Knowing 'who she is' based on 'where she is': The effect of co-speech gesture on pronoun comprehension

WHITNEY GOODRICH SMITH[Δ] AND CARLA L. HUDSON KAM[μ*]

[Δ]*University of Maryland*
[μ]*University of British Columbia*

Abstract

We examine whether pronoun interpretation is affected by naturalistic co-speech gesture. Participants in three conditions watched narrations containing ambiguous pronouns. In one condition the narrator produced gestures consistent with order-of-mention; in another, they conflicted with order-of-mention; and in the third, she did not gesture. Results showed that when the gestures conflicted with order-of-mention participants were much less likely to interpret the pronoun as referring to the first-mentioned character. In a second experiment we ruled out the possibility that participants were simply picking up on differences within the speech itself. These results extend previous work on gesture and language processing by showing that the information in gesture can influence the way people interpret words which by their nature are ambiguous, and that this influence is similar to that of well-known speech internal cues.

Keywords
gesture, language comprehension, pronoun resolution, referent processing

1. Introduction

Throughout the world, wherever you see people speaking, you see them spontaneously moving their hands—gesturing while they speak (Goldin-Meadow 2003; Kendon 1981). Although there is a long-running debate about why people gesture (i.e. whether gesture is 'for the speaker,' e.g. Butterworth and Hadar 1989; Krauss et al. 2000; or 'for the listener,' e.g. Alibali et al. 2001;

* Correspondence address: Whitney Goodrich Smith, University of Maryland, College Park, Department of Hearing and Speech Sciences, Room 0100 Lefrak Hall, College Park, MD 20742, USA. E-mail: wgs@umd.edu.

Language and Cognition 4–2 (2012), 75–98
DOI 10.1515/langcog-2012-0005
1866–9808/12/0004–0075
© Walter de Gruyter

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Bavelas et al. 1992; Kendon 1994), there is evidence that gesture conveys information and that listeners make use of this information during language comprehension. For instance, when congruent with the message conveyed in speech, gestures improve comprehension (Beattie and Shovelton 1999; Campana et al. 2005); but they need not be congruent to influence comprehension. There is also evidence that listeners integrate information presented uniquely in gesture with that presented uniquely in speech such that the final interpretation includes information from both sources (Cassell et al. 1999; Singer and Goldin-Meadow 2005; Goodrich and Hudson Kam 2009).

The exact nature of the integration of information from gesture and speech has been debated (e.g. Hadar and Krauss 1999; McNeill 1992), however, it is clear that the two are not interpreted independently (e.g. Willems et al. 2007; Kelly et al. 2010; Krauss et al. 1991).[1] Krauss et al. (1991), for instance, showed that adult listeners interpreted the meaning of co-speech gestures differently when shown a muted video of a speaker gesturing while talking than when they saw the same video with sound. Indeed, only those who heard the sound version interpreted the gestures in a way that was consistent with the content of the speech. This relationship goes the other way as well; various studies have demonstrated that gesture can influence the way that ambiguous or hard-to-hear speech is interpreted (Rogers 1978; Thompson and Massaro 1994).

The current study expands on this work, exploring the role gesture plays on listeners' interpretation of ambiguous speech when the ambiguities arise due to the content of the speech, rather than the speech signal itself. We investigate this with respect to pronouns, word forms that by their very nature are often ambiguous and known to depend on a variety of probabilistically informative cues—that is, aspects of the language which are only partially correlated with reference patterns—for their interpretation (Arnold 1998, 2001). Notably, the kinds of gestures we investigate are likewise probabilistically available, insofar as they do not occur with every noun phrase in natural speech contexts. Thus, not only might we expect pronoun interpretation to be amenable to the influence of gesture, we might expect gesture to influence interpretation similarly to previously examined cues. The work is therefore relevant to questions

1. Although it is generally agreed that speech and gesture influence each other, the mechanism behind this is less clear. In particular, there has been much debate about whether speech and gesture form a unified system (as put forth by Kendon 1980, and McNeill 1992), or are two systems with information integrated at some later stage of processing (for instance, gesture serving as an auxiliary 'support system' for speech, Krauss 1998). Most of the recent work addressing this issue supports the unified system view, both from the perspective of production (e.g. So et al. 2009) and perception (Kelly et al. 2010; Kelly et al. 2004; Wu and Coulson 2007). As the present study does not speak directly to this question (one system versus two), however, we will not discuss it further.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

of gesture-speech integration, in that it extends what we know about the kinds of integrations listeners can perform, as well as to questions of language comprehension more broadly, extending ideas about the integration of cues during the comprehension process.

1.1. *Background*

In discourse, the same referents are often referred to multiple times. The way an entity is referred to when it is introduced is rarely the way it is referred to later in the same discourse. Instead, speakers and signers often shift from using names or full noun phrases to pronouns.[2] In (1), for instance, the speaker shifts from using the name *Ann* to the pronoun *she*.

(1) Ann is having a picnic in the park. She has a lot of food with her, but she is especially excited about the cookies.

Pronouns are often shorter and easier to pronounce than full NPs, making speaking more efficient. However, they can also introduce a degree of ambiguity, potentially making comprehension more difficult. This is particularly true when we use pronouns after describing more than one person. For example, in (2), it is not clear whether the visit occurred during John's or Sam's summer vacation.[3]

(2) John visited Sam during his summer vacation.

Listeners are often unperturbed by this referential of ambiguity, and quickly and easily assign interpretations to seemingly ambiguous pronouns using a variety of cues such as gender (Arnold et al. 2000), grammatical function (Gorden et al. 1993), and order-of-mention (Arnold et al. 2000). The influence of a cue such as gender is quite obvious; if there are two previously mentioned characters, but only one matches the gender of the pronoun, the pronoun will be interpreted as referring to the gender-matched character. The influence of the other cues is somewhat more subtle, however, with interpretation being guided, rather than determined by them. In the case of grammatical role, for instance, pronouns tend to be interpreted as referring to characters introduced as grammatical subjects (see Arnold 1998, for a review). This is not an absolute pattern in pronoun interpretation, however. Indeed, sometimes the character introduced as object may be just as likely, or even more likely, to be interpreted as the referent of the pronoun (Arnold 2001; Kehler 2002; Kousta 2008),

2. Speakers do not always switch to using pronouns, as evidenced by lexical entrainment—a phenomenon which occurs during a conversation when people converge upon, and repeat, similar terms for objects (e.g. Brennan and Clark 1996).
3. The potential ambiguity only exists for the listener; clearly, the speaker knows who 'his' refers to.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

especially when the pronoun is not a subject pronoun (see e.g. Chambers and Smyth 1998).[4] Order-of-mention is another cue listeners have been found to be sensitive to. In particular, studies have found that listeners tend to interpret a pronoun as being co-referential with the first-mentioned character. For instance, Gernsbacher and Hargreaves (1988) found that participants were faster at recognizing a probe for the first-mentioned noun phrase. Arnold et al. (2000) subsequently found that listeners were faster at identifying the referent of an ambiguous pronoun when it was the first-mentioned character in a story than when it was the second. As with grammatical role, however, people do not always interpret the pronoun as referring to the first-mentioned character (Järvikivi et al. 2005).

Of course, cues usually co-occur in natural speech. Studies that have unpacked the role of multiple, and at times conflicting, cues to pronoun resolution have shown that adult listeners are especially likely to interpret ambiguous pronouns as referring to first-mentioned entities when order-of-mention agrees with an additional cue to meaning, such as gender (Arnold et al. 2000), or grammatical subject (Järvikivi et al. 2005). Importantly, however, listeners will usually rely on other cues when they conflict with order-of-mention, suggesting that order-of-mention is a weaker cue than the others examined (cf. MacWhinney et al. 1984). Järvikivi et al. (2005), for instance, examined the influence of grammatical role and order-of-mention on subject pronoun resolution in Finnish—a language with free word order and a gender-neutral third person singular pronoun. They found effects for both cues, but the effect for grammatical role was stronger and seen earlier in the eye-tracking data. Nouns introduced as subjects attracted more looks than those introduced as objects immediately after the pronoun was heard, but object nouns introduced in OVS sentences attracted more looks than object nouns introduced in SVO sentences. However, this order-of-mention effect was about 200 ms slower than the grammatical role effect. This work suggests that rather than relying on a single cue, such as a salient first-mentioned entity (e.g. Gernsbacher and Hargreaves 1988), or a preference for grammatical subjects (e.g. Crawley et al. 1990), listeners are relying on several strategies in combination.

These results make sense from the perspective of a constraint-based theory of language comprehension in which multiple relevant cues interact to make some interpretations more likely than others. With respect to pronoun interpretation specifically, according to Arnold's (1998, 2001, 2010) Expectancy Hypothesis, the relevant cues are anything that makes something more likely

4. There are actually two different tendencies interacting here that we have lumped together in our brief discussion of grammatical role. One is the tendency of nouns introduced as subjects to be more salient, something central to Centering Theory (Grosz et al. 1995). The other is syntactic parallelism (Chambers and Smyth 1998).

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

to be mentioned in that position in the discourse (meaning point in time, position in the sentence, etc.), based on the listener's experience with the language. Different cues have more or less impact on interpretation, depending in part on their overall reliability, which stems from the correlations present in the language. Moreover, the impact of any individual cue on interpretation can differ according to the other cues available in that instance. That is, a cue that may only weakly impact interpretation on its own may have a large impact when it occurs with another cue. Of particular relevance for the present work, the Expectancy Hypothesis predicts that even rare or atypical cues can have an impact on interpretation (Arnold, Altmann et al. 2004; Arnold, Hudson Kam et al. 2007). A cue need only be relevant and interpretable by the listener to affect interpretation.

Here we explore whether gestures related to reference can also serve as cues to pronoun interpretation. Research shows that speakers sometimes produce what we refer to as *coreferential localizing* gestures when describing multiple entities (Foraker 2011; Foraker and Goldin-Meadow 2007; So et al. 2005; So et al. 2009). That is, speakers may gesture in a location in space when referring to an entity and then gesture back to that same location when referring to the same entity later in the discourse. For example, while saying the name *John*, a speaker might place her right hand palm down at her right side, and gesture back towards that same space when mentioning John again later on.

These gestures can be deictic, such as points, or more iconic gestures encoding the shape or movement of an entity (Foraker and Goldin-Meadow 2007). What is crucial about these gestures is not their form, but the way they point back to a space previously associated with a character. This metaphoric mapping between characters and spatial loci gives rise to spatial coreference patterns that reflect the pronoun coreference intended by the speaker. This is similar to the idea of gesture catchments described by McNeill (2000), where recurrent features of a gesture (in this case, consistent use of space) suggest a common discourse theme. If listeners are able to associate the space in which a gesture occurred with the character mentioned while the gesture was produced and can remember which spaces were associated with which characters, this spatial representation may then affect how the pronouns are interpreted. The listener can link the referent with the pronoun according to the spatial location gestured to. In this way they are relevant, and so fulfill one of the conditions for cues. Previous research has revealed gesture to be a useful cue when listeners are faced with ambiguity from other sources, such as pragmatically ambiguous deixis (Tfouni and Klatsky 1983), or phonetically ambiguous speech (Thompson and Massaro 1994). For instance, in a study by Thompson and Massaro (1994), four and nine-year old children saw two objects, a ball and a doll, and heard either /bɔl/, /dɔl/, or an ambiguous blend of the two syllables, along with a pointing gesture to one of the objects. Perhaps unsurprisingly,

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

children were more likely to select the ball when they heard /bɔl/ along with a point at the ball than when they heard /bəl/ and a point to the doll. Of particular relevance to the current work, when presented with the ambiguous intermediate blend of /bɔl/ and /dɔl/, participants relied on the gesture to resolve the ambiguity. Thus, even children can rely on pointing gestures to resolve ambiguity in speech. However, it is unknown whether listeners are sensitive to information contained in more abstract gestures (specifically, the consistent use of space as a metaphorical mapping for referents), and can rely on these gestures to inform their interpretation of an ambiguous pronoun. If listeners are indeed sensitive to this information it would provide further evidence for the direct influence of gesture on the interpretation of speech.

It is also interesting to consider the question of how gesture may interact with other cues. For instance, when presented with one cue to pronoun reference in gesture and a conflicting speech internal cue (order-of-mention), which modality do listeners rely on? And when the two modalities agree, are listeners even more likely to select the first-mentioned name? Based on the pronoun literature, specifically the results of Arnold et al. (2000) and Järvikivi et al. (2005), we would expect the influence of order-of-mention to be particularly strong when it co-occurs with a concurring gestural cue to pronoun reference (see also Arnold, Altmann et al. 2004) as the two cues would converge on the same interpretation, thereby strengthening it. We would also expect that listeners would go against order-of-mention when faced with a conflicting gestural cue to reference; order-of-mention is a cue that is known to be rather easily overridden. The degree of this shift is difficult to predict, however. Coreferential gesture is a cue that it is not always available to listeners, although they are more common than one might think. One study found that on average, 35% of the gestures speakers produced were associated with character-specific locations (So et al. 2009). On this measure, we might expect gesture to influence interpretation only weakly, at least on its own, as listeners have had relatively fewer opportunities to learn to use it (as compared to a cue that is necessarily present in every sentence). However, unlike order-of-mention which is only a tendency in the language – not all pronouns refer to first-mentioned entities – it is probably the case that coreferential gestural cues are consistent in terms of how they relate to reference when they occur. That is, we assume that speakers are consistent in their spatial mapping such that space associated with one character is not used later when speaking about a different character.[5]

5. Spatial consistency at the level of the speaker might be true only if we consider relative rather than absolute space. A speaker might, for instance, start talking while standing one way, but shift her weight or move such that, although the side associated with each character remains the same (e.g. left vs. right), the absolute location associated with one character becomes associated with a different character (e.g. Kimbara 2007).

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Thus, coreferential gesture is highly informative when it occurs. On this measure, coreferential gesture might be expected to strongly influence interpretation and so completely override the order-of-mention tendency. Thus the results of our study have the potential to inform us about the strength of gestural cues relative to other cues during interpretation.

Results from the gesture literature paint a slightly different picture of what we might expect. Numerous studies have demonstrated that when information in gesture and speech conflict, performance suffers. Kelly et al. (2010), for example, found that while adult listeners were faster and more accurate at responding to action primes when the speech and gesture contained the same information (e.g. a 'chopping' gesture produced with the word *chop*), accuracy on the task decreased as the information in speech and gesture became increasingly incongruent. Predictions with respect to cases where the gesture accords with the order-of-mention tendency align in the two cases—both the pronoun and gesture literatures would predict increased first-mentioned responding. When gesture conflicts with the speech internal cues, in contrast, we would expect performance to suffer, possibly by making it difficult to assign any interpretation to the pronoun. However, we want to be careful not to pit the predictions from the two literatures against each other too strongly. The studies examining gesture-speech conflict have examined cases where the meaning inherent to the word conflicted with the meaning (usually iconic) conveyed by the gesture. It is not clear that the pronouns and coreferential gestures in our study can be viewed this way—order-of-mention is not part of the meaning of the pronoun, rather, it is a tendency that exists in speech (that listeners make use of). Likewise with the gestures: their meaning is contextually constructed, not inherent in the individual gestures themselves. Thus, we might not expect the same kind of result as other gesture-speech conflict studies where the inherent meanings encoded by the word and gesture conflict.

2. Experiment 1

To investigate the impact of gesture on pronoun interpretation we showed participants short narrations containing multiple referents and ambiguous pronouns, and then probed pronoun interpretation. In some of the vignettes the narrator gestured using naturalistic co-speech gestures. As the narrator mentioned each character, she produced a gesture that localized the reference in space, space which might now be linked or associated with the referent, and thus, potentially available for later references to the same character. This is much like pronominal reference or agreement in signed languages (Padden 1988), although not at all systematic or conventionalized as in actual signed languages. Sometimes the locations of the gestures were consistent with the speech-internal cue (order-of-mention), and other times the two cues conflicted.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

If listeners are able to extract and interpret localization information from the gestures, we would expect them to treat this information in a similar manner as other probabilistic cues to pronoun interpretation. That is, in the absence of an additional cue (in this case, those seeing no gesture) we would expect listeners to show a general tendency to interpret the pronoun as the first mentioned referent. In contrast, we would expect those exposed to gestures contradicting order-of-mention to show more second-mentioned responses than other participants. It is also possible that participants who were exposed to gestures that reinforced order-of-mention might show higher first-mentioned response rates than those who saw no gestures. If listeners do not extract and interpret localization information from co-speech gesture we would expect the responses to be similar across all three conditions, as the narrations themselves were identical, and thus provided the same verbal cues to pronoun interpretation.

2.1. *Method*

2.1.1. *Participants.* Thirty undergraduate students (18 female) from the University of California, Berkeley took part in this experiment. All were native English speakers (defined as having learned English before age 5), although many were bilingual. Crucially, none of the participants had studied any signed languages. Participants received course credit for participating.

2.1.2. *Materials and procedure.* Participants were seated in front of a video monitor on which they watched eight video clips of a narrator (the first author) telling stories, each lasting approximately 15 s. The narrator was shown from the waist up; her hands and face were clearly visible. The narrator is an experienced gesture coder, and while the speech and gesture were more characteristic of story telling than conversation (e.g. the prosody of the speech was slightly exaggerated, and the gestures were slightly larger), every effort was made to ensure the stimuli were as naturalistic as possible. Five of the eight vignettes were experimental items, each containing an ambiguous pronoun. These stories contained two characters of the same gender. The characters were first introduced, then mentioned again by name before the narrator described an action performed by only one of the characters using a pronoun.

Importantly, aside from order-of-mention, other speech internal cues to pronoun interpretation were not present in the experimental narrations; as mentioned, both characters were of the same gender, neither character was introduced as the sole grammatical subject, and the case of the ambiguous pronoun either matched the case used previously for both of the participants or neither participant. (The latter was true of one item, where the ambiguous pronoun was a possessive.) Thus, the only potential cues to pronoun interpretation present in the stimuli were order-of-mention and the gestures. Participants also

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Table 1. *Sample stimuli*

Condition	Narrative
Ambiguous pronoun	Annie and Sarah are having a picnic in the park. They have a lot of food with them. Annie is carrying the picnic basket, and Sarah has a blanket to sit on. She's excited about the cookies.
Unambiguous pronoun	Bobby gave Andrea a new CD for her birthday. When they listen to the CD, Bobby sings, and Andrea dances

heard three filler stories with two characters but without ambiguous pronouns, in an effort to prevent them from discerning what we were investigating. Example stimuli are presented in Table 1.

Following each narration, the image paused and there was a voice-over with three questions. For the experimental items, one of the questions asked about the identity of the referent of the ambiguous pronoun. For instance, the question for the narrative presented in Table 1 was *Who is excited about the cookies?* The remaining questions served as distractors and asked about minor details from the story, for example, *What are Annie and Sarah doing?* and *Who is carrying the picnic basket?* The order of the questions was rotated for each narration, so that the question related to the ambiguous pronoun alternated between being the first, second, or last question asked. Participants were instructed to listen carefully and to answer each question aloud as it was asked.

After listening to all eight narrations, participants were asked what they thought the study was investigating. This allowed us to see whether participants noticed anything unusual about the gestures used in the study. The most typical responses were "memory for details" and "differences between adults' and children's memory." (The consent form described our study as part of a larger project examining differences between children and adults in language comprehension.) Very few participants made any comments about the gestures.

A digital audio recorder was used to record responses. Each session lasted approximately 15 minutes.

2.1.3. *Experimental manipulation.* Participants were randomly assigned to one of three conditions.[6] All participants, regardless of condition, heard the same narrations and questions. What varied was the presence/location of

6. We used a between-subjects design because we are also interested in children's ability to use gestures to interpret pronouns. It would have been difficult to collect enough data of each type (OoM, AOoM, NG) from young children in a within-subjects design—the study would simply be too long for them. To facilitate comparison of child patterns to adult patterns, the studies needed to be as similar as possible. Thus, we were somewhat constrained in our design.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

A B C

Figure 1. *Examples of (A) localizing gesture 1, (B) localizing gesture 2, and (C) coreferential gesture.*

gestures that co-occurred with the ambiguous pronoun. In the Order-of-Mention (OoM) condition, the narrator produced gestures that matched the order-of-mention for the noun phrases. For example, when introducing the first character of the story, the narrator produced a localizing gesture to her right side, for instance, by placing her right hand palm down on the left side of her body, while saying the name of the character. When introducing the second character, the narrator produced a gesture on her left side with her left hand. When producing the ambiguous pronoun, the narrator gestured back towards the space where she had initially localized the first character (the right side). An example of this is shown in Figure 1.[7]

Four of the five narrations used similar gestures (i.e. flat palms placed to the narrator's left and right sides), while one narration made use of space in a slightly more complicated manner. This narration described Donald being at the bottom of a hill, with one hand placed in the same location as the other gestures, and Mickey at the top of the hill, with the hand placed at the narrator's head level. When the narrator said *Donald is walking up the hill*, she produced a diagonal climbing gesture with two fingers 'walking' up the hill. An items analysis (reported with the results) suggests that participants did not interpret this narration differently from the other four.

The Against-Order-of-Mention (AOoM) condition was similar, except that when producing the ambiguous pronoun the narrator gestured toward the location that went with the second-mentioned character. That is, if the first-mentioned character was initially localized on the left side, the narrator produced a gesture on her right side (the side previously associated with the second-mentioned character) in conjunction with the ambiguous pronoun.

7. These pictures show the end-points of dynamic gestures.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Finally, the No-Gesture (NG) condition was used to establish the baseline rate of order-of-mention interpretations in response to our stimuli. In these narrations the narrator kept her hands in her lap while speaking.

2.1.4. *Coding.* Responses to the experimental questions were coded as either identifying the first-mentioned character as the referent for an ambiguous pronoun, and so following the order-of-mention tendency, or identifying the second-mentioned character. Occasionally, participants did not unambiguously identify either character as the referent of the pronoun. Most often these were refusals to answer (e.g. 'I don't know'), but sometimes participants responded that both participants performed the action in question. These responses were classified as *other*.

2.2. *Results and discussion*

Figure 2 shows the percentage of participant responses that fell into the following categories: *first-mentioned character*, *second-mentioned character*, and *other*. Error bars represent standard error rates. Recall that the NG condition was intended to establish the base rate of order-of-mention type of responses to our stimuli. As expected, participants were consistent with other studies examining order-of-mention (e.g. Arnold et al. 2000), and interpreted the ambiguous pronoun as referring to the first-mentioned character on average 68% of the time (with one participant selecting the first mentioned character 20% of the time, one at 40%, and the remaining at least 60% of the time). Thus, on average, participants demonstrated a first-mentioned bias despite the differences in the current stimuli from previous studies examining order-of-mention (i.e. rather than reading or hearing sentences describing pictures, participants watched videos of a narrator telling a story).

With this established, we performed a one-way ANOVA comparing the number of first-mentioned responses across the three conditions both by

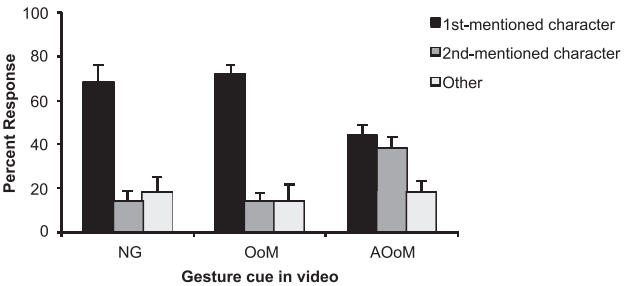

Figure 2. *Mean responses by type and condition.*

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

participant (F_1) and by item (F_2). The effect of condition was marginally significant by participant (F_1 (2, 27) = 2.89, p = .07), and significant in the item analysis (F_2 (2, 12) = 12.45, p = 0.001). We then went on to compare the conditions to each other. These were the tests of primary concern, as we were interested in whether specific types of gestures changed interpretations relative to the NG condition.

We had predicted that if participants were using the information from the localizing and co-referring gestures to interpret pronouns, participants in the AOoM condition should show fewer first-mentioned responses than other participants. This was indeed the case: Participants in the AOoM condition interpreted the ambiguous pronoun as referring to the first-mentioned character less than half of the time (44%, SD = 22), significantly less often than participants in the OoM condition (M = 72%, SD = 21; $t_1(18) = 2.83, p = .011$; $t_2(8) = 4.51$, $p = 0.001$) and the NG condition (M = 68%, SD = 27; $t_1(18) = -2.23$, $p = .03$; $t_2(8) = 4$, $p = .003$) . We also anticipated that participants in the OoM condition, who saw co-referring gestures that accorded with order-of-mention tendencies, might produce more first-mention responses than participants in the NG condition; however, this was not the case. Although they produced slightly more first-mention responses than the NG participants, the difference between the two groups is not significant ($t_1(18) = .381$, $p = .70$; $t_2(8) = 0.73$, $p = .48$).

We also hypothesized that participants in the AOoM condition would interpret the pronoun as referring to the second-mentioned character more often than participants in the other two groups. As there are three possible answer types, this cannot simply be inferred from the previous analysis. A one-way ANOVA comparing the number of second-mention responses across the three conditions was significantly different by participant (F_1 (2, 27) = 7.53, p = .03), and by item (F_2 (2, 12) = 25.22, p = 0.01). As in our analysis of the first-mention responses, we then went on to compare the conditions to each other. Participants in the AOoM condition selected the second-mentioned character 38% of time (SD = 23), which is significantly more than in the OoM (M = 14%, SD = 13; $t_1(18) = 3.4$, $p = .03$; t_2 (8) = 5.37, $p < .001$) and NG conditions (M = 14%, SD = 14; $t_1(18) = 3.4$, $p = 0.03$; $t_2(8) = 7.55$, $p < .001$). Consistent with the first-mention results, there was no difference between the OoM and NG conditions in terms of second-mention responses ($t_1(18) = -0.342$, $p = 0.73$; t_2 (8) = 0.364, $p = 0.72$).

Turning to the other responses, there was no significant difference across the three groups ($F_1(2, 27) = 0.12$, $p = 0.78$; F_2 (2, 12) = 0.44, $p = 0.65$). This is unsurprising, given how few other responses there were overall (as is evident in Figure 2). However, it is still possible that such responses were distributed differently in the three groups, so we also examined whether the number of participants who produced any such responses varied by condition. Five out of ten participants in the NG condition produced at least one such response, com-

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Table 2. *Number of participants selecting first-mentioned character by item.*

Condition	Item number				
	1	2	3	4	5
NG	9	7	6	7	6
OoM	10	7	9	6	7
AOoM	6	4	3	3	2

pared to seven participants in the AOoM condition and four in the OoM condition. This may suggest that participants were more confused when faced with gestures that went against the first-mentioned bias, due to a conflict in the interpretations suggested by the two sources of information. However, the difference between the groups was not significant (χ^2 (df = 2, N = 16) = .87, p = .64).

To explore the possibility that participants in the AOoM condition were learning to use the gestures as a cue to reference, we examined responses to the first experimental question only.[8] The percentages of participants that interpreted the ambiguous pronoun as referring to the first-mentioned character in the first narrative were 90% (NG), 100% (OoM), and 60% (AooM). Thus, the same patterns are evident in these data, and in fact, the differences between the AOoM group and the other two groups are even larger when only the first experimental narrative is examined.

We also examined the distribution of participant responses over time. Table 2 shows the number of participants that selected the first-mentioned character for each experimental question in each condition. If one looks only at the data from the AOoM condition, it does appear that they are learning to use the gestures, as they give fewer first-mention responses over time. However, this same trend is apparent in all three conditions, the first-mention tendency is stronger in the earliest items. Reflecting this overall trend, a chi-square analysis approached significance in the overall analysis (χ^2 (df = 4, N = 150) = 8.60, p = .072). However, the number of participants giving first-mention responses was not significantly different across the items in any individual condition. It is not clear why the effect is stronger initially. It could be something about the particular stories, or it could be that participants were following more general tendencies of interpretation early on, and second-guessing themselves or changing strategies in response to being in an experiment as time went on. It is difficult to say on the basis of these results alone. Whatever the case, however, these data suggest that participants in the AOoM condition were not

8. Note that participants could only learn to use the gestures as a cue in a general sense: they could not learn anything about the relative reliability of the cues as they were never privy to the true or intended mapping between pronoun and antecedent.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

learning to use the gestures in this study. Rather, the ability (and tendency) to use gestures as a cue to pronoun interpretation was something they brought with them.

As mentioned previously, four of the five narrations used similar gestures (i.e. flat palms placed to the narrator's left and right sides), while one narration made use of space in a slightly more complicated manner. We conducted one further analysis to ensure that this item was not interpreted differently from the others. Specifically, we compared responses for the two kinds of items (more static vs. more dynamic gestures) across conditions and found no significant difference according to the kind of gesture used (OoM condition: $t\,(18) = 0.37$, $p = 0.71$; AoM condition: $t\,(18) = 1.22$, $p = 0.24$). Note, however, that this was a post-hoc analysis, and we did not design the study to assess the impact of different kinds of gesture. The narrator simply produced the gestures that were most natural to her for the given story.

Thus, it seems that the gestures are influencing people's interpretation of the ambiguous pronoun. When people saw gestures that indicated that the second-mentioned character was the intended referent, they were more likely to interpret the pronoun as the second-mentioned character and in so doing overcome their first-mentioned bias, at least to some degree. However, contrary to our expectations the presence of a gesture consistent with the first-mentioned bias did not increase participants' tendency to interpret the pronoun as the first-mentioned character.

3. Experiment 2

Although the narrator made every attempt to keep the narrations as similar as possible in Experiment 1, we also wanted the narrations to sound naturalistic and therefore did not manipulate the speech or video after recording. It is possible that the speech itself differed as the gestures differed, and that it was these speech differences, rather than the gestures, that influenced interpretation. Prosodic effects on language processing are well known (Cutler et al. 1997). Of particular relevance, stress or accent differences can shift interpretation such that something that would normally be dispreferred as the referent of a pronoun (Akmajian and Jackendoff 1970; Maratsos 1973), or other anaphoric device (Dahan et al. 2002) instead becomes the preferred referent. Thus, it was important to rule out the possibility that our results were driven by prosodic cues rather than gesture as we suggest. To investigate this we ran a second experiment in which we exposed listeners to the audio portion only of the stimuli used in Experiment 1. If participants' responses were driven by the narrator's speech as opposed to her gesture, we would expect the responses to be roughly the same across the three conditions. However, if participants were influenced by the gesture in Experiment 1, we would expect differences between responses

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

in the OoM and AOoM conditions in Experiment 1 and Experiment 2, but not in the NG condition.

3.1. *Method*

3.1.1. *Participants.* Thirty additional undergraduate students (20 female) from the University of California, Berkeley took part in this experiment for course credit. As in Experiment 1, all were native English speakers, with no sign language experience.

3.1.2. *Materials and procedure.* This experiment was identical to Experiment 1 in every way except that the computer monitor was turned off so that participants could only hear the audio version.

3.2. *Results*

Figure 3 shows the percentage of participant responses that fell into each of the three categories: first-mentioned character, second-mentioned character, and other. One-way ANOVAs comparing the three conditions were not significant for any response type (first-mentioned responses: F_1 (2,27) = 1.64, p = 0.21; F_2 (2,12) = 1.02, p = .39; second-mentioned responses: F_1 (2,27) = 1.93, p = 0.18, F_2 (2,12) = 1.04, p = 0.37; other: F_1 (2,27) = 0.75, p = 0.92, F_2(2, 12) = 0.21, p = 0.81), so we did not compare the conditions to each other.

As Experiment 2 was aimed at discerning any unintended differences in the audio version across conditions in Experiment 1, that is, to ensure that any differences we found in Experiment 1 were due to the gestures, we also compared the results for each condition in Experiment 1 with the results from the same 'condition' in Experiment 2. This was a second check on the impact of the gestures, directly comparing participants exposed to the same audio information. Figure 4a shows the percent of first-mentioned character responses for

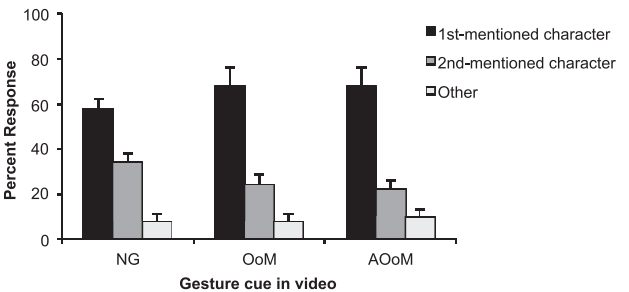

Figure 3. *Mean first- and second-mentioned character responses by condition, audio only.*

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

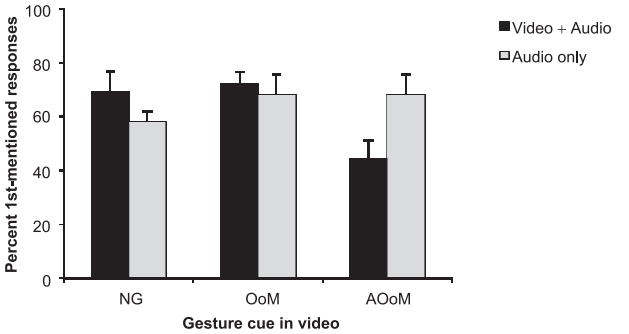

Figure 4a. *Experiments 1 and 2: Mean first-character responses by condition.*

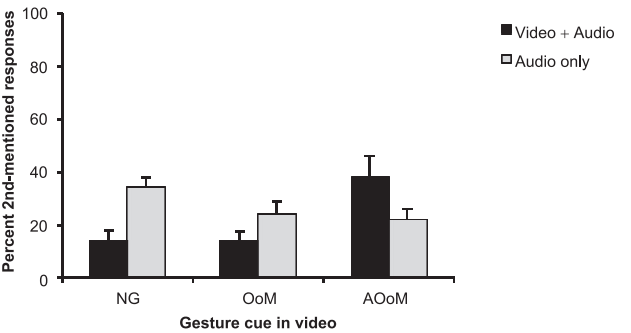

Figure 4b. *Experiments 1 and 2: Mean second-character responses by condition.*

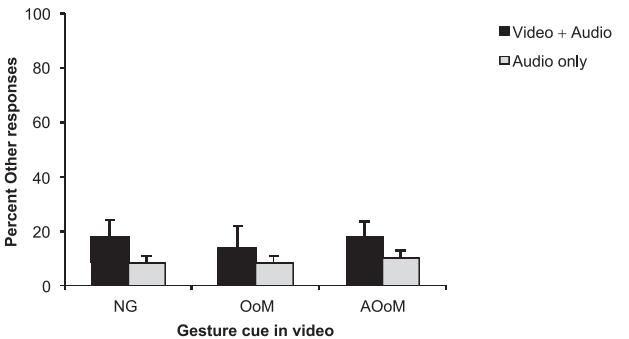

Figure 4c. *Experiments 1 and 2: Mean other responses by condition.*

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Experiment 1 and Experiment 2, Figure 4b shows the second-mentioned character responses, and Figure 4c shows the other responses. Looking first at first-mentioned responses, participants in the NG condition who saw the video in Experiment 1 did not respond significantly differently from participants in Experiment 2 who only heard the audio ($t_1(18) = 1.02$, $p = 0.31$; $t_2(8) = 0.89$, $p = 0.40$). Likewise, there was no significant difference between the two OoM conditions ($t_1(18) = 0.49$, $p = 0.62$; $t_2(8) = 0.28$, $p = 0.78$). However, there was a significant difference in first-mentioned responses in the AOoM conditions ($t_1(18) = -3.04$, $p = 0.007$; $t_2(8) = 4.47$, $p = 0.002$).

Turning to the second-mentioned character responses, again there was no significant difference between Experiment 1 and Experiment 2 in the OoM condition ($t_1(18) = 1.38$, $p = 0.18$, $t_2(8) = 0.68$, $p = 0.54$). There was a significant difference across the two experiments in the NG condition by participant ($t_1(18) = 3.31$, $p = 0.04$), but not by item ($t_2(8) = 0.89$, $p = 0.39$). In the AOoM condition the difference between the two experiments was significant in both analyses ($t_1(18) = 2.07$, $p = .05$; $t_2(8) = 6.33$, $p < .001$). Finally, there were no significant differences between the number of other responses in the two experiments across the three conditions (OoM: $t_1(18) = 0.70$, $p = 0.49$), $t_2(8) = 0.0$, $p = 1$; AOoM: $t_1(18) = 1.1$, $p = 0.27$, $t_2(8) = 6.48$; $p = 0.179$; NG: $t_1(18) = 1.41$, $p = .175$, $t_2(8) = 1.44$, $p = 0.186$).

If participant responses in Experiment 1 were due to speech-internal cues such as stress, rather than the gestures as we intended, we would not expect to see differences between the responses in Experiment 1 and Experiment 2. However, if the gestures were what was influencing responses in Experiment 1, we would expect differences between the two experiments, since the gesture cue was removed in Experiment 2. Specifically, we would expect differences in the AOoM condition, as this was where the effect of gesture was seen in Experiment 1. Results showed that there were no differences between the OoM condition in first-mention or second-mention responses, or in the NG condition in first-mention responses. Somewhat unexpectedly, the second-mentioned responses in the NG conditions were significantly different in the two experiments, although only in the by-participant analysis. This is surprising, given the fact that in the video there were no additional cues (at least, not intentionally) to pronoun reference in the NG condition. This difference may be due in part to differences in the number of other responses in the two experiments for the NG condition as there were slightly more in Experiment 1 than in Experiment 2, although recall that this difference was not significant. It is not clear what additional cue might have been present in the video in the NG condition that could have led to this difference, however. It seems likely to us that this is just an accidental finding due to individual differences.

Importantly, and most relevant to the question we set out to investigate, we did find a significant difference between both first- and second-mentioned

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

responses in the AOoM condition, suggesting that participants were more likely to go against the first-mention bias when they saw an alternative referent suggested in the narrator's gesture. The difference between responses in the AOoM condition of Experiment 1 and Experiment 2 fits with our expectations, suggesting that at least in this condition, it was the information contained in the gesture—not the speech—that affected participant responses in Experiment 1.

4. Discussion

We set out to investigate whether gesture can affect how the speech it accompanies is interpreted. In particular, we examined whether coreferential localizing gestures inform the interpretation of ambiguous pronouns. The data from the AOoM condition provide strong evidence that the answer is yes: despite a clear bias to interpret the pronouns in our stimuli as referring to the first-mentioned character, as demonstrated by the responses of participants in the NG condition in Experiment 1, and all conditions in Experiment 2, when people saw a gesture that went against order-of-mention, they often interpreted the pronoun as referring to the second-mentioned character. However, the impact of the gestures was not so strong as to completely overcome the first-mention bias, as participants in the AOoM condition still interpreted the pronouns as referring to the first-mentioned character about half the time. We had anticipated that the opposite would also be true, namely that when exposed to a gesture that matches order-of-mention participants would be even more likely to select the first-mentioned character. Somewhat to our surprise, seeing a gesture did not increase the proportion of first-mentioned responses in the OoM condition.

We wondered whether there was variability in participants' tendencies to notice the gestures, and whether this could explain the subtle effect. To investigate this we had another ten participants perform the same task while being eye-tracked.[9] The stimuli had a visual angle of 1.409 degrees and the gestures were typically outside of the foveal area. Clearly humans are quite able to detect motion parafoveally (Palmer 1999), and so this does not preclude participants from encoding left or right without looking directly at the narrator's hands. Nevertheless, we thought that participants' attention might be attracted by the movement of the hands (Goodwin 1986; Kendon 1990; Streeck 1993) and that overt eye movements during the presentation of the narrations might correlate with responses. Consistent with previous studies (Argyle and Cook

9. We used an ASL Pan/Tilt Model 6R remote eye tracker. Participants sat approximately 25 inches away from a computer monitor displaying the same stimuli used in Study 1. Their gaze locations were monitored while they listened to the narrations and answered the questions out loud.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

1976; Gullberg and Holmqvist 2006; Kendon 1990; Rutter 1984), we found very few actual looks to the gestures, and no relationship between looks during the narrations and responses. Interestingly, listeners frequently looked back to the spot where the localizing gesture had been produced while responding—another indication (in addition to their responses) that they had encoded something about the location of the gestures. However, the data during responding was not collected systematically, and so we are unable to analyze it further, although it opens up interesting possibilities for further study.

This is not the first work to show the influence of visual information on speech processing. The McGurk effect, for instance, is a well-known demonstration of this influence (McGurk and McDonald 1976). When shown a video of a speaker producing one phoneme (for example, /g/) with the audio of a different phoneme dubbed over the speech (/b/), participants perceive a third, intermediary phoneme (/d/). The influence of visual information on speech perception in not limited to the face, however. Munhall et al. (2004) found that being able to see head movements (such as bobbing and swaying) which are correlated with speech sound production can improve the perception of speech sounds, even in the absence of facial movements. We have demonstrated that the impact of visual information goes beyond the perceptual (i.e. speech processing) to the referential (the referent associated with a pronoun).

It is as yet unclear exactly how gesture exerts its influence. On the one hand, coreferential localizing gestures might function like the lexical features on pronouns, which serve as very explicit restrictions on possible co-reference (e.g. a masculine pronoun cannot refer to a female, any more than a gesture associated with one character's location could direct the listener to another character's identity). However, gestures are not part of the form being interpreted, and in this are quite different from linguistic artefacts such as nominal gender and number marking. Clark (2009) discusses processes by which children figure out what an interlocutor is attending to or thinking about and describes two things which make interpretation easier for the child, namely the actual presence of the object or event, or physical co-presence, and the explicit mention of the object or event, or conversational co-presence. As interpreting pronouns involves discerning the speaker's intentions, these are potentially relevant concepts. Considered from this perspective, gestures might be better viewed as cues to reference provided by physical rather than conversational co-presence, since they are not explicit mentions of (properties of) the entity being referred to, but rather, visual representations of the entity. Alternatively, the influence of gestures might be more indirect. A great deal of work has shown how cues such as grammatical function (Gordon et al. 1993), emphatic stress (Maratsos 1973), and order-of-mention (Arnold et al. 2000; Gernsbacher and Hargreaves 1988; McWhinney 1977) make some referents more accessible than others, and in turn more accessible referents are preferentially selected by listeners as the

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

referent of the pronoun. Although the cues just mentioned are all within the speech itself, the expectancy hypothesis of Arnold and colleagues (Arnold 1998, 2001, 2008; Arnold, Altmann et al. 2004) explicitly allows for cues outside of the speech to influence expectancy as well, as they can also provide cues as to the speaker's intentions. The gestures we studied here serve this function, providing a spatial cue for the intended referent.

One question that immediately arises, however, is why, if the gestures are such a direct link with what the speaker is thinking, the information contained in the gestures did not deterministically guide interpretation. The answer to this, we suggest, lies in the amount of experience listeners have relying on these specific gestures. Many of the cues listeners are known to rely on for pronoun interpretation are based on patterns in their input that they have had years to observe. Indeed, young children do not initially use the same cues as adults—they must learn to do so (Arnold, Brown-Schmidt et al. 2004; Arnold, Brown-Schmidt et al. 2007). Since speakers do not always produce these sorts of gestures alongside ambiguous pronouns (So et al. 2009), even adult listeners may not have had enough experience with coreferential gestures to weight them as heavily as cues known to be reliable from years of experience.

In any experimental design there is always a difficult balance to maintain between controlled methodology and realistic stimuli, and our study is no different. As noted earlier, the narrator's gestures were rather exaggerated and it is possible that participants were more aware of the use of spatial localization than they might have been under naturalistic conditions, and so our effect is larger than it would otherwise be. However, we are confident that the salience of the gestures in our stimuli was not the underlying cause of our findings. Our eye-tracking results suggest that participants were not actively directing their gaze to the gestures, as we might expect if the gestures were noticeably odd. Moreover, none of the participants reported noticing the gestures in the post-experiment debriefing interview. Thus, we feel that our findings represent the ability of listeners to integrate information presented through a visual modality with information presented in speech. It may also be relevant that we used story-telling stimuli rather than interactive conversation. Again, this was primarily due to our desire to maintain consistency across participants; interactive conversation would have necessarily led to inter-participant variation in the gestures experienced by the listener. However, it is not clear why story-telling would lead participants to pay more attention to the gestures than conversation. Indeed, one can just as easily propose that it might lead them to pay less attention to the gestures, by being less engaged in the story than they might be during a live interaction, for instance. The possible differences between sensitivity to gesture in interactive and non-interactive situations is a very interesting question, but one that is beyond the scope of the present study. Thus, while gestures might have more or less of an impact for the listener in conversation

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

than in story-telling, our main point remains, that they do have an impact. It is admittedly possible (although we feel unlikely) that the narrations may have also differed in other visual cues, such as the speaker's eye gaze or facial expressions. We do not, however, have any explicit predictions regarding how these other visual factors might influence the OoM tendency, other than to predict that a shift in speaker eye-gaze or a head tilt toward the gestures might additionally serve to highlight them. But the gestures would still be playing the main role, since these unintentional cues would presumably signal the listener to attend to the information conveyed in the hand movements themselves.[10]

The current work then adds to previous research demonstrating that speakers produce coreferential localizing gestures while referring to multiple referents (e.g. So et al. 2005) by showing that listeners are sensitive to these gestures. More broadly, it extends previous work demonstrating that listeners can integrate information from speech and gesture to form a unified interpretation (e.g. Cassell et al. 1999; Singer and Goldin-Meadow 2005), by showing that the information from gesture is not simply additive. That is, gesture can do more than improve comprehension when it accords with speech-internal information or impair comprehension when it conflicts with that information; it can also affect the way the speech itself is interpreted. Moreover, it does so in ways similar to other speech internal cues listeners are known to use when the intended meaning of the speech is ambiguous.

References

Akmajian, A. & R. Jackendoff. 1970. Coreferentiality and stress. *Linguistic Inquiry* 1. 124–126.

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

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

Arnold, J. E. 1998. *Reference form and discourse patterns*. Stanford, CA: Stanford University dissertation.

Arnold, J. E. 2001. The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* 31. 137–162.

Arnold, J. E. 2008. Reference production: Production-internal and Addressee-oriented Processes. *Language and Cognitive Processes* 23. 495–527.

Arnold, J. E. 2010. How speakers refer: the role of accessibility. *Language and Linguistic Compass* 4. 187–203.

Arnold, J. E., R. Altmann, M. Fagnano & M. K. Tanenhaus. 2004. The Old and thee, uh, new. *Psychological Science* 9. 578–582.

Arnold, J. E., S. Brown-Schmidt & J. Trueswell. 2007. Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes* 22. 527–565.

10. Unless, of course, the speaker was tilting her head to one side for one character and the other side for the other character, which would also be indexically informative, but this does not appear to be the case in our stimuli.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Arnold, J. E., S. Brown-Schmidt, J. Trueswell & M. Fagnano. 2004. Children's use of gender and order-of-mention during pronoun comprehension. In J. Trueswell & M. Tanenhaus (eds.), *Approaches to studying world-situated language use. Bridging the language-as-product and language-as-action traditions,* 261–282. Cambridge, MA: MIT Press.

Arnold, J. E., J. Eisenband, S. Brown-Schmidt & J. Trueswell. 2000. The rapid use of gender information: Evidence of the time course of pronoun resolution from eye tracking. *Cognition* 76. B13–B26.

Arnold, J. E., C. Hudson Kam & M. K. Tanenhaus. 2007. If you say thee uh- you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition* 33. 914–930.

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

Beattie G. W. & H. K. Shovelton. 1999. Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica* 123. 1–2.

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

Butterworth, B. & U. Hadar. 1989. Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review* 96. 168–174.

Campana, E., L. Silverman, M. K. Tanenhaus, L. Bennetto & S. Packard. 2005. Real-time integration of gesture and speech during reference resolution. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*. 378–383.

Cassell, J., D. McNeill & K. E. McCullough. 1999. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics and Cognition* 7. 1–33.

Chambers, C. G. & R. Smyth. 1998. Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language* 39. 593–608.

Clark, E. V. 2009. *First language acquisition*. Cambridge: Cambridge University Press.

Crawley, R. A., R. J. Stevenson & D. Kleinman. 1990. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* 194. 245–265.

Cutler, A., D. Dahan & W. van Donsellar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40. 141–201.

Dahan, D., M. K. Tanenhaus & C. G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47. 292–314.

Foraker, S. 2011. Gesture and discourse: How we use our hands to introduce and refer back. In G. Stam, M. Ishino & R. Ashley (eds.), *Integrating gestures*, 279–292. Amsterdam: John Benjamins.

Foraker, S. & S. Goldin-Meadow. 2007. Gesture and discourse: How we use our hands to refer back. Poster presented at 20th Annual CUNY Conference on Human Sentence Processing, San Diego, CA. March.

Gernsbacher, M. A. & D. J. Hargreaves. 1988. Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language* 27. 699–717.

Goldin-Meadow, S. 2003. *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language.* New York: Psychology Press.

Goodrich, W. & C. L. Hudson Kam. 2009. Co-speech gesture as input in verb learning. *Developmental Science* 12. 81–87.

Goodwin, C. 1986. Gestures as a resource for the organization of mutual orientation. *Semiotica* 62. 29–49.

Gordon, P. C., B. J. Grosz & L. A. Gilliom. 1993. Pronouns, names and the centering of attention in discourse. *Cognitive Science* 17. 311–347.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Grosz, B. J., A. K. Joshi & S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21. 203–226.
Gullberg, M. & K. Holmqvist. 2006. What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition* 14. 53–82.
Hadar, U. & R. Krauss. 1999. Iconic gestures: the grammatical categories of lexical affiliates. *Journal of Neurolinguistics* 12. 1–12.
Järvikivi, J., R. P. van Gompel, J. Hyönä & R. Bertram. 2005. Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science* 16. 260–264.
Kehler, A. 2002. *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.
Kelly, S. D., C. Kravitz & M. Hopkins. 2004. Neural correlates of bimodal speech and gesture comprehension. *Brain and Language* 89. 253–60.
Kelly, S. D., A. Özyürek & E. Maris. 2010. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21. 260–267.
Kendon, A. 1980. Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (ed.), *Relationship of verbal and nonverbal communication*, 207–228. Berlin: Mouton de Gruyter.
Kendon, A. 1981. Geography of gesture. *Semiotica* 37. 129–163.
Kendon, A. 1990. *Conducting interaction.* Cambridge: Cambridge University Press.
Kendon, A. 1994. Do gestures communicate? A review. *Research on Language and Social Interaction* 27. 175–200.
Kimbara, I. 2007. Indexing locations in gesture: Recalled stimulus image and interspeaker coordination as factors influence gesture form. In S. Duncan, J. Cassel & E. Levy (eds.), *Gesture and the dynamic dimension of language*, 213–220. Amsterdam: John Benjamins.
Kousta, S. T. 2008. *A constraint-based approach to pronoun resolution. University of Cambridge Research Centre for English and Applied Linguistics Working Papers* 12. 33–69.
Krauss, R. M. 1998. Why do we gesture when we speak? *Current Directions in Psychological Science* 7. 54–60.
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.
Krauss, R. M., P. Morrel-Samuels & C. Colasante. 1991. Do conversational hand gestures communicate? *Journal of Personality and Social Psychology* 61. 743–754.
MacWhinney, B. 1977. Starting points. *Language* 53. 152–168.
MacWhinney, B., E. Bates & R. Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23. 127–150.
Maratsos, M. P. 1973. The effects of stress on the understanding of pronominal coreference in children. *Journal of Psycholinguistic Research* 2. 1–8.
McGurk, H. & J. MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748.
McNeill, D. 1992. *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.
McNeill, D. 2000. (ed.). *Language and gesture: Window into thought and action.* Cambridge: Cambridge University Press.
Munhall, K. G., J. A. Jones, D. E. Callan, T. Kuratate & E. Vatikiotis-Bateson. 2004. Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science* 15. 133–137.
Padden, C. 1988. *Interaction of morphology and syntax in American Sign Language.* NY: Garland.
Palmer, S. 1999. *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press

Rogers, W. T. 1978. The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research* 5. 54–62.
Rutter, D. R. 1984. *Looking and seeing: The role of visual communication in social interaction*. Chichester: Wileys.
Singer, M. A. & S. Goldin-Meadow. 2005. Children learn when their teacher's gestures and speech differ. *Psychological Science* 16. 85–89.
So, W. C., M. Coppola, V. Licciardello & S. Goldin-Meadow. 2005. The seeds of spatial grammar in the manual modality. *Cognitive Science* 29. 1029–1043.
So, W. C., S. Kita & S. Goldin-Meadow. 2009. Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science* 33. 115–125.
Streeck, J. 1993. Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs* 60. 275–299.
Tfouni, L. & R. Klatzky. 1983. A discourse analysis of deixis: Pragmatic cognitive, and semantic factors in the comprehension of 'this', 'that', 'here', and 'there'. *Journal of Child Language* 10. 123–133.
Thompson, L. A. & D. W. Massaro. 1994. Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology* 57. 327–354.
Willems, R. M., A. Özyürek & P. Hagoort. 2007. When language meets action: The neural integration of gesture and speech. *Cerebral Cortex* 17. 2322–2333.
Wu, Y. C. & S. Coulson. 2007. How iconic gestures enhance communication: An ERP study. *Brain and Language* 101. 234–245.

https://doi.org/10.1515/langcog-2012-0005 Published online by Cambridge University Press