



Construing events first-hand: Gesture viewpoints interact with speech to shape the attribution and memory of agency

Dana Michelle Chan¹ · Spencer Kelly¹

Accepted: 21 December 2020 / Published online: 7 January 2021
© The Psychonomic Society, Inc. 2021

Abstract

Beyond conveying objective content about objects and actions, what can co-speech iconic gestures reveal about a speaker's subjective relationship to that content? The present study explores this question by investigating how gesture viewpoints can inform a listener's construal of a speaker's agency. Forty native English speakers watched videos of an actor uttering sentences with different viewpoints—that of low agency or high agency—conveyed through both speech and gesture. Participants were asked to (1) rate the speaker's responsibility for the action described in each video (encoding task) and (2) complete a surprise memory test of the spoken sentences (recall task). For the encoding task, participants rated responsibility near ceiling when agency in speech was high, with a slight dip when accompanied by gestures of low agency. When agency in speech was low, responsibility ratings were raised markedly when accompanied by gestures of high agency. In the recall task, participants produced more incorrect recall of spoken agency when the viewpoints expressed through speech and gesture were inconsistent with one another. Our findings suggest that, beyond conveying objective content, co-speech iconic gestures can also guide listeners in gauging a speaker's agentic relationship to actions and events.

Keywords Language comprehension · Modality effects · Co-speech gesture · Cued recall

Research on the comprehension of co-speech iconic gestures—gestures representing object attributes, actions, and spatial relations—is built on the foundation that they describe concrete information about the world (Goldin-Meadow, 2005; Kendon, 1986; McNeill, 1985). Indeed, gestural descriptions of objects and actions can greatly influence listeners' interpretations of a speaker's meaning (for reviews, see Church et al., 2017; Hostetter, 2011). But in addition to objectively describing things in the physical world, might gestures also provide a glimpse into a speaker's *subjective relationship* to that information? We explore this question in the context of how viewpoints associated with co-speech iconic gestures shift construal of “who did what” with objects and actions.

It has been well documented over the past 3 decades that a speaker's iconic gestures add meaningful information to speech during language production (for more on gesture's

function in producing language, see Church et al., 2017). Careful observations of spontaneously produced gestures have revealed that speakers use their hands to depict relevant semantic information about objects and actions related to the accompanying speech, supporting the theory that gesture and speech share the same computational processes (McNeill, 1985, 1992). Bolstering this theory are studies showing that speakers' mental representations of objects and actions are most fully revealed only through a combination of gesture and speech, as gestures often add relevant and complementary information to what is verbally expressed (Debreslioska & Gullberg, 2017; Goldin-Meadow, 2005; Hostetter & Alibali, 2008, 2019; Kita & Özyürek, 2003). If gestures reveal such pertinent information about a speaker's knowledge of object and action attributes, are listeners also perceptive of these cues?

People do in fact incorporate the semantic content of gestures when building their own account of events, as confirmed by studies demonstrating that listeners integrate aspects of a speaker's iconic gestures into their own spoken and written recall of previously heard descriptions, explanations, and narratives (Beattie & Shovelton, 1999a, 1999b; Goldin-Meadow et al., 1992; Kelly et al., 1999; So et al., 2012). Additionally, the content contained in co-speech iconic gestures influences

✉ Dana Michelle Chan
dchan@u.northwestern.edu

¹ Department of Psychological and Brain Sciences, Neuroscience Program, Colgate University, 13 Oak Dr, Hamilton, New York 13346-1338, USA

the online processing of accompanying speech during language comprehension (Holle & Gunter, 2007; Kelly et al., 2010; Özyürek et al., 2007; Wu & Coulson, 2007). When speech and iconic gestures both convey similar content, memory traces for an utterance can even be strengthened (Cohen, 1989; Engelkamp, 1998). This “enactment effect” occurs not just when someone produces iconic gestures themselves (Engelkamp et al., 1994; Russ et al., 2003; Stevanoni & Salmon, 2005), but also when they view the gestures of others (Cutica & Bucciarelli, 2008; Feyereisen, 2006, 2009; Iani & Bucciarelli, 2017; Kelly et al., 2009; So et al., 2012). Such evidence points to the integration of iconic gesture and speech in language comprehension and memory, with listeners combining information from both channels to fully inform their understanding and memory of narrated events, descriptions of objects, and performance of actions.

Whereas many studies have focused on how iconic gestures objectively capture features of the external world, gesture can also reveal a speaker’s subjective relationship to that world (Debreslioska et al., 2013; Hostetter & Alibali, 2019; Masson-Carro et al., 2016; McNeill, 1985, 1992; Parrill, 2010; Parrill & Stec, 2018). This means that in addition to enriching the content of spoken sentences, gestures may also shed light on how speakers see themselves with respect to what they are describing. David McNeill (1992) differentiated two types of gestures that show a speaker’s subjective perspective: character viewpoint and object viewpoint gestures. Speakers adopting a first-person perspective would commonly produce a character viewpoint gesture, where the speaker’s hands act as an equivalent to the agent’s hands to mimic the agent’s action (e.g., two hands gesturing dropping an object). Speakers adopting a third-person perspective would instead produce an object viewpoint gesture, with their hands tracing the motion trajectory of the object (e.g., a balled fist depicting the path of an object dropping).

How do these different gestural viewpoints align with what is said in speech? Given that English speakers use transitive and intransitive sentences to mark agentic roles in an event (Fausey & Boroditsky, 2010; Fausey et al., 2010), it is likely that gestures also signal that relationship. Indeed, Fey Parrill (Parrill, 2010; Parrill & Stec, 2018) has shown that differences in gesture viewpoints align with linguistic differences in agentivity, with character viewpoint gestures going mostly with transitive sentences (e.g., saying “I dropped the vase,” while gesturing two hands letting go of an object) and object viewpoint gestures going mostly with intransitive sentences (e.g., saying, “The vase dropped,” while making a balled fist and moving it downward; see also Debreslioska et al., 2013). In this way, speakers can mark their own agentic relationship to actions and objects through multiple channels—through speech *and* gesture.

Although there are numerous studies demonstrating enhanced recall using iconic gestures that are congruent with the action uttered in a sentence, only been a few

studies have specifically manipulated the gesture viewpoint associated with those actions. Most of these past studies have explored the role of gesture viewpoints in the recall and interpretation of story content (Beattie & Shovelton, 1999a, 1999b; Cassell et al., 1999; Merola, 2009), and this work has shown that people are sensitive to a gesture’s agentivity. For example, Merola (2009) found that young children (~ 5–6 years old) remember aspects of a story better when a teacher produces character viewpoint versus object viewpoint gestures. However, to our knowledge, no study has explored how listeners negotiate viewpoints in gesture and speech that are explicitly pitted against one another. In other words, what are the relative contributions of speech and gesture viewpoints in how people construe a speaker’s subjective relationship to actions and events?

To answer this question, we borrowed an “enactment paradigm” used by previous experiments on the role of iconic gestures in speech comprehension and memory (e.g., Feyereisen, 2006, 2009). Specifically, we presented participants with gesture–speech pairs that conveyed either high or low agentic viewpoints. As an example in speech, the transitive sentence “I dropped the vase” conveys a highly agentic viewpoint, while the intransitive sentence “The vase dropped” conveys a minimally agentic viewpoint (such that it is not clear who or what did the dropping). Meanwhile, a highly agentic gesture would typically be a character viewpoint gesture (e.g., two hands letting go of a vase), whereas a minimally agentic gesture would typically be an object viewpoint gesture (e.g., a fist following the path of a falling vase). Combining these two viewpoints produces two variables with two levels each: agency in speech (high and low) and agency in gesture (high and low). To assess how speech and gesture viewpoints interact with one another in assessments of a speaker’s agency, the study employed two tasks: (1) explicitly reporting how personally responsible the speaker was for the event (encoding task) and (2) completing a cued recall of the spoken sentences on a surprise memory test (recall task).

Our first prediction is based on past findings regarding English speakers’ judgements and memory of accidental and intentional events (Fausey & Boroditsky, 2010; Fausey et al., 2010). This past work by Caitlin Fausey and colleagues has shown that even when an event is presented in a nonagentic manner (e.g., a video of someone accidentally dropping a vase), English speakers still often construe it in an agentic way (e.g., saying that “the person dropped the vase”). Based on this bias toward agentivity, we predicted a main effect of agency in speech in the encoding task—specifically, there should be greater attributions of responsibility for sentences that are high versus low in spoken agency. This would serve as a basic manipulation check that participants indeed differentiated degrees of agency in the spoken sentences. For the recall task, we predicted that because English speakers more naturally conceive of events in agentic terms, they would also

produce better memory for sentences whose agency in speech is high versus low when later asked to recall them.

Our second prediction explores this agency bias in gesture. If gestures are also viewed as a reliable source of agency (Debreslioska et al., 2013; Hostetter & Alibali, 2019; Masson-Carro et al., 2016; McNeill, 1985, 1992; Parrill, 2010; Parrill & Stec, 2018), we predicted a main effect of agency in gesture, such that gestures with highly agentic viewpoints would produce higher attributions of responsibility in the encoding task. The recall task is less straightforward. Because the recall task explicitly taps into memory for *speech* content, it is not clear what a main effect of gesture would mean. Therefore, we did not make any predictions about a main effect of agency in gesture on memory for speech during the recall task. Rather, we expected an interaction effect, which we describe next.

Our third prediction explores the interaction between agency in speech and agency in gesture. Focusing first on the encoding task, when agency in speech is high, we predicted only a slight dip in responsibility ratings when agency in gesture was low versus high. This is motivated by past work showing that there is a strong bias in English speakers toward attributing high levels of agency to speech describing actions (Fausey & Boroditsky, 2010; Fausey et al., 2010). Therefore, any deviation of that agency through gesture should cause only minor disruption. However, when agency in speech is low, we predicted that gestures of high agency would buoy up responsibility ratings, whereas gestures of low agency would considerably drag them down.

Regarding the recall task, based on research showing that the meaning behind gesture can affect memory for speech content (Beattie & Shovelton, 1999a, 1999b; Cassell et al., 1999; Cutica & Bucciarelli, 2008; Iani & Bucciarelli, 2017; Kelly et al., 2009; Merola, 2009; So et al., 2012), we predicted that opposing levels of agency in speech and gesture would disrupt correct recall of speech. That is, when speech and gesture both convey the same level of agency, correct recall of sentences would be strong, but when the two channels conflict, correct recall should suffer.

Method

Participants

Forty participants were recruited from a liberal arts school in northeastern United States (23 females, 17 males; 18–21 years).¹ All participants were self-reported monolingual native English speakers; all received course credit for compensation.

¹ We calculated the sample size for a two-way ANOVA using the ‘pwr2’ package in R. Our goal was to obtain a 0.99 power to detect a large effect size of 0.75 at the standard 0.05 alpha error probability [ss.2.way(a = 2, b = 2, alpha = 0.05, beta = 0.01, f.A = 0.75, f.B = 0.75, B = 100)]. R calculated an *n* of 9, resulting to a total sample size of 36.

Materials

Stimuli The stimuli in the study were videos of a native English speaker narrating sentences accompanied by gesture. The independent variables of interest were agency in speech and agency in gesture.² Agency in speech could either be high or low. Spoken narrations that are high in agency are transitive sentences, which in the current study follow the order *subject (I)–verb–object* (e.g., “I shattered the mirror”). Spoken narrations that are low in agency are intransitive sentences, which follow the order *object–verb* (e.g., “The mirror shattered”). Agency in gesture could also either be high or low. Gestures that are high in agency are character viewpoint gestures, which explicitly show an agent carrying out the action and are thus consistent with speech that is also high in agency. Character viewpoint gestures in the current study show the perspective of an agent acting on an object, either directly (e.g., dropping an object) or occasionally with a tool (e.g., using an axe). Gestures that are low in agency are object viewpoint gestures, which convey an object’s motion without any reference to an agent and are thus more consistent with speech that is low in agency. Object viewpoint gestures in the current study demonstrate the consequences of an object being acted upon, from the perspective of the object itself (e.g., an object falling or breaking). Figure 1 offers a visual representation of the four speech–gesture conditions.

Stimuli videos were filmed with the same actress speaking with an even tone and a neutral facial expression in front of a blank wall. The actress narrated either a transitive or intransitive sentence, with either a character viewpoint or an object viewpoint gesture. Gesture movements were timed to match the moment the actress uttered the verb, regardless of sentence construction. The video clips were edited using Final Cut Pro, and each video lasted approximately 3 seconds.

We opted to dub the videos for equal audio quality in order to ensure that any changes in the participants’ perception of speech was due to gesture and not to any differences in speech intonation (Krahmer & Swerts, 2007). The program Audacity was used to edit the background noise in the videos and to create the same speech audio for both versions of the videos where the actress enacted either a character viewpoint or an object viewpoint gesture. The speech audio was extracted from the video that reflected the most natural prosody where the speaker did not overly exaggerate any of the words in the sentence and was dubbed over the videos of both gesture conditions. The extracted audio file was aligned with the original video’s sound waves so that the mouth movements matched the sound.

There was a total of 24 sentences (see Table 1) and 96 videos. These 24 sentences were chosen through preliminary

² Our labels for the independent variables differ from what was presented in the preregistration, although our predictions remain the same.

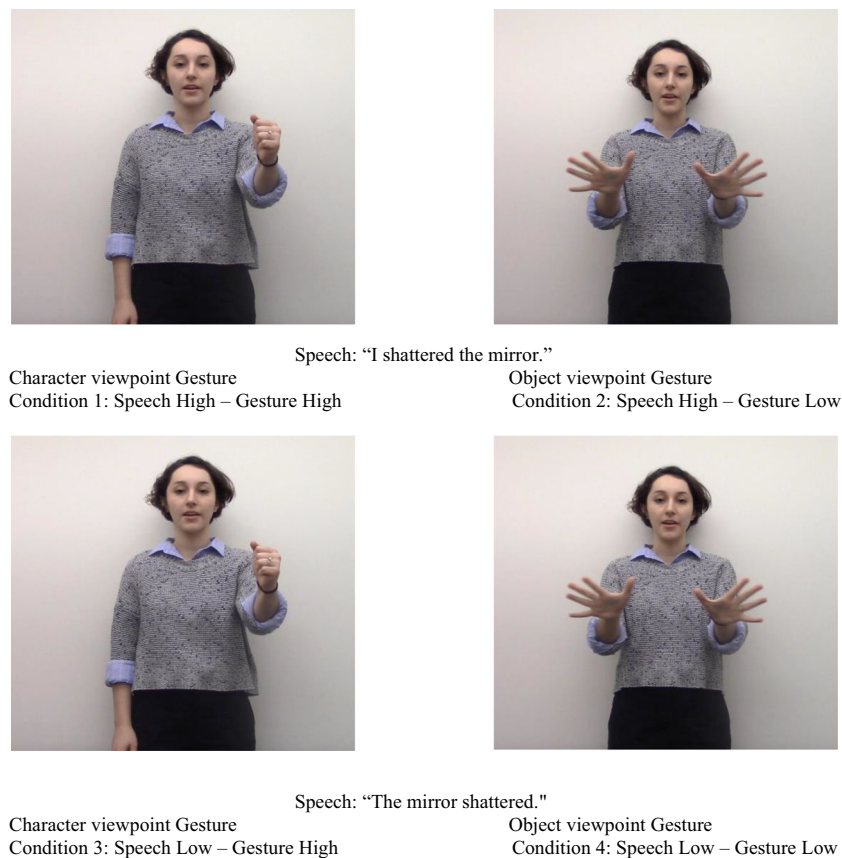


Fig. 1 Gesture–speech conditions. The videos consisted of a native English speaker narrating transitive (speech agency: high) or intransitive (speech agency: low) sentences alongside their character viewpoint (gesture agency: high) and object viewpoint (gesture agency: low) gestures. In this example, the character viewpoint gesture showcases

one-on-one discussions with 14 undergraduate students from an introductory psychology course. The experimenter showed students videos for the Speech High + Gesture High and Speech Low + Gesture Low conditions of each sentence and asked the students to describe and compare the relationship between speech and gesture for both conditions. Sentences in which the gesture–speech relationship were described to be unclear or ambiguous by most of the students were eliminated from the final set of videos.

Participants in the actual experiment were shown only one speech–gesture condition per sentence for a total of 24 videos per participant and six unique sentences per condition. The videos were presented using a computer survey built with Qualtrics, which automatically randomized the order of the videos being displayed. Four versions of the Qualtrics survey were created to equally represent each sentence in all four speech–gesture conditions across participants, resulting in 10 participants per version.

Recall test The recall test consisted of a sheet of paper that listed all 24 object names as memory cues (e.g., the word *mirror* would be used as a cue for the stimuli in Fig. 1).

the manner in which the event was performed (a fist “hitting” the air to represent an agent applying force to the mirror), while the object viewpoint gesture depicts the outcome of the event (both palms spreading out to represent the shards of glass as they scatter)

Written instructions were provided at the top of the form, and blank spaces were presented after each object name for the participants to write down their responses.

Procedure

The researcher assigned a participant to one of the four versions of the Qualtrics survey before they arrived. The researcher informed the participant that the purpose of the study was “to look at the ways people talk about events in their daily lives,” which was aimed at encouraging participants to focus on the speech and to prevent them from directing heightened attention to the gestures. Study instructions were read to the participants after they signed the consent form.

Participants were provided with a set of four unique practice trials at the beginning of the computer survey to ensure that they understood the procedure (see Table 1). If the participant had no questions about the procedure, the researcher exited the room to allow them to proceed with the experimental trials. The study consisted of two tasks: encoding and recall.

Table 1 Transitive and intransitive sentences used as speech stimuli

	Transitive sentence	Intransitive sentence
Practice 1	I <i>cracked</i> the egg	–
Practice 2	–	The candle <i>extinguished</i>
Practice 3	I <i>mixed</i> the sauces	–
Practice 4	–	The cookie <i>crumbled</i>
1	I <i>bounced</i> the ball	The ball <i>bounced</i>
2	I <i>broke</i> the pencil	The pencil <i>broke</i>
3	I <i>closed</i> the door	The door <i>closed</i>
4	I <i>crashed</i> the car	The car <i>crashed</i>
5	I <i>dropped</i> the vase	The vase <i>dropped</i>
6	I <i>flattened</i> the box	The box <i>flattened</i>
7	I <i>lowered</i> the window shade	The window shade <i>lowered</i>
8	I <i>moved</i> the plates	The plates <i>moved</i>
9	I <i>opened</i> the can	The can <i>opened</i>
10	I <i>overturned</i> the table	The table <i>overturned</i>
11	I <i>peeled</i> the sticker	The sticker <i>peeled</i>
12	I <i>popped</i> the balloon	The balloon <i>popped</i>
13	I <i>poured</i> the water	The water <i>poured</i>
14	I <i>rolled</i> the hoop	The hoop <i>rolled</i>
15	I <i>shook</i> the plant	The plant <i>shook</i>
16	I <i>shattered</i> the mirror	The mirror <i>shattered</i>
17	I <i>splattered</i> the paint	The paint <i>splattered</i>
18	I <i>split</i> the wood	The wood <i>split</i>
19	I <i>stretched</i> the elastic	The elastic <i>stretched</i>
20	I <i>swung</i> the bat	The bat <i>swung</i>
21	I <i>tore</i> the paper	The paper <i>tore</i>
22	I <i>turned</i> the key	The key <i>turned</i>
23	I <i>unzipped</i> the zipper	The zipper <i>unzipped</i>
24	I <i>waved</i> the flag	The flag <i>waved</i>

Encoding task Each stimulus video played automatically, and participants were only allowed to view each video once. After watching each video, participants were shown a question that read “On a scale from 0% to 100%, how responsible is she for this event?” and were presented with a slider scale on the same screen to provide their response. This dependent measure was based on how Fausey and Boroditsky (2010) operationally defined agency, which is the amount of responsibility attributed by the listener to the speaker for the action described in the sentence. Participants watched and assessed the speaker’s responsibility for all 24 videos one at a time and had no timing constraints on their responses.

Recall task After the participant indicated that they had finished all the videos, the researcher returned to the study room and provided them with a 2-minute distractor task by asking them to list as many countries as possible. After this distractor task, the experimenter presented the participant with a surprise

recall test. Specifically, they were asked to write down the *exact sentences* that they heard in the video. To help prompt memory, each item included a cue with a single key word (e.g., *mirror*). The researcher instructed the participant not to leave any items blank, encouraging them to guess, where possible. The participant was debriefed after the recall test and thanked for their participation. The entire study took approximately 30 minutes to complete.

Data coding and analysis

Encoding task The average percentage of responsibility assigned to the speaker was computed for each speech–gesture condition per participant. A 2×2 (Agency in Speech \times Agency in Gesture) repeated-measures analysis of variance (ANOVA) was performed for responsibility ratings. A significant interaction effect was followed by two orthogonal and planned paired *t* tests for high vs. low agency in gesture within each level of agency in speech.

Recall task In the recall task, participants wrote down from memory all 24 sentences they heard in the videos. Responses to the cued recall test were coded as correct recall of spoken agency only when participants wrote down the *exact* sentence as it was said in the videos. We also coded for the incorrect recall of spoken agency, wherein participants had answered with the opposite level of agency in speech from what was presented in the videos. Participants who demonstrated an incorrect recall of spoken agency misidentified the agent in the sentence by writing down an incorrect word order, as the only way to distinguish between agency levels in speech was through word order (i.e., sentences that were both high and low in spoken agency contained the same verb and the same object). As an example, if the original sentence was “The vase dropped,” an incorrect recall of spoken agency would be “I dropped the vase” or “I threw the vase.”

The same 2×2 repeated measures ANOVA and a priori paired *t* tests were run on both correct and incorrect recall of spoken agency.

Results

Encoding task

Consistent with our predictions, the influence of spoken agency on responsibility ratings was significant by subjects, $F(1, 39) = 208.03$, $p < .001$, $\eta_p^2 = 0.842$, and by items, $F(1, 23) = 370.12$, $p < .001$, $\eta_p^2 = 0.941$. Participants attributed lower levels of responsibility to the speaker when spoken agency was low versus high. The effect of agency in gesture on responsibility ratings was also significant by subjects, $F(1, 39) = 22.35$, $p < .001$, $\eta_p^2 = 0.364$, and by items, $F(1, 23) = 13.50$, $p < .001$, η_p^2

= 0.370, with participants rating the speaker as more responsible for the action when gestures demonstrated high as opposed to low agency. There was also a significant interaction effect of agency in speech and agency in gesture, by subjects, $F(1, 39) = 12.16, p = .001, \eta_p^2 = 0.238$, and by items, $F(1, 23) = 7.49, p = .01, \eta_p^2 = 0.246$. When spoken agency was high, responsibility ratings only slightly dipped when speech was paired with gestures that were low versus high in agency, by subjects, $t(39) = 1.82, p = .08$, and by items, $t(23) = 0.16, p = .87$. However, when spoken agency was low, highly agentic gestures substantially increased responsibility ratings compared to gestures that were low in agency, by subjects, $t(39) = 4.32, p < .001, \eta_p^2 = 0.324$, and by items, $t(23) = 3.21, p = .004, \eta_p^2 = 0.309$ (see Fig. 2).

Recall task

Correct recall There was a main effect of spoken agency on correct recall by subjects, $F(1, 39) = 7.12, p = .01, \eta_p^2 = 0.154$, and by items, $F(1, 23) = 7.94, p = .01, \eta_p^2 = 0.257$. When agency in speech was low, participants were able to correctly

recall the exact sentences more often than when agency in speech was high. However, differences in correct recall scores were not driven by agency in gesture, both by subjects, $F(1, 39) = 0.21, p = .65$, and items, $F(1, 23) = 0.10, p = .75$. Lastly, there was no interaction between agency in speech and agency in gesture, by subjects, $F(1, 39) = 1.60, p = .21$, and by items, $F(1, 23) = 1.87, p = .19$ (see Fig. 3).

Incorrect recall There was no significant main effect of agency in speech by subjects, $F(1, 39) = 0.07, p = .90$, and by items, $F(1, 23) = 0.02, p = .90$, and no significant main effect of agency in gesture, by subjects, $F(1, 39) = 0.82, p = .37$, and by items, $F(1, 23) = 0.41, p = .53$. However, there was a significant interaction between agency in speech and agency in gesture, by subjects, $F(1, 39) = 8.31, p = .006, \eta_p^2 = 0.176$, and by items, $F(1, 23) = 8.62, p = .007, \eta_p^2 = 0.273$. When agency in speech was high, incorrect recall of speech greatly increased with gestures that were low in agency, by subjects, $t(39) = -2.81, p = .008, \eta_p^2 = 0.168$, and by items, $t(23) = -2.81, p = .01, \eta_p^2 = 0.255$. When spoken agency was low, incorrect recall was not affected by whether agency in gesture

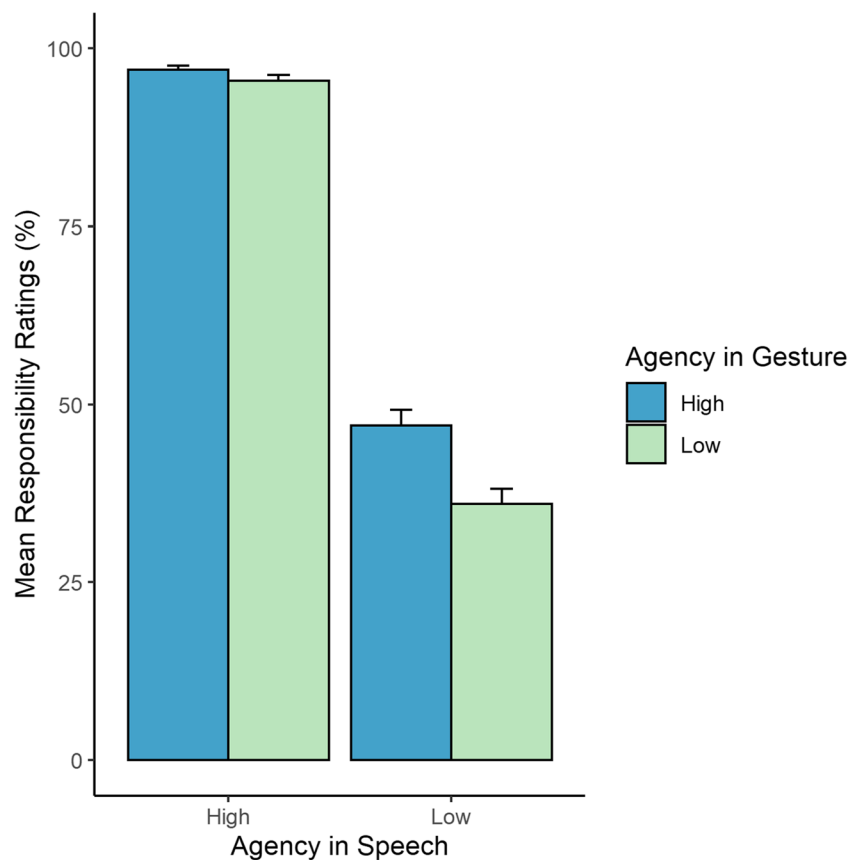


Fig. 2 Responsibility ratings in the encoding task with standard errors. Highly agentic speech produced higher responsibility ratings than speech that is low in agency. Highly agentic gestures also resulted in higher responsibility ratings than gestures that were low in agency. Additionally, agency in gesture interacted with agency in speech to

change responsibility ratings. Gestures that were low in agency slightly lowered responsibility ratings when agency in speech was high, but highly agentic gestures greatly increased responsibility ratings when agency in speech was low

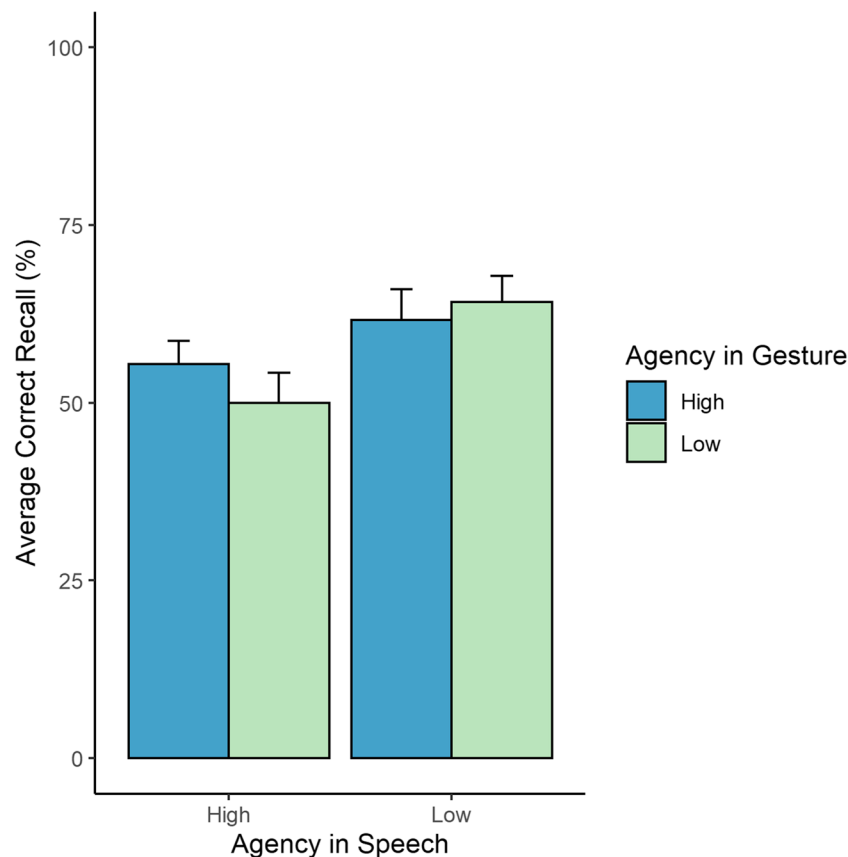


Fig. 3 Correct recall of spoken agency with standard errors. Speech that was low in agency had higher percentages of correct recall than speech that was high in agency, regardless of the level of agency in gesture. Percentages of correct recall were not different between the two levels of gestural agency

was high or low, by subjects, $t(39) = 1.24$, $p = .22$, and by items, $t(23) = 1.16$, $p = .26$.

To explore this potential effect of gesture–speech congruency on memory for speech, we compared the differences in incorrect recall scores when speech and gesture were either congruent or incongruent in their level of agency. Incorrect recall of spoken agency was minimal when gesture and speech were congruent (Speech High + Gesture High and Speech Low + Gesture Low), but errors significantly increased when gesture and speech were incongruent (Speech High + Gesture Low and Speech Low + Gesture High), by subjects, $t(39) = -2.88$, $p = .006$, $\eta_p^2 = 0.176$, and by items, $t(23) = -2.94$, $p = .007$, $\eta_p^2 = 0.273$ (see Fig. 4).

Discussion

With regard to the encoding task, participants assigned higher levels of responsibility to the speaker when agency in speech was high and also when agency in gesture was high. Moreover, highly agentic, character viewpoint gestures produced large increases in responsibility ratings when spoken agency was low. These findings suggest that both speech and gesture are used to make judgments of agency. With regard to

the recall task, participants actually remembered sentences better when agency in speech was low. Although we did not find any interaction between agency in speech and gesture in the correct recall of sentences, we found that incorrect recall increased when the level of agency in speech and gesture did not match. This suggests that incongruency between speech and gesture may have caused listeners to conflate spoken and gestural agency in memory.

Encoding of agency Past studies found that a highly agentic framing of events (i.e., through the transitive sentence construction) prompted English speakers to attribute higher levels of blame to individuals involved in the event compared with a nonagentic framing (Fausey & Boroditsky, 2010). We extend these observations by demonstrating that agentivity conveyed through another modality, hand gesture, also influences attributions of a person’s level of responsibility. As expected, when the level of agency in speech was high, listeners in our study based their responsibility ratings almost entirely on the speaker’s words even when the hands conveyed lower agency through an object viewpoint gesture (consistent with Fausey & Boroditsky, 2011). However, listeners were swayed by the heightened agentivity offered by gesture when they heard intransitive sentences with character viewpoint gestures. This

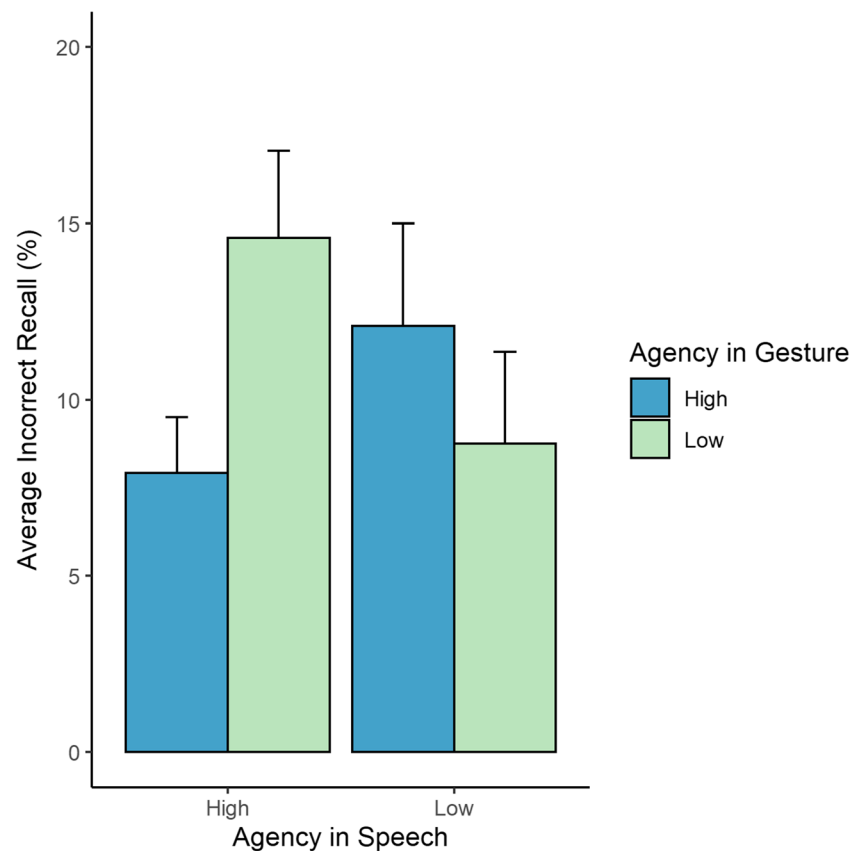


Fig. 4 Incorrect recall of spoken agency with standard errors. Agency in gesture interacted with agency in speech to change incorrect recall. The percentage of incorrect recall increased when speech was paired with a gesture of the opposing level of agency

indicates that listeners may also have an agency bias when processing gesture: When speech is low or ambiguous in personal agency, listeners use gestures high in agency to attribute more responsibility to a speaker.

Recall of agency Contrary to our predictions, our results showed that participants were more accurate in recalling sentences when speech was low in agency. This could potentially be ascribed to a novelty effect in memory in which people are more likely to remember distinct, novel materials over familiar ones (Kishiyama & Yonelinas, 2003; Poppenk et al. 2010), particularly as intransitive sentences are not a common feature of everyday conversations for English speakers (Roland et al., 2007). A novelty effect for sentences that were low in spoken agency may actually indirectly support past findings regarding how English speakers interpret intentional and accidental events. Cross-linguistic studies have demonstrated that English speakers tend to approach events through agentic lens, mirroring their greater use of agentic language to describe both intentional and accidental events (Fausey & Boroditsky, 2011; Fausey et al., 2010). Because English speakers often do not describe events in a nonagentic manner, the presence of numerous intransitive sentences over the course of the study may have caused such

sentences to stand out to listeners and become more memorable.

Interestingly, there was no interaction between agency in speech and agency in gesture in the correct recall of sentences, which may seem inconsistent with past studies showing that memory for spoken sentences was enhanced when information expressed in gesture matched that of speech (Cohen, 1989; Cutica & Bucciarelli, 2008; Engelkamp & Krumnacker, 1980; Engelkamp, et al., 1994; Feyereisen, 2006, 2009; Iani & Bucciarelli, 2017; Kelly et al., 2009; So et al., 2012; Russ, et al., 2003; Stevanoni & Salmon, 2005). This so-called enactment effect has been most prominently demonstrated in memory for action phrases and sentences. Free recall of spoken phrases and sentences improved when participants were asked to either produce gestures themselves (subject-performed task [SPT]) or watch a speaker produce the gestures (experimenter-performed task [EPT]; Cohen, 1989; Engelkamp, 1998; Engelkamp et al., 1994). When the congruency between gesture content and speech content was manipulated, results from past studies show that congruent gestures support better recall than incongruent gestures or no gestures at all (and indeed, incongruent gestures often produced worse recall than no gestures at all; Cutica & Bucciarelli, 2008; Feyereisen, 2006; Iani & Bucciarelli,

2017; Kelly et al., 2009; So et al., 2012). The present study differs from previous enactment studies in two major ways.

First, the current study required participants to respond with the *exact* sentences that they heard, whereas much of the previous work allowed for paraphrases (Cutica & Bucciarelli, 2008; Feyereisen, 2006; Iani & Bucciarelli, 2017). Hence, our strict instructions and more stringent criteria for correct recall may have limited our ability to detect any influence of gesture on correct recall for speech.

Second, there was an important difference in how gesture “congruency” was manipulated in the present study. Past experiments have often created truly incongruent gestures that convey completely different semantic content than what was expressed in speech. Such gestures can be considered “event incongruent” in that they depict a separate and unrelated event from speech. For example, in Feyereisen (2006), the sentence, “He closed the book before the end of the story” was coupled with an incongruent “jumping” gesture. This is a complete—and artificial—semantic disjuncture between the two modalities, which is common in many gesture comprehension experiments (Church et al., 2017). In contrast, the semantic relationship explored in the present study—viewpoint “incongruency”—is a more natural case of when multiple agentic relationships are captured in speech and gesture. These viewpoint incongruent gestures reflected the same events as their congruent counterparts, but simply offered differing levels of personal agency for the speaker. For example, the character viewpoint gesture (hand hitting) and object viewpoint gestures (glass shattering) for the sentence “She shattered the mirror” both refer to the same event of glass breaking. The only difference is the perspective on the event—in one case, the perspective is from the person doing the hitting, and in the other, it is from the mirror doing the shattering (see Fig. 1). Compared with previous studies (e.g., Feyereisen, 2006; Kelly et al., 2009), it is plausible that our viewpoint incongruent gestures did not disrupt memory for speech because they still conveyed information about the *same* event.

This “viewpoint incongruity” between gesture and speech is interesting to consider in light of other studies that have manipulated the degree to which gestures and speech are semantically disconnected (Cutica & Bucciarelli, 2015; Kelly et al., 2010). For example, Cutica and Bucciarelli (2015) created a fully “unrelated” gesture condition by delaying a video track by 20 seconds from the corresponding speech track, thus totally eliminating any meaningful relationship between speech and gesture. Not surprisingly, these unrelated gestures produced worse memory for speech than “related” gestures (with no video delay), but interestingly, memory for speech accompanied by these unrelated gestures was no different from speech alone. This suggests that when the semantic distance between gesture and speech becomes too great, listeners can entirely disregard them without disruption to speech processing or memory. In a different vein, Kelly et al. (2010) manipulated the strength of incongruency between

speech and gesture in an online priming task, finding that when gestures were weakly incongruent to speech (saying “chop,” while gesturing *cut*), processing of targets was quicker and more accurate than when gestures were strongly incongruent (saying “chop,” while gesturing *twist*).

Taking all these studies together, it appears that processing gesture–speech relationships operates on a continuum: When gestures are too unrelated to speech—through temporal and semantic distancing—they can easily be ignored (Cutica & Bucciarelli, 2015; see also Habets et al., 2011). Moving along the continuum, when gestures occur together with speech, but carry completely different semantic content (i.e., they are “event incongruent”), they greatly disrupt comprehension and memory for speech (e.g., Feyereisen, 2006). And a bit further down the line, when gesture and speech co-occur, but convey slightly different semantic content about the same event, as with the weakly incongruent gestures from Kelly et al. (2010), they disrupt speech to a much lesser extent than strongly incongruent ones. And even further down the line, we arrive at the present study: When gesture and speech co-occur, and both focus on different agentic perspectives of the *same* event (i.e., they are “viewpoint incongruent”), they do not disrupt comprehension or memory at all. In fact, it is entirely possible that they may even enhance it. For example, Kelly et al. (1999) found that sentences accompanied by “complementary” iconic gestures (e.g., saying “My brother went to the gym,” while making a basketball-shooting gesture) help people recall the spoken message better than when no gestures accompanied the speech. Running with this in a future study, it would be interesting to compare speech accompanied by “viewpoint incongruent” gestures to speech accompanied by no gestures. One might predict that these gestures—like the complementary gestures from above—might boost correct recall of speech compared to speech alone. Regardless of whether this particular prediction is born out, it is safe to conclude that not all gesture–speech relationships are created equal in language comprehension.

The story gets even more nuanced when considering the data from the “incorrect recall” measure. Even though participants did not produce many incorrect recollections (~12% overall), we found an interaction effect between spoken agency and gestural agency on the errors that participants did make. We observed that participants produced many more memory mistakes when speech was paired with gestures of the opposing level of agency. For example, when presented with the sentence “The mirror shattered,” a participant might incorrectly write down “She shattered the mirror,” if the speech was accompanied by a viewpoint-incongruent character gesture (a hand making a hitting gesture). This suggests that gestural information on viewpoint can slip into memory when listeners fail to remember exact details from speech. Hence, listeners construct their account of events by considering information presented by both modalities, relying on the

visual modality to help reconstruct memory when memory for speech is fuzzy (Broaders & Goldin-Meadow, 2010).

It is interesting to consider the current study in association with the literature on event representations in sign language. Signers also construct viewpoints in two main ways—through entity and handler constructions (Cormier et al., 2012). Entity constructions depict the physical characteristics of an object (e.g., an upside down V to indicate bipedal legs) while handler constructions represent the manner in which an object is manipulated (e.g., two hands “holding and turning a page” for newspaper), and these distinctions seem to map onto McNeill’s object viewpoint and character viewpoint gestures (Cormier et al., 2012). As with spoken language, signers describing events in a transitive manner often use handler constructions, while intransitive expressions go with entity constructions (for a review, see Stec, 2012). Hence, signers are intentional with using their hands to communicate their perspective on events, and it would be interesting to explore whether different viewpoints in sign also influence how an addressee may interpret and remember the signer’s subjective perspective.

In conclusion, gestures are capable of not only relaying the objective content of events, but also revealing a speaker’s own personal sense of agency in relation to those events. Building on McNeill’s observations about gesture viewpoint in language production 35 years ago (McNeill, 1985), listeners consider gestural viewpoints alongside speech in forming judgements and remembering details about who did what to objects out in the world. In this way, multimodal cues gleaned from both speech and gesture feature prominently in how listeners construct a comprehensive account of events communicated to them.

Acknowledgements We would like to thank Shiri Spitz for her help in creating the video stimuli, Pomelo Wu for her assistance in data collection, and members of the Center for Language and Brain at Colgate University for their helpful comments and suggestions.

References

- Beattie, G., & Shovelton, H. (1999a). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology*, 18(4), 438–462. <https://doi.org/10.1177/0261927X99018004005>
- Beattie, G., & Shovelton, H. (1999b). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, 123(1/2), 1–30. <https://doi.org/10.1515/semi.1999.123.1-2.1>
- Broaders, S. C., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science*, 21(5), 623–628. <https://doi.org/10.1177/0956797610366082>
- Cassell, J., McNeill, D., & McCullough, K. (1999). Speech–gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition*, 7(1), 1–34. <https://doi.org/10.1075/pc.7.1.03cas>
- Church, R. B., Alibali, M. W., & Kelly, S. D. (Eds.). (2017). *Why gesture? How the hands function in speaking, thinking and communicating*. Amsterdam: John Benjamins Publishing.
- Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review*, 1(1), 57–80. <https://doi.org/10.1007/BF01326550>
- Cormier, K., Quinto-Pozos, D., Sevcikova, Z., & Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: Comparing depicting constructions and viewpoint gestures. *Language and Communication*, 32, 329–348. <https://doi.org/10.1016/J.LANGCOM.2012.09.004>
- Cutica, I., & Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science*, 32, 921–935. <https://doi.org/10.1080/03640210802222039>
- Cutica, I., & Bucciarelli, M. (2015). Non-determinism in the uptake of gestural information. *Journal of Nonverbal Behavior*, 39(4), 289–315. <https://doi.org/10.1007/s10919-015-0215-7>
- Debreslioska, S., & Gullberg, M. (2017). Discourse reference is bimodal: How information status in speech interacts with presence and viewpoint of gestures. *Discourse Processes*, 50(1), 431–456. <https://doi.org/10.1080/0163853X.2017.1351909>
- Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, 50(7), 431–456. <https://doi.org/10.1080/0163853X.2013.824286>
- Engelkamp, J. (1998). *Memory for actions*. United Kingdom: Psychology Press/Taylor & Francis.
- Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 27(4), 511–533.
- Engelkamp, J., Zimmer, H.D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory and Cognition*, 22(1), 34–39. <https://doi.org/10.3758/bf03202759>
- Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, 17(5), 644–650. <https://doi.org/10.3758/PBR.17.5.644>
- Fausey, C.M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, 18, 150–157. <https://doi.org/10.3758/s13423-010-0021-5>
- Fausey, C.M., Long, B.L., Inamori, A., & Boroditsky, L. (2010). Constructing agency: The role of language. *Frontiers in Psychology*, 1, 162. <https://doi.org/10.3389/fpsyg.2010.00162>
- Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, 18(2), 185–205. <https://doi.org/10.1080/09541440540000158>
- Feyereisen, P. (2009). Enactment effects and integration processes in younger and older adults’ memory for actions. *Memory*, 17(4), 374–385. <https://doi.org/10.1080/09658210902731851>
- Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Harvard University Press.
- Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children’s hands to read their minds. *Cognition and Instruction*, 9(3), 201–219. https://doi.org/10.1207/s1532690xc0903_2
- Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, 23(8), 1845–1854. <https://doi.org/10.1162/jocn.2010.21462>
- Holle, H., & Gunter, T.C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, 19(7), 1175–1192. <https://doi.org/10.1162/jocn.2007.19.7.1175>

- Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137(2), 297–315. <https://doi.org/10.1037/a0022128>
- Hostetter, A.B., & Alibali, M.W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Hostetter, A.B., & Alibali, M.W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, 26(3), 721–752. <https://doi.org/10.3758/s13423-018-1548-0>
- Iani, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, 96, 110–121. <https://doi.org/10.1016/j.jml.2017.05.004>
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40(4), 577–592. <https://doi.org/10.1006/jmla.1999.2634>
- Kelly, S.D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gestures lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, 24(2), 313–334. <https://doi.org/10.1080/01690960802365567>
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2), 260–267. <https://doi.org/10.1177/0956797609357327>
- Kendon, A. (1986). Some reasons for studying gesture. *Semiotica*, 62(1/2), 1–28.
- Kishiyama, M. M., & Yonelinas, A. P. (2003). Novelty effects on recollection and familiarity in recognition memory. *Memory & Cognition*, 31(7), 1045–1051. <https://doi.org/10.3758/BF03196125>
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16–32. [https://doi.org/10.1016/S0749-596X\(02\)00505-3](https://doi.org/10.1016/S0749-596X(02)00505-3)
- Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396–414. <https://doi.org/10.1016/j.jml.2007.06.005>
- Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, 31(3), 430–440. <https://doi.org/10.1080/23273798.2015.1108448>
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92(3), 350–371. <https://doi.org/10.1037/0033-295X.92.3.350>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Merola, G. (2009). The effects of the gesture viewpoint on the students' memory of words and stories. In M. Sales Dias, S. Gibet, M. M. Wanderley, & R. Bastos (Eds.), *Gesture-based human-computer interaction and simulation* (pp. 272–281). Springer.
- Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, 19(4), 605–616. <https://doi.org/10.1162/jocn.2007.19.4.605>
- Parrill, F. (2010). Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25(5), 650–668. <https://doi.org/10.1080/01690960903424248>
- Parrill, F., & Stec, K. (2018). Seeing first person changes gesture but saying first person does not. *Gesture*, 17(1), 158–175. <https://doi.org/10.1075/gest.00014.par>
- Poppenk, J., Köhler, S., & Moscovitch, M. (2010). Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1321–30. <https://doi.org/10.1037/a0019900>
- Roland, D., Dick, F., & Elman, J.L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348–379. <https://doi.org/10.1016/2Fj.jml.2007.03.002>
- Russ, M.O., Mack, W., Grama, C.R., Lanfermann, H., & Knopf, M. (2003). Enactment effect in memory: Evidence concerning the function of the supramarginal gyrus. *Experimental Brain Research*, 149(4), 497–504. <https://doi.org/10.1007/s00221-003-1398-4>
- So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall?. *Language and Cognitive Processes*, 27(5), 665–681. <https://doi.org/10.1080/01690965.2011.573220>
- Stec, K. (2012). Meaningful shifts: A review of viewpoint markers in co-speech gestures and sign language. *Gesture*, 12(3), 327–360. <https://doi.org/10.1075/gest.12.3.03ste>
- Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior*, 29(4), 217–233. <https://doi.org/10.1007/s10919-005-7721-y>
- Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, 101(3), 234–245. <https://doi.org/10.1016/j.bandl.2006.12.003>

Open practices statement The data for the study are available from the corresponding author upon request. Stimuli videos are available (<https://osf.io/8bd3n/>), and the experiment in this study was preregistered (<https://osf.io/juc8p>).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.