting in dialogue (e.g. requesting evidence for understanding), including beats, simple, rhythmic gestures that do not encode semantic content (Alibali et al. 2001; McNeill 1992), and what would have been classified as interactive gestures (Bavelas et al. 1992). Ambiguous gestures were gestures that could not be clearly classified as representational or meta-narrative. Two coders independently categorized the gestures in 25% of the corpus and agreed on 75.8% of their judgments on how to classify a certain gesture type. There was no evidence that speakers' expectations and addressees' attentiveness affected the frequency of specific gesture types differently. This may have been the case because there were significantly more representational than other types of gestures in this task: of the 332 gestures produced, 62.80% were classified as being representational, 16.94% were meta-narrative, and 20.25% were ambiguous. There were significantly more representational gestures per narrative element than meta-narrative gestures ($F(1, 36) = 49.50, p < .001$)

2.3.1. *Gesture frequency.* Two measures of gesture frequency were analyzed: the first measure considered the number of gestures produced for each narrative element, while the second normalized the number of gestures for a narrative element by the number of words for that element. Hand movements were coded as discrete gestures using the annotation tool ELAN (Brugman and Russel 2004). Following Seyfeddinipur (2006), the beginning of a gesture was defined by the onset of movement (characterized by a blurred image of the hand in the video frame). The end of a gesture was defined by the retreat of the hands to a rest position, which could be the location where the gesture had started (e.g. the lap) or any other location at which the hand had not moved for a period of more than 30 frames. For our purposes, a gesture was defined as having only one stroke (expressive part of a gesture, McNeill 1992). However, when the same gesture stroke was repeated many times in a cyclical manner these repetitions counted as only one gesture. In all other cases, hand movements with more than one stroke were divided into separate gestures. If a gesture was interrupted before the meaningful unit was completed, the gesture was excluded from further analyses. Audio was available during gesture identification.

Two coders (the first author and a naive research assistant), blinded to experimental conditions (the speaker's expectation of the addressee and the addressee's actual attentiveness), jointly identified gestures for 8% of the dataset for training purposes. Reliability in identifying gestures was determined by comparing judgments made independently by the two coders for an additional 25% of the corpus. Agreement between the two coders was established as the likelihood of both coders identifying a narrator's behavior as a gesture. Coders agreed on 78.8% of their judgments. After establishing reliability, remaining disagreements were discussed and resolved among the coders. After this process of calibration, each coder then coded half of the remaining data.

2.3.2. *Gesture location: Extremely peripheral gestures.* Beyond gesture frequency, qualitative aspects of gesture were also assessed. First, the location of gestures was judged according to how centrally versus peripherally they were executed with respect to the speaker's body. Using conventions for segmenting gesture space (McNeill 1992), each gesture was classified as being executed in one of four locations relative to the speaker's body: (1) Center-Center, (2) Center, (3) Periphery, and (4) Extreme Periphery. Figure 1 illustrates the breakdown of the different gesture locations. When judging gesture

or ambiguous gestures (F (1, 36) = 49.53, $p < .001$). The preponderance of representational gestures in our joke-telling task is consistent with the distribution of gesture types in other narrative tasks, including cartoon narrations, in which characters' physical actions were critical to the story (e.g. Alibali et al. 2001).

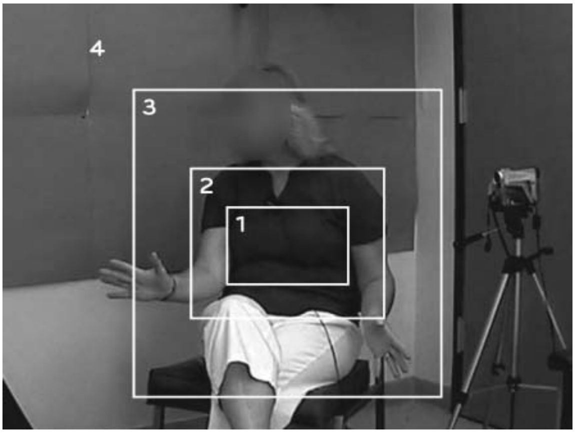

Figure 1. *Gesture space segmented according to periphery (based on McNeill 1992).*

location, coders considered the most extreme location of hands, instead of the location of the hands during a gesture stroke. For two-handed gestures that were asymmetrical, coders judged gesture location according to the hand that ended up in the more extreme location.

Two additional coders (the second author and a second naive assistant), blinded to conditions, watched the segmented gestures with sound and coded them independently for gesture location. One coder did 100% of the coding and the second coder did 25% of the coding (coding the middle sequence of the joke for 30 participants), following a training session of coding together 1/12 of the corpus. Reliability between the two coders on gesture location and the other two qualitative measures (gesture height and gesture size, described next) was established by calculating agreement between coders' judgments for the 25% of the corpus they both coded using Cohen's Kappa (Cohen 1968). Kappa values were interpreted based on Landis and Koch (1977: 165). Diverging coding decisions were resolved by discussing them until consensus was reached among the coders.

Since addressees are known to pay more overt attention to gestures in the body's periphery (Gullberg and Holmqvist 1999), we were interested in whether speakers would gesture in the extreme periphery of gesture space in response to addressees' (expected or actual) attentiveness. In our subsequent analyses we therefore focused only on the proportion of narrative elements that included a gesture in the extreme periphery. Narrative elements with multiple gestures were considered to have a gesture in the extreme periphery if at least one of these gestures was in that location. The coders showed substantial agreement on whether a gesture was in the extreme periphery, $K = .80$.

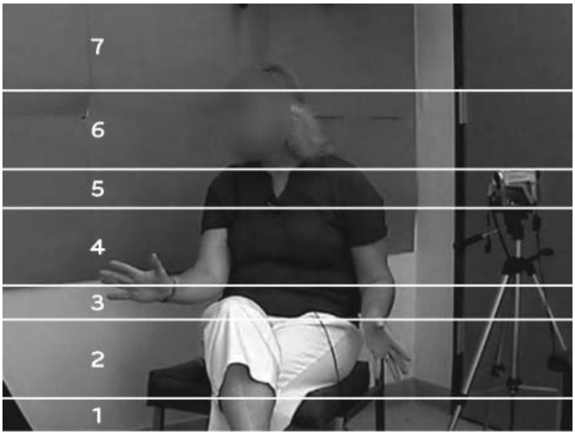

Figure 2. *The same gesture, with gesture space segmented according to height (based on McNeill 1992).*

2.3.3. *Gesture height: High and low gestures.* In addition to judging how peripherally speakers gestured with respect to their body, coders also judged how high in gesture space speakers executed their gestures. Using conventions for segmenting gesture space (McNeill 1992), each gesture was classified as belonging to one of seven segments according to how high it was executed: (1) Lower Extreme Periphery, (2) Lower Periphery, (3) Lower Center, (4) Center-Center, (5) Upper-Center, (6) Upper Periphery, and (7) Upper Extreme Periphery (see Figure 2). Using similar guidelines as those for gesture location, when coding gesture height, coders considered the highest location of the hands (rather than the height during the stroke phase). For asymmetrical two-handed gestures, coders considered the height of the hand reaching the highest location.

Again, we focused on gestures involving exaggerated movement; this time in terms of whether they were executed high or low in gesture space. Specifically, in our subsequent analyses we considered the proportion of narrative elements with a gesture in two areas of gesture space. First, we considered the proportion of narrative elements with gestures in the Upper-Center or higher (*high gestures*). Secondly, we considered the proportion of narrative elements with a gesture in Lower Periphery or Lower Extreme Periphery (*low gestures*). High gestures may reflect exaggerated adjustments in response to addressees' (expected or actual) attentiveness, while conversely low gestures may reflect the reverse pattern of adjustment. For narrative elements with more than one gesture, we first computed the average height for the element (with Lower Extreme Periphery = 1, Lower Periphery = 2, and so on), and then classified it as having a low gesture if the average height was 2 or lower and as having a high gesture if the average height of the gestures was 4 or higher.

The two coders showed almost perfect agreement in classifying gestures as either low (i.e. executed in Lower Periphery or Lower Extreme Periphery), medium (i.e. executed in Lower Center or Center-Center), or high (i.e. executed in Upper-Center, Upper Periphery, or Upper Extreme Periphery), $K = .86$.

2.3.4. *Gesture size.* The final qualitative measure of gesture was gesture size, which was defined as the amount of space that the speakers' hands spanned across while gesturing. It involved both the displacement of the speakers' hands while gesturing (e.g. the length of an upward trajectory of a single hand) and also the space between the speakers' hands in two-handed gestures. Coders gave a judgment for the gesture size of each gesture using a 1–7 scale. When judging the dislocation of the hands for two-handed gestures where the movement was asymmetrical, the coders considered the dislocation of the hand that moved the most. For narrative elements with more than one gesture, the average gesture size was computed.

The two coders showed substantial agreement in giving identical or nearly identical judgments to the gestures' size, $K = .73$. Coders differed by at most one score in 97% of all coding decisions.

2.4. *Analyses and baselines for the gesture corpus*

Analyses of mention of narrative element, gesture frequency, gesture type, gesture height and location were conducted as 2 × 2 ANOVAs with the addressees' attention (attentive addressee vs. distracted addressee) and the speakers' expectation (expecting attentive addressee vs. expecting distracted addressee) as between-subjects factors. The distribution of the variables of narrative elements realized by speakers, gesture location and gesture height did not conform to the assumption of normality that underlies the parametric ANOVA. For these variables, a nonparametric adjusted rank transform test was performed (Leys and Schumann 2010). Where appropriate, we also report correlations between speakers' gestural and spoken behavior and perform 2 × 2 ANOVAs, with speech measures as covariates.

We established that the subset of the corpus coded here for gesture was similar to the larger speech corpus (Kuhlen and Brennan 2010), as follows (see Table 1 for means and standard deviations). As in the larger corpus, counting *ands* was distracting for addressees in the subset corpus; addressees who counted *ands* gave significantly less feedback than those who did not, $F(1, 35) = 4.29$, $p < .05$. This suggests that feedback is the mechanism by

Table 1. *Means and standard deviations for baseline measures addressee feedback, original narrative elements and number of words used by speakers for corpus sampled for this study.*

Measures	Attentive addressee		Distracted addressee		*Total*	
Addressee feedback per 100 speaker words[a]						
Speaker expects attentive addressee	4.22	(6.29)	2.33	(5.20)	3.33	(5.87)
Speaker expects distracted addressee	4.31	(5.76)	2.30	(4.77)	3.25	(5.34)
Total	4.26	(6.03)	2.32	(4.96)		
Percentage of original narrative elements realized by speaker						
Speaker expects attentive addressee	80.00	(40.15)	78.63	(41.16)	79.35	(40.56)
Speaker expects distracted addressee	76.98	(42.26)	83.85	(36.95)	80.47	(39.72)
Total	78.52	(41.15)	81.38	(39.01)		
Number of words used by speaker[a]						
Speaker expects attentive addressee	9.50	(3.85)	9.18	(3.42)	9.35	(3.65)
Speaker expects distracted addressee	9.35	(3.63)	9.21	(3.87)	9.28	(3.75)
Total	9.43	(3.74)	9.20	(3.66)		

[a] Measure records behavior for each narrative element that was realized.

which an addressee's attentiveness shapes a speaker's behavior. Also as in the larger corpus (ibid.), narrative elements were just as likely to be mentioned across all of the experimental conditions; similarly, in the subset corpus, speakers were equally likely to mention a particular narrative element regardless of their expectations about addressees' level of attention, $F(1, 35) = .02$, *n.s.*, or their addressees' actual attentiveness, $F(1, 35) = .00$, *n.s.*, with no interaction between those two factors, $F(1, 35) = 1.44$, *n.s.* This baseline equivalence in content (narrative elements) realized across experimental conditions also extended to the number of words used to express each narrative element in the sub-corpus: word counts were similar, regardless of expectation of addressees' level of attention $F(1, 35) = .01$, *n.s.* and of addressees' attentiveness, $F(1, 35) = .30$, *n.s.*, again with no interaction, $F(1, 35) = .30$, *n.s.* Therefore, any effects of speakers' expectations and addressees' attentiveness on gesturing are not artifacts due to differences in amount of speech, since word counts are the same across conditions.

3. Results

3.1. *Adaptation in gesture frequency*

Figure 3 shows the mean frequency of gesturing for all four experimental conditions. Overall, speakers gestured with comparable frequency whether interacting with attentive or distracted addressees (.80 gestures per narrative ele-

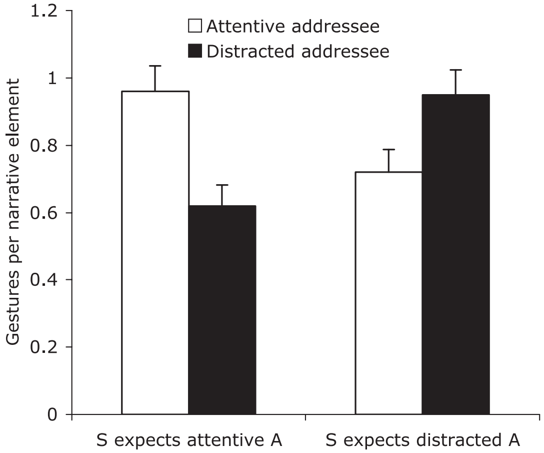

Figure 3. *Gesture frequencies, according to addressees' attentiveness and speakers' expectations (bars represent standard error).*

ment in both conditions), *F* (1, 35) = .01, *n.s.*, and whether expecting distracted or attentive addressees (.80 and .81 gestures per narrative element, respectively), *F* (1, 35) = .00, n.s. However, depending on their expectations, speakers reacted differently to addressees' displays of attention, as shown by the interaction of these two factors, *F* (1, 35) = 7.55, *p* < .01. Specifically, speakers gestured more frequently when their expectations were consistent with addressees' behavior: they gestured more when they expected attentive addressees and addressees were indeed attentive, and when they expected distracted addressees and addressees were indeed distracted. When speakers' expectations mismatched addressees' attentiveness, speakers gestured less frequently. Not surprisingly, the same pattern held when gestures were normalized by words rather than by narrative elements. Speakers produced similar rates of gestures per word when interacting with attentive as with distracted addressees, *F* (1, 35) = .11, *n.s.*, and whether expecting attentive and distracted addressees, *F* (1, 35) = .001, *n.s.*, but gestured with greater frequency when the feedback they received matched their expectations, *F* (1, 35) = 4.51, *p* < .05 (see Table 2).

The adjustments speakers made in gesturing, as shaped by their own expectations and their addressees' feedback, were not a direct consequence of any adjustments they made in amount of speech. Although the number of words, when entered as a covariate, marginally predicted the number of gestures, *F* (1, 34) = 3.69, *p* = .06, the interaction between speakers' expectations and addressees' feedback remained significant, *F* (1, 34) = 7.79, *p* < .01.

Table 2. *Means and standard deviations for gesture size and height relative to speakers' expectations and addressees' attentiveness.*

Measures	Attentive addressee		Distracted addressee		*Total*	
Gestures per 100 speaker words						
Speaker expects attentive addressee	10.76	(8.22)	7.93	(9.13)	9.43	(8.75)
Speaker expects distracted addressee	8.74	(8.06)	11.25	(8.39)	10.07	(8.31)
Total	9.78	(8.19)	9.73	(8.87)	9.76	(8.53)
High gestures[a]						
Speaker expects attentive addressee	0.33	(0.48)	0.23	(0.42)	0.28	(0.45)
Speaker expects distracted addressee	0.25	(0.44)	0.18	(0.39)	0.22	(0.42)
Total	0.29	(0.46)	0.21	(0.41)	0.25	(0.43)
Low gestures[a]						
Speaker expects attentive addressee	0.22	(0.42)	0.43	(0.50)	0.33	(0.46)
Speaker expects distracted addressee	0.29	(0.46)	0.40	(0.49)	0.35	(0.48)
Total	0.26	(0.44)	0.42	(0.50)	0.34	(0.47)
Gesture size[b]						
Speaker expects attentive addressee	2.93	(1.32)	2.42	(1.52)	2.72	(1.42)
Speaker expects distracted addressee	2.64	(1.21)	2.40	(1.14)	2.50	(1.17)
Total	2.80	(1.27)	2.41	(1.30)	2.61	(1.30)

[a] Gesture height is based on mean proportion of narrative elements that included a high, or low gesture respectively.

[b] Gesture size was rated on a scale from 1 (small)–7 (large).

3.2. *Qualitative adaptations in gesturing*

After establishing that gesture frequency increased when addressees' attentiveness matched speakers' expectations, we examined whether these two factors affected qualitative aspects of gesturing—how peripherally, how highly, and how largely speakers gestured. Speakers produced a higher proportion of narrative elements having extreme peripheral gestures *only* when addressees were expected to be attentive and indeed were attentive, as revealed through a significant interaction, $F(1, 35) = 4.26, p < .05$. In fact, in this condition speakers were more than twice as likely to gesture in the extreme periphery than in all other conditions (19% vs. 8%), reliably different according to a nonparametric Mann-Whitney Test performed as post hoc contrast, $U = 79.5, p < .05$. (There was a marginal main effect of expectation, $F(1, 35) = 3.45, p = .07$, but this was driven entirely by the interaction). Figure 4 shows the mean proportion of all narrative elements containing gestures in which a gesture was executed in the extreme periphery, according to addressees' attentiveness and speakers' expectations.

As for the other qualitative aspects of gesture, their patterns of results were not reliable. Addressees' attentiveness had a marginal effect on speakers' gesture size: speakers tended to produce marginally larger gestures with attentive

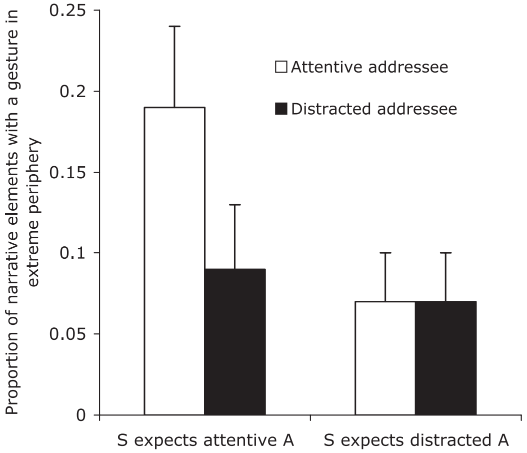

Figure 4. *Mean proportions of narrative elements with a gesture in the extreme periphery according to addressees' attentiveness and speakers' expectations (bars represent standard error).*

addressees than with distracted ones, $F(1, 35) = 3.84$, $p = .06$. They produced (numerically but not reliably) a higher frequency of high gestures (in the upper-center of the gesture space or higher) to attentive than distracted addressees, $F(1, 35) = 3.72$, $p = .06$, and conversely, a marginally higher frequency of low gestures (in lower periphery and extreme lower periphery) to distracted than attentive addressees, $F(1, 35) = 2.98$, $p = .09$. Speakers' expectations about addressees' attentiveness did not affect the height or the size of gestures, nor was the interaction between attentiveness and expectation significant. Table 2 shows the proportion of narrative elements with high gestures, low gestures, and the gestures' mean size for all experimental conditions.

4. Discussion

Our findings demonstrate that gesturing is shaped simultaneously by top-down and bottom-up factors. Speakers gestured more frequently when their expectations matched their addressees' actual behavior (when they expected attentive addressees and their addressees were indeed attentive, and when they expected distracted addressees and their addressees were indeed distracted). Top-down expectations appear to frame bottom-up interpretation of addressees' moment-by-moment feedback. When partners in communication are correct in their assumptions about one another, particularly concerning each other's ability to attend to the primary discourse task, they may find the interaction more engaging (see De Jaegher et al. 2010); when one is unaware of the other's distinct goal, coordination may be less fluent (see Wilkes-Gibbs 1986 for similar observations about conversations in which partners are unaware of one another's distinct goals). The difference between our findings and those of Richardson et al. (2009), who found closer coordination when there was a perceived mismatch between interlocutors' perspectives, may be due to differences in whether a behavior was able to be used as a communicative signal; the gestures in our study were used by speakers to affect (or to try to affect) addressees' behavior. How speakers adapt to their addressees may therefore vary depending on the conversational context and task goals.

In our study, top-down and bottom-up information about the partner affected both the formulation and the motoric execution of gestures: speakers' expectations and addressees' attentiveness affected not only how *much* they gestured, but also *how* they gestured. Specifically, speakers produced more gestures in the extreme periphery of their gesture space only when they expected attentive addressees and their addressees were indeed attentive. These quantitative and qualitative adaptations in gesture parallel adaptations that we have previously found in speech (Kuhlen and Brennan 2010), and yet were not simply artifacts due to the narrative content expressed or the numbers of words spoken: even when looking only at those passages in which narrative content did not differ systematically across conditions, and with differences in numbers of words controlled, speakers still modified whether and how they gestured depending on whether they could accurately anticipate their addressees' behavior. This suggests that even if gestures arise from the same underlying representations as speech and are temporally coordinated with it at various points, partner-specific adaptations in gesture are not a direct consequence of adaptations in speech (even if they pattern similarly).

Our results extend previous findings showing that addressees shape speakers' gesturing (e.g. Alibali et al. 2001; Bavelas et al. 1992; Gerwing and Bavelas 2004; Holler and Stevens 2007; Jacobs and Garnham 2007; Özyürek 2000, 2002) by specifically testing (in an unconfounded way) the interaction of a top-down and a bottom-up factor, both having to do with the same phenomenon (addressee attentiveness). In fact, in our study, speakers' expectations and addressees' feedback each (by themselves) had a limited effect on gesturing. Unlike the finding of Jacobs and Garnham (2007) that speakers gestured more to attentive than to distracted addressees, we did not find a main effect of addressees' attentiveness on speakers' gesture frequency. Because, in Jacobs and Garnham's study, speakers interpreted addressees' feedback in light of whether addressees could see what they were describing, integrating these two sources of information may have been more difficult than in our study since they concerned different characteristics of addressees (their visual availability and attentiveness, whose interaction was, notably, only marginally significant). As a result, in making inferences about their addressees' needs, speakers may have relied primarily on addressees' feedback, and adapted the frequency of their gestures accordingly. In our study, where both top-down and bottom-up information concerned the addressees' attentiveness, it was the interaction of speakers' expectations with addressees' feedback that accounted for most adaptations in gesturing. This confirms that gesturing is not simply an automatic response to feedback, but is affected by how that feedback is construed. Speakers not only monitor their addressees' behavior, but interpret it flexibly, based on what they know or expect.

This study supports the conclusion that adaptation to a partner is flexible rather than entirely automatic or 'dumb'; feedback signaling distraction was interpreted differently in the light of speakers' top-down expectations. In the present study, when speakers had an attribution for their addressees' distraction —that they were working on a second task—they gestured with comparable frequency as when they narrated to attentive addressees. In the mismatched conditions, speakers may have reacted, either explicitly or implicitly, by withdrawing from addressees they found disengaged (distracted addressee expected to be attentive) or by not putting any extra effort into entertaining an addressee who seemed to be able to attend adequately to two tasks at once (attentive addressee expected to be distracted). This could explain the quantitative finding of lower gesture rates for these two conditions, and higher gesture rates for both conditions in which both partners were mutually committed to the same goals. In contrast, for the qualitative measure, we found a somewhat different pattern, with more dramatic (peripheral) gestures produced only to attentive addressees expected to be attentive, which may reflect additional accommodation based on mutual engagement in the narrative task with such addressees, and on a lack of co-engagement with the other (actually—or assumed—distracted) addressees.

Concerning the extent to which top-down expectations may be gradually updated by incoming bottom-up cues, this is still an open question. Our narrative corpus does not lend itself well to addressing this question, in that a single joke, with its build-up to a punchline, lacks the rhetorical opportunities to support observing changes in expectations over time. Tasks without such narrative structure could be employed in future studies to investigate whether speakers' gestural strategies, adopted on the basis of expectations, are revised during the course of an interaction on the basis of addressee feedback.

Our findings are consistent with the previous analysis of the speech that accompanied these gestures (Kuhlen and Brennan 2010). In that study, speakers spent more time in the interaction when speakers' expectations matched addressees' feedback. This parallels our current finding that speakers also produce more gestures per narrative element in these conditions, strengthening our interpretation that speakers put more effort into narrating when their expectations of their addressees are matched by addressees' behavior in the interaction. Kuhlen and Brennan also found that when (and only when) speakers narrated to attentive addressees whom they had also expected to be attentive, they used more additional narrative details; in the current study, they used more peripheral gestures as well. Together, these adjustments suggest that speakers did not produce gestures in the periphery of their gesture space to attract their addressees' attention. Instead, these gestures may simply go hand in hand with a more vivid or engaged style of narrating. Having an attentive addressee is not enough by itself to lead to a vivid narration: expecting an attentive addressee is also necessary. Both of these apparently lead to a more engaged, and therefore a more coordinated, interaction.

Our evidence for speakers flexibly adapting their gestures in light of both feedback cues and expectations corroborates the idea that the behavior of conversational partners is closely coordinated (De Jaegher et al. 2010; Shockley et al. 2009). Speakers monitor cues about addressees' uptake and engagement, interpreting those cues against their prior expectations and knowledge, to facilitate the interpersonal 'mind-reading' needed to successfully coordinate a joint activity.

Appendix

Original text of the "Atheist" joke. The thirteen target narrative elements selected for gesture analysis are highlighted in bold print.

1. An atheist was taking a walk through the woods,
2. admiring all that evolution had created.
3. "What majestic trees!
4. What powerful rivers!
5. What beautiful animals!", he said to himself.
6. As he was walking along the river,
7. **he heard a rustling in the bushes behind him.**
8. **When he turned to see what the cause was,**
9. **he saw a 7-foot grizzly charging right towards him.**
10. He ran as fast as he could.
11. He looked over his shoulder and saw that the bear was closing,
12. He ran even faster,
13. crying in fear.
14. He looked over his shoulder again and the bear was even closer.
15. His heart was pounding
16. and he tried to run even faster
17. **He tripped and fell on the ground.**
18. **He rolled over to pick himself up, but**
19. **saw the bear right on top of him,**
20. **reaching for him with his left paw**

21. and raising his right paw to strike him.
22. At that moment, the atheist cried out "Oh my God! . . . ."
23. Time stopped.
24. The bear froze.
25. The forest was silent.
26. Even the river stopped moving.
27. As a bright light shone upon the man,
28. a voice came out of the sky,
29. "You deny my existence for all of these years;
30. teach others I don't exist;
31. and even credit creation to a cosmic accident.
32. Do you expect me to help you out of this predicament?
33. Am I to count you as a believer?"
34. The atheist looked directly into the light
35. "It would be hypocritical of me to suddenly ask You to treat me as Christian now,
36. but perhaps could You make the bear a Christian?"
37. "Very well," said the voice.
38. The light went out.
39. The river ran again.
40. And the sounds of the forest resumed
41. And then the bear dropped his right paw
42. . . . brought both paws together . . .
43. bowed his head
44. and spoke: "Lord, for this food which I am about to receive, I am truly thankful."

References

Alibali, M. W., D. C. Heath & H. J. Myers. 2001. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language* 44. 160–188.

Alibali, M. W., S. Kita & A. J. Young. 2000. Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes* 15. 593–613.

Argyle, M. & M. Cook. 1976. *Gaze and mutual gaze.* Cambridge: Cambridge University Press.

Bavelas, J. B., N. Chovil, D. A. Lawrie & A. Wade. 1992. Interactive gestures. *Discourse Processes* 15. 469–489.

Bavelas, J., L. Coates & T. Johnson. 2000. Listeners as co-narrators. *Journal of Personality and Social Psychology* 79. 941–952.

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

Brennan, S. E. 1991. Conversation with and through computers. *User Modeling and User-Adapted Interaction* 1. 67–86.

Brennan, S. E. 2005. How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (eds.), *Approaches to studying world-situated language use:*

Bridging the language-as-product and language-action traditions, 95–129. Cambridge, MA: MIT Press.

Brennan, S. E., Z. Chen, C. A. Dickinson, M. B. Neider & G. J. Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106. 1465–1477.

Brennan, S. E., A. Galati & A. K. Kuhlen. 2010. Two minds, one dialog: Coordinating speaking and understanding. In B. Ross (ed.), *The psychology of learning and motivation (Vol. 51)*, 301–344. Burlington: Academic Press.

Brugman, H. & A. Russell. 2004. Annotating multimedia/multi-modal resources with ELAN. Paper presented at the LREC 2004, Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal.

Church, R. B. & S. Goldin-Meadow. 1986. The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23. 43–71.

Clark, H. H. & S. E. Brennan. 1991. Grounding in communication. In L. B. Resnick, J. Levine & S. D. Teasley (eds.), *Perspectives on socially shared cognition*, 127–149. Washington, DC: APA.

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

Cohen, L. 1968. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70. 213–220.

de Jaegher, H., E. di Paolo S. & Gallagher. 2010. Can social interaction constitute social cognition? *Trends in Cognitive Sciences* 14. 441–447.

de Ruiter, J. P. A. 2000. The production of gesture and speech. In D. McNeill (ed.), *Language and gesture: Window into thought and action*, 284–311. Cambridge: Cambridge University Press.

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

Goodwin, C. 1979. The interactive construction of a sentence in natural conversation. In G. Psathos (ed.), *Everyday language. Studies in ethnomethodology*, 97–121. New York: Irvington.

Gullberg, M. & K. Holmqvist. 1999. Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition* 7. 35–63.

Gullberg, M. & S. Kita. 2009. Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior* 33. 251–277.

Hanna, J. E. & S. E. Brennan. 2007. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language* 57. 596–615.

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

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

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

Kendon, A. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26. 22–63.

Kraljic, T., S. E. Brennan & A. G. Samuel. 2008. Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107. 54–81.

Kraljic, T., A. G. Samuel & S. E. Brennan. 2008. First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science* 19. 332–338.

Krauss, R. M., Y. Chen & R. Gottesman. 2000. Lexical gestures and lexical access: A process model. In D. McNeill (ed.), *Language and Gesture*, 261–283. Cambridge: Cambridge University Press.

Kraut, R. E., S. H. Lewis & L. W. Swezey. 1982. Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology* 43. 718–731.

Kuhlen, A. K. & S. E. Brennan. 2010. Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes* 47. 567–587.

Landis, J. R. & G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33. 159–174.

Leys, C. & S. Schumann. 2010. A nonparametric method to analyze interactions: The adjusted rank transform test. *Journal of Experimental Social Psychology* 46. 684–688.

Liszkowski, U., K. Albrecht, M. Carpenter, M. & M. Tomasello. 2008. Infants' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior and Development* 31. 157–167.

McNeill, D. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. 2000. Catchments and context: non-modular factors in speech and gesture production. In D. McNeill (ed.), *Language and gesture*, 312–328. Cambridge: Cambridge University Press.

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

Özyürek, A. 2000. The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (ed.), *Language and gesture*, 64–83. Cambridge: Cambridge University Press.

Özyürek, A. 2002. Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language* 46. 688–704.

Richardson, D. C., R. Dale & J. M. Tomlinson. 2009. Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science* 33. 1468–1482.

Russell, A. W. & M. F. Schober. 1999. How beliefs about a partner's goal affect referring in goal-discrepant conversations. *Discourse Processes* 27. 1–33.

Seyfeddinipur, M. 2006. *Disfluency: Interrupting speech and gesture.* (MPI Series in Psycholinguistics, 39). Nijmegen, NL: Max Planck Institute of Psycholinguistics.

Shockley, K., D. C. Richardson & R. Dale. 2009. Conversation and coordinative structures. *Topics in Cognitive Science* 1. 305–319.

Teufel, C., D. M. Alexis, H. Todd, A. J. Lawrance-Owen, N. S. Clayton & G. Davis. 2009. Social cognition modulates the sensory coding of observed gaze direction. *Current Biology* 19. 1274–1277.

Teufel, C., P. C. Fletcher & G. Davis. 2010. Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Science* 14. 376–382.

Wilkes-Gibbs, D. 1986. *Individual goals and collaborative actions: Conversations as collective behavior*. Stanford, CA: Stanford University dissertation.

Yngve, V. H. 1970. On getting a word in edgewise. *Papers from the 6th Regional Meeting of the Chicago Linguistic Society*, 567–578. Chicago, IL: Chicago Linguistic Institute.