Representational gesture and memory: an empirical analysis

Tara Ghazi
Pembroke-Kings Programme, University of Cambridge
University of California, Berkeley

First Reader & Faculty Mentor

Dr. Eve Sweetser
Department of Linguistics
University of California, Berkeley

Second Reader & Faculty Mentor

Dr. Terrence Deacon
Department of Anthropology
University of California, Berkeley

ABSTRACT

Gestures are nonverbal, visible indexes of cognition (Hostetter & Alibali, 2008), interesting in part for what they reveal of mental representations on behalf of a speaker during communication. While study of those performing gesture is extensive, a great deal is yet to be understood about the role gesture serves for those viewing it. In an empirical study we explore how viewing gesture that accompanies speech (co-speech gesture) impacts memory and mental representation of a story scene relayed in English to both native and non-native English speakers. We find that gesture is significantly associated with accurate memory of an episodic story scene – an effect driven by non-native, but not native, speakers of English. Across participants, gesture provides surprisingly minimal aid to accurate recall of gesture-only story details. We consider differences in viewpoint along with spatial and temporal directionality norms for native English speakers as potential explanatory factors.

INTRODUCTION

Gesture is a common accompaniment to verbal communication across languages and cultures (Goldin-Meadow & Beilock, 2010). This study examines the explicit conveyance of information to clarify and supplement speech for a listener (Cassell, McNeill, & McCullough, 1999), is focused upon herein. “Spoken words […] are inherently ambiguous” (Kendon, 2000, p. 51). Gesture is a solution for that communicative dilemma.

Working memory is the term used to describe the subgroups of memory that function together as the moment to moment workhorse of cognition (Baddeley & Logie, 1999; Shah & Miyake, 1999). By connecting sensory experience to mental representation, and manipulating and integrating incoming information, working memory acts as the bridge between the transient and longer term memory stores (Baddeley, 2007). It is important to consider paradigms of working memory, as they may provide a means for understanding how gesture holds utility for the listener.

BACKGROUND

GESTURE

Gestures described as “representational” are those whose function is to represent or depict (Kita, 2000). For instance, placing one fist on top of the other and swinging arms in a reenactment of a batter striking a ball illustrates a “character viewpoint” – one representational gesture subtype. A storyteller slowly moving their hand in an upward arc, and then plunging it downwards uses an “observer viewpoint” representational gesture subtype to illustrate the speed and drama of a rollercoaster ride. “Deictic” gesture is used when a speaker defines the boundaries of an imagined space and shares it with a listener by using discrete motions towards the regions of space in front of them. In these examples, gestures clarify and enrich an accompanying verbal account (Beilock & Goldin-Meadow, 2010), bringing the listener’s mental representation of a scene into closer alignment with that of the speaker.

The communicative transfer from the gesturing speaker to the viewing listener is dependent upon listener cognition. Working memory is a critical part of cognition (Shah & Miyake, 1999), and therefore important in the communicative transfer. To understand how that transfer may be occurring, two current models of working memory – the Baddeley and Hitch Multicomponent Model, and the Cowan Embedded-Processes Model – were considered (Baddeley & Logie, 1999; Cowan, 1999).

MODELS OF MEMORY

THE MULTICOMPONENT MODEL

Baddeley and Hitch first proposed the *multi-component model of working memory* in 1974 (Baddeley, 2007). The model describes four sub-systems that comprise working memory and operate distinctly under this model from long-term memory. Of these sub-systems, the *phonological loop* handles meaningful speech and sound. Visual sensory information, including viewed gesture as well as maintained mental imagery, is handled by the *visuospatial sketchpad* (Wesp, Hesse, Keutmann, & Wheaton, 2001). An *episodic buffer* links, coordinates, and mediates information between the two. These three sub-modules link bi-directionally to a fourth and final module, the *central executive*. This *central executive* handles three aspects of attentional control – focusing, dividing and switching – while also serving as the conduit to long-term memory storage (Baddeley, 2007).

THE EMBEDDED-PROCESSES MODEL

Under the *Embedded-Processes Model*, proposed by Cowan, working memory is a system of dynamic parts continually regulating incoming information. Information from the environment is routed through a short-term sensory store before being transferred to working memory for manipulation. Previously stored information may be retrieved from long term memory, where it is elevated to an activated level within working memory. The model does not distinguish between modalities of sensory information – auditory or visual – but rather channels and prioritizes based on significance and novelty

(Cowan, 1999). Critical information is raised to a level of highest priority and most constrained capacity – that of *focus*. Focus, and what shifts in and out of focus, is managed by a *central executive*, akin to the controller of the Multi-Component Model. Separate from the level of focus, two other activation levels exist, which an individual may or may not be consciously aware of. Both novel incoming information and information from long-term stores, if they are independently successful in passing through the filtering constraints of the working memory system, and arrive at an activation level, have the potential to be retained and hence stored or re-stored again in long-term memory. Only information being focused upon is a person necessarily aware of, whereas information in the other two dynamic levels may indeed be remembered, but without an individual being conscious of it (Cowan, 1999).

MEMORY AND GESTURE

A speaker who gestures while talking - thus exhibiting co-speech gesture - is presenting information simultaneously in two modalities: the auditory through verbal utterances, and the visual through their gesture. An individual attending to a gesturing speaker therefore, will have two channels of information available to them. Even when listeners are unaware of attending to gesture, it has been demonstrated that they do assimilate unspoken with spoken content (Cassell et al., 1999; Goldin-Meadow, Wein, & Chang, 1992). Attentional awareness therefore, may not be prerequisite to the memory of gestural information. Per both the Multi-Component and Embedded-Processes Models of working memory, both incoming verbal and visual information must pass through working memory to be stored long-term (Baddeley, 2007; Baddeley & Logie, 1999; Cowan, 1999), and long-term storage is required for retrieval (Foster & Jelicic, 1999; Glenberg, 1997).

Under the Multi-Component Model, the *phonological loop* processes incoming auditory information; co-speech gesture is processed separately by way of the *visuospatial sketchpad*. The two modalities are coordinated, and the semantic contents of each are brought together by conjoined processing of the *episodic buffer* and the *central executive*. Information cycling through the *visuospatial sketchpad* – which may induce co-speech gesture on the part of a speaker – has been interpreted as a form of internal mental rehearsal which aids memory (Wesp et al., 2001). In addition to a communicative purpose, this raises a cognition function for gesture - but on behalf of the speaker, not the listener.

Under the Embedded-Processes Model, visual and verbal information are routed to working memory via universal channels that do not distinguish between modalities. As with the Multi-Component Model, a *central executive* manages attentional control - raising a portion of each modality of content onto the central stage of focus at any given time. Novel information is given attentional priority over that which is familiar or repetitive, meaning new content is more likely to be "activated" and thus more likely to be stored in long-term memory (Cowan, 1999). Under this model, a listener exposed to a story coupled with co-speech gesture may therefore be more likely to remember story content due to its novelty. Counter to the Embedded-Processes Model though, the dual funneling of auditory and visual information into the same channels could create competition between the two modalities for attentional resources.

Study of representational gesture use provides an important window into a speaker's mental processes (Cartmill, Beilock, & Goldin-Meadow, 2012; Hostetter & Alibali, 2008). It has also been argued that gestures function not just as communicative devices, but as aids to working memory on the part of the gesturing speaker (Wesp et al., 2001). On the part of the listener, viewing gesture can clarify ideas and impact thoughts (Beilock & Goldin-Meadow, 2010). How this occurs, and what role working memory has in this process, are both questions yet to be answered. To move towards an understanding of these processes, we raise the question of whether co-speech gesture impacts *what* a listener remembers.

We hypothesize that a listener presented with co-speech gesture will exhibit greater and more accurate memory than a listener presented with speech only. In the case where verbal information is not concrete, we expect a listener's working memory to enable integration of gestural information into their mental construct of a scene to compensate. Predicted by the Multi-Component Model, the *central executive* would prioritize content from the visuospatial sketchpad – gesture – when incoming information to the *phonological loop* is ambiguous. Interpreted through the Embedded-Processes Model, the lack of competition from verbal information would render novel gestural information more likely to be raised to an attentional center stage – increasing the likelihood it will be stored in long-term memory.

To test this hypothesis, we conduct an empirical study in which an episodic story is relayed to participants. Verbal details in the story were intentionally ambiguous, with the presence or absence of clarifying gesture employed as the independent variable. Participant recall of verbal details and memory of the depicted scene was assessed, and analysis of differences in performance was carried out.

EMPIRICAL STUDY

METHODS

PARTICIPANTS

Twenty-six students from the *Pembroke-King's International Programme* at the University of Cambridge took part in the experiment. All participants were between the ages of 18 and 42, with a mean age of 21. Participants self-identified as native or non-native English speakers, but all had college-level English fluency per criteria for international student admission to the program. Fourteen of the total participants, 11 females and 3 males, were native English speakers (E). Twelve of the total participants, 7 females and 5 males, were non-native English speakers (O).

Participants were recruited from a program attended by students originating from a range of countries and cultural backgrounds. Controlling for all potential differences was not feasible. This was acknowledged about the participant population. Participants were randomly assigned to control and test groups as a pool. Native language was considered vital to consider during data analysis, but did not impact group assignment. On completion of the experiment, data from 1 E female and 1 O female was excluded due to a data collection error in the first case and participant failure to follow experimental directions in the second case. All participants had normal hearing and vision, with or without correction.

MATERIALS

The first group watched a video that included both speech and gesture usage (SG). The second group watched a video with speech only and no gesture (SO). Both videos showed a speaker, the experimenter, in front of a neutral background (*Figure 1, Figure 2*), verbally relaying a short story (*Figure 8*). Video recording, rather than live experimenter presentation, both of which are utilized in study of gesture (see (Kelly, Özyürek, & Maris, 2009) and (Kelly, McDevitt, & Esch, 2009) for examples) was chosen for consistency of experience across participants. In the SO video, no arm or hand gestures are present. In the SG video, 15 separate gestures occur, 12 of which are representational (*Figure 8*). In a study of gesture and working memory, Wagner and colleagues showed that conflicting verbal and gestural content adversely impact participant memory (Wagner, Nusbaum, & Goldin-Meadow, 2004). . To avoid this effect, only gesture that reinforced and elucidated speech was utilized in the SG video. Effort was made by the experimenter to use similar vocal intonation and emphasis in each case, however each video was separately recorded. Each video was 43 seconds in length.

The story relayed in both SO and SG videos, was novel to this study (*Figure 8*). It is intentionally episodic and action-based to facilitate the use of gesture, previously shown to naturally co-occur with speech at a greater rate during spatial descriptions (Lavergne & Kimura, 1987). Story length was limited to six, active grammar sentences comprised of common American English words (*Figure 8*). The arc of the story followed a logical progression of sequential events to facilitate comprehension. The main character in the story was described as a child, with the gender-neutral moniker "C." During the interview phase of the experiment, all participants confirmed the story was easily understood.

Figure 1: Frame from SG Video: Speech with Gesture

Figure 2: Frame from SO Video: Speech Only

After the participants watched the video, they were provided with a tray holding 4 paper origami objects with which to perform a physical scene reconstruction task (SR-P). Since participants differed in cultural background as well as native language, physical scene reconstruction was chosen as paradigm to minimize cultural biases (Rogoff & Waddell, 1982). SR-P objects included a House, a Character C, a Puppy, and a Ball. On completion of the SR task, the participant was provided with a pointer object to indicate the vantage point from which they mentally visualized the story scene. Prior studies of mental scene reconstruction in response to reading an episodic narrative have shown that participants may assume a 1st person perspective envisioning the story scene unfolding around them (i.e. internal viewpoint), or they may assume a 3rd person perspective as a detached viewer who watches the story scene unfold at a distance (i.e. external viewpoint) (Bryant, Tversky, & Franklin, 1992). Through the pointer object, viewpoint was established.

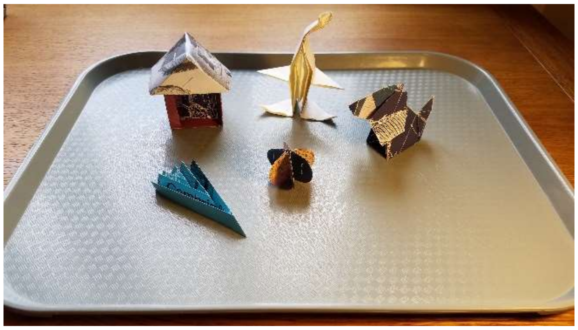


Figure 3: Scene reconstruction task objects and tray. Listed clockwise from the top left: House, Character, Puppy, Ball, Pointer.

Participant accuracies in recollection of story events and story object positioning were assessed by three means, each novel to this study. This included a questionnaire instrument (*Figure 9*), a physical scene reconstruction task (*Figure 10*), and an interview conducted by the experimenter (*Figure 11*) in which participants explained their recall of events and choices in positioning objects. The questionnaire instrument was comprised of 11 questions. Two of the questions required participants to make inferences about object positioning that were communicated only through deictic co-speech gesture in the SG video. SO participants lacked access to this information. Correct SO participant responses to those two questions therefore, were expected to be at chance. Similarly, correct completion of the SR-P task was dependent on gestural information relayed only through the SG video – where gesture indicated which story object was positioned to each side of the character. SO participants were also expected to perform at chance in the SR-P task.

For baseline comparison between participants, verbal and visual memory was assessed with a 25-word list (*Figure 12*), and an image, respectively. The 25 words were a random selection from a longer list of common nouns individuals learning English as a second language encounter. The image chosen was a black and white line drawing of a complex scene that contained at least 30 identifiable noun objects. Memory tasks were separated to account for separate sensory processing modalities.

PROCEDURE

Participants were randomly assigned to one of two groups; both groups progressed through the same experimental phases. A single participant completed the experiment at a given time, within a 30-minute time-period. The room where the experiment was conducted was a private, artificially lit space with a single window. Environmental cues have been shown to impact participants frame of reference, a possible confound when the task of interest involves positioning objects within a scene (Li & Gleitman, 2002). Thus, view of the outside was closed off by drawn curtains.

At the start of the experiment, the participant was instructed to take a seat at the single desk in the room, on the surface of which was a laptop only. The participant was advised that instructions would be provided on the screen, and that the experimenter would be present to facilitate and answer any questions they had. The experimenter remained present in the room throughout the entire course of the experiment in each case. On-screen instructions were presented via *Qualtrics*, which followed a survey format designed by the experimenter.

Each participant was presented with an online consent form and was advised a recorded interview would occur. The stated reward for participation was entry into a raffle for one of several small prizes, each valued at £15 or less, to be held after completion of all scheduled experimental sessions. All participants provided informed consent.

After consent was provided, the experimenter brought onto screen the SG or SO video as determined by random group assignment. The participant initiated viewing when they were ready. After the participant watched the video, the experimenter placed a tray holding 4 paper origami objects (House, Character, Puppy, Ball) on the desk while the participant was presented on screen with instructions for the physical scene recreation task (SR-P). On completion of object positioning, pointer task instructions (*Figure 10*Figure 10) followed, and the experimenter placed the pointer on the desk for the participant. Next, the participant was prompted on-screen to answer a series of questions presented in randomized order, without time constraints to test their memory of story details. On completion of questionnaire items (*Figure 9*), the experimenter requested the participant move to an adjacent seat and began the recorded interview (*Figure 11*).

Following the interview, the experimenter provided verbal instructions and administered verbal and visual memory tests. By conducting these tests after the main experiment, it was expected that all participants had an equal possibility of being primed by the previous tasks. In the verbal memory test, participants were instructed to write down all words they recalled hearing, in any sequence, after the experimenter read aloud the list of 25 words (*Figure 12*). In the visual memory test participants were instructed to write down all objects they recalled after viewing an image for 30 seconds. Finally, participants completed a brief demographic questionnaire and were then debriefed by the experimenter.

RESULTS

In the verbal memory test, SG participants scored a mean of 10.2 words recalled with a standard deviation of 2.85. SO participants scored a mean of 10.0 words recalled with a standard deviation of 2.77. In the visual memory test, SG group participants scored a mean of 14.7 noun items recalled with a standard deviation of 3.01; SO participants a mean of 14.3 noun items recalled with a standard deviation of 2.81. No significant baseline difference in demonstrated verbal or visual memory ability between participant groups was noted (*Table 2, Table 3***Error! Reference source not found.**).

SG and SO groups were subdivided into native (E) and non-native (O) English speakers to enable further analysis. E and O participant response accuracy to questionnaire items dependent upon verbal content only was compared to determine if any language barrier may have impacted comprehension. Accuracy was 67% for E participants in the SG group (ESG) and 64% for O participants in the SG group (OSG). Accuracy was 65% for E participants in the SO group (ESO), and 54% for O participants in the SO group (OSO). These differences were not significant (*Table 4, Table 5*).

Questionnaire responses were grouped into those that related to gestural content and those that did not. In the SG group, 58% of participants responded correctly to story questions reinforced by co-speech gesture. In the SO group, 45% of participants did so. The difference is not considered significant, $p < 0.13$, but is in the expected direction (*Table 7*).

Participant performance in the SR task was assessed by evaluating the participant's SR-P and their mental construct of the scene as explained during the interview (SR-I). For both SR-P and SR-I, objects ordered left to right: Ball → Character → Puppy, as positioned in a group in front of House, were considered correct (*Figure 4*). For example, if during the interview a participant described object positioning correctly, but had failed to correctly position them during the physical task, the SR-I response for that participant was considered correct and their SR-P response was considered incorrect. SR-I and SR-P tasks for both SG and SO groups were coded based on these criteria.

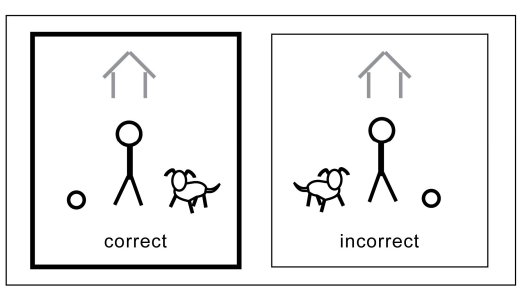


Figure 4: Positioning scored as correct or incorrect in SR tasks. Variances other than correct sequence not depicted were also scored as incorrect.

For the SR-P task, 25% of SG participants correctly completed it while in the SO group, 17% did so. As can be seen in Table 1, for the SR-I task, 67% of SG participants correctly completed it, while 25% of SO group participants did so. The difference in correct SR-I task completion between the groups is considered significant, $p < 0.05$ (*Figure 5*).

Table 1: Comparison of correct SR-I responses between SG and SO groups

	SG	*SO*
Mean	0.6667	0.2500
Variance	0.2424	0.2045
Observations	12	12
P(T<=t) two-tail	0.0420	

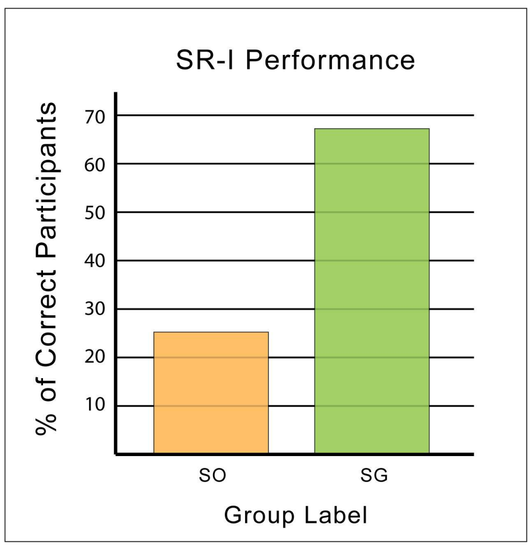
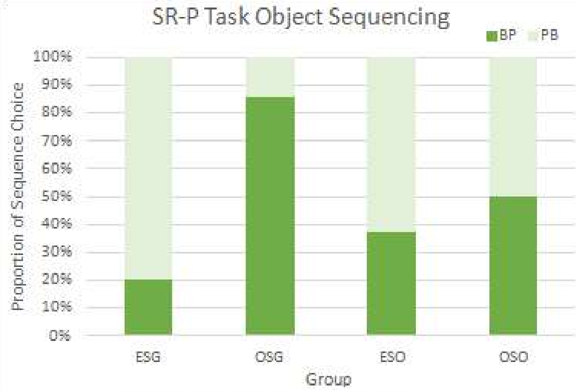
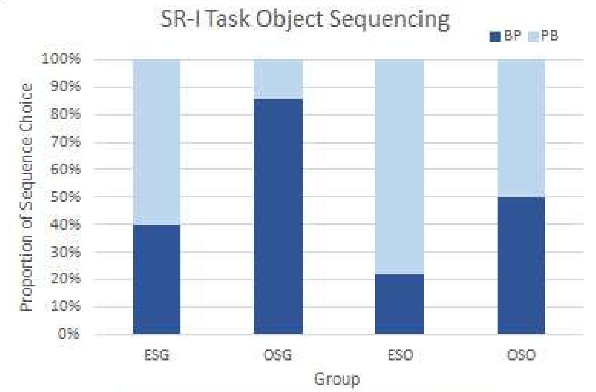


Figure 5: Comparison of SG and SO groups during SR-I task. Difference between groups is significant, p<0.5.

In the SG video, deictic gestures established the positions of the Ball and Puppy. The position of the character was not made explicit through gesture, but could be inferred by the story content. To separate the impact of gesture alone then, an analysis of participant positioning of the Ball and Puppy, regardless of C positioning, were considered. OSG participants were more accurate than ESG participants in recalling the correct sequence, with 86% OSG and 20% ESG participants correct in the SR-P task, and 86% OSG and 40% ESG participants correct in the SR-I task (*Figure 6*).



SR-P			
GROUP	%BP	BP COUNT	TOTAL
ESG	0.20	1	5
OSG	0.86	6	7
SG	0.58	7	12
ESO	0.38	3	8
OSO	0.50	2	4
SO	0.42	5	12



SR-I			
GROUP	%BP	BP COUNT	TOTAL
ESG	0.40	2	5
OSG	0.86	6	7
SG	0.67	8	12
ESO	0.25	2	8
OSO	0.50	2	4
SO	0.33	4	12

Figure 6: SR-P and SR-I Task responses considering only Ball (B) and Puppy (P) placement. Ball to the left and Puppy to the right (BP) is the correct response. The reverse (PB) is incorrect. SG and SO groups performed at chance for this task. Differences between ESG and OSG groups are of note.

The viewpoint a participant assumes within a story scene has direct impact upon their recall of object positioning within the scene, with those assuming an external viewpoint possibly having greater accuracy in recall of element and character positioning (Abelson, 1975). Within SG and SO groups, 42% of participants in each stated during the interview that they had assumed an internal viewpoint at some point when listening to the story. This is in line with prior research positing that an internal viewpoint poses a lower load to working memory, and that participants often assume internal viewpoints when reading narratives that describe space around a character (Bryant et al., 1992). Where participants ultimately chose to place the pointer object in the SR-P however, was considered their dominant viewpoint. In the SG group, 58% of participants positioned the pointer object facing *towards* the house - an external viewpoint. In the SO group, 25% of participants did so. Though the sample size is small, with n = 12 in each case, the difference between groups in this second aspect of perspective is approaching significance, $p < 0.11$.

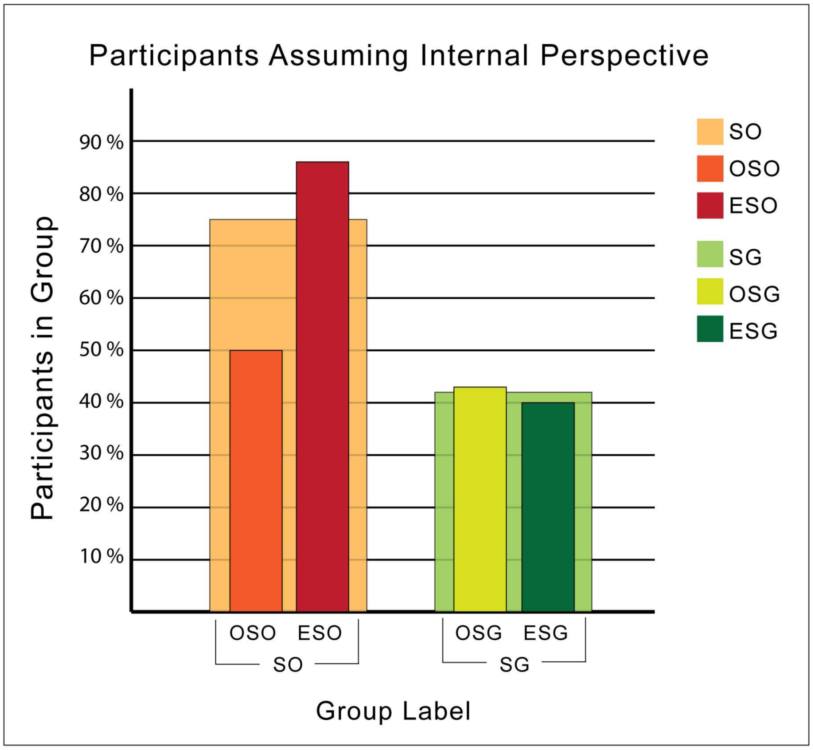


Figure 7: Analysis of participant viewpoint between groups.

Upon closer analysis of the SO group, it was noted that only one of the eight ESO participants had correctly reconstructed the scene during the interview – well below chance. The single correctly responding participant was also the only one out of eight to indicate an external viewpoint, hence viewing the scene in front of them. The remaining seven ESO participants had assumed an internal viewpoint. Six of those seven made the same error of reversing ball and puppy positioning.

DISCUSSION

We hypothesized that listeners who had access to gestural information, the SG group, would demonstrate greater memory for details than the SO group - listeners who had access solely to ambiguous verbal information, facilitated by listener working memory. By using different tasks to assess what listeners remembered, we could identify several factors that seemed to be impacting participant responses.

For questionnaire items, no significant difference was found between the SG and SO groups. Though no ceiling effect was present, it is possible that the questionnaire was not a sensitive enough tool to reveal differences in perception between participants. In which case, further questionnaire revision and testing would be warranted. Additionally, because questionnaire and SR-P tasks were not counterbalanced, in each case the questionnaire was completed after the SR-P. Participants may have been better able to recall story events specifically related to the scene they had just reconstructed. The presence of tangible items in scene reconstruction meant that participants had a visual referent available to them while completing the questionnaire. This may have impacted their recollection of other story elements. Repeating the study with a larger participant group, counterbalancing tasks, and assessing the impact of scene presence or absence after reconstruction would address these concerns.

In the SR-I task, a significant difference between the SG and SO groups was found. The task was designed to reveal individual differences among participants – vital in a participant group that was diverse in both language and cultural background – and succeeded in doing so. The verbally relayed story was intentionally ambiguous in object placement. Were gesture not playing a role, no difference in correctly reconstructing the scene would be expected between the groups. However, this was not the case. Initial analysis supported that listeners viewing gesture had indeed incorporated that information into their memory of the scene. The sensitivity of this task shed light on several differences across participant groups that could be further analyzed.

Within the SG group, where 67% had correctly answered the SR-I task, it was found that of the O participants, 86% had been correct and of the E participants, only 40% had been correct (*Figure 6*). This difference between native and non-native English speakers is far below chance. Even considering the low sample size of ESG (n = 5), this is a significant difference. These results might indicate that within the SG group gesture was a communicative factor impacting memory only among non-native English speakers.

This raised the further questions: what factors may have contributed to non-native English speaker's performance in the SR-I task? Was the OSG subset of the SG group better able to incorporate information from gesture, or was there something else common to the group that gave them an advantage? A weakness in the study is the presence of more than a single variable of difference between groups, which makes difficult causal attribution of differences in performance solely to the presence or absence of gesture.

Among the OSG group, all 4 participants who had assumed an external viewpoint were correct in the SR-I task, compared with 2 out of 3 from the ESG group. It is possible that assuming an external viewpoint correlated with greater accuracy in recall of spatial relationships between objects as was the case in an earlier study (Abelson, 1975). Another consideration is the strong connection to story characters that can develop with an internal viewpoint, which may lead to distortion of recalled details (Bower, 1978). The external viewpoint then, could be considered the more objective of the two.

Of the other OSG group members, 2 out of the 3 who assumed an internal viewpoint did correctly complete the SR-I task, yet none of the ESG members who shared that viewpoint did so. Within the SO group, ESO participants overwhelmingly adopted an internal viewpoint. Additionally, all 7 of the 8 who did so were incorrect in the SR-I task. OSO participants were evenly divided by both viewpoint and correctness in the SR-I task. These results suggest that English speakers may have been biased to assume an internal viewpoint. A follow-up study, analyzing viewpoint alone among a larger group of English speakers would be required for a definitive answer.

Aside from viewpoint, another factor considered was the impact that perception of temporality may have had upon English speaking participants. Story content unfolded temporally, as is the case with all verbally relayed narratives. Participants heard a sequence that first, mentioned a Ball, and second, mentioned a Puppy. When individuals create the mental map of a scene, they have been shown to do so serially – in order of the verbal presentation of information (Tversky, 1991). As participants reconstructed the scene, it is possible they did so by employing a strategy of recalling, in sequence, the major events, and reconstructing step by step accordingly. In a study of English and Mandarin speakers, English speakers responded more quickly to left positioned items in a horizontal pair when it represented "earlier," whereas Mandarin speakers showed both a left bias and a top vertical bias (Boroditsky, Fuhrman, & McCormick, 2011). For English speakers then, it might seem natural to position the first item in a sequence, to the left. O group participants, like the Mandarin speakers in the study, may have been more flexible in their positioning of the first item they recalled.

The directionality conventions of written language may also contribute to this “left first” bias. In one study of how directionality conventions of English, Chinese and Taiwanese writing systems impacted performance on spatial tasks – Taiwanese participants -–exposed to both horizontal writing and vertical writing - showed greater flexibility in sequencing than did English and Chinese participants who were strongly biased to horizontal, left to right sequencing (Chan & Bergen, 2005). Supporting this idea is the account of one participant who, during the interview after the SR-I task, explained that because she was accustomed to reading left to right, it seemed most natural to her to place the objects in sequence from left to right – also the order in which she remembered hearing them introduced. Exploring that impact of gesture on memory through a task that allowed for positioning of items beyond the horizontal plane, and with a difference type of story, would be a way to examine if a directionality bias is always present for English speakers, or if it only arises when verbal information is ambiguous.

CONCLUSION

In this study, unique access to a diverse participant population allowed for some comparison across language and cultural groups. Self-reported native languages across participants included Arabic, Cantonese, Chinese, English, Kazakh (Kazakhstan), Lebanese, Mandarin and Tagalog. Countries of origin included Australia, Canada, China, Egypt, Hong Kong, Kazakhstan, Lebanon, the Philippines, Singapore and the United States. Future studies isolating and comparing subsets of these populations to establish what cultural and language differences are linked to memory of gesture would be of interest, and key in answering more pointed questions about these specific populations.

In a commentary response to Arthur Glenberg’s article, *What memory is for*, Karl MacDorman raised the critique that Glenberg’s discussion of memory failed to account for the role that “internal feedback” – the meshing of newly formed memories with pre-existing experience - plays (Glenberg, 1997). This study, which has raised several questions for further pursuit, has certainly provided an example of that which is pre-existing, interfering with that which is novel – at least for speakers of English.

We found, as hypothesized, that gesture was indeed associated with accuracy in memory for story details. Unanticipated were the differences in how participants assumed perspective within their own story scene visualizations. These differences, emerging not only between those who did and did not view co-speech gesture, but also between native and non-native speakers of English, demonstrate that individual predispositions may impact not only what is encoded in memory, but also perhaps, perception of the same events. Teasing apart the circumstances in which "internal feedback" is a factor strong enough to override novel information, and for whom, are questions to be addressed in future research.

ACKNOWLEDGEMENTS

Thanks to Dr. John Williams, *Department of Theoretical and Applied Linguistics, University of Cambridge* for his supervision and guidance throughout this study, and to the *University of Cambridge,* host institution of the *Pembroke-Kings Programme,* which brought together a diverse range of students from around the world and provided the supportive atmosphere within which this study was carried out. Thanks to Professor Eve Sweetser, *Department of Linguistics,* and Professor Terrence Deacon, *Department of Anthropology, University of California, Berkeley,* for their guidance and mentorship. Finally, thank you to the Gesture and Multi-Modalities Group (GMMG) for their attention, discussion, and critical feedback.

REFERENCES

Abelson, R. P. (1975). Does a story understander need a point of view? (pp. 140–143). Presented at the Proceedings of the 1975 workshop on Theoretical issues in natural language processing, Association for Computational Linguistics.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Baddeley, A. D., 1934- Working memory. (2007). Working memory, thought, and action, *45*, xviii, 412 p., [2] p. of plates.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: the multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory : mechanisms of active maintenance and executive control*. Cambridge University Press.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture Changes Thought by Grounding It in Action. *Psychological Science*, *21*(11), 1605–1610.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*(1), 123–129. https://doi.org/10.1016/j.cognition.2010.09.010

Bower, G. H. (1978). Experiments on story comprehension and recall□. *Discourse Processes*, *1*(3), 211–231.

Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, *31*(1), 74–98. https://doi.org/10.1016/0749-596X(92)90006-J

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 129–143. https://doi.org/10.1098/rstb.2011.0162

Cassell, J., McNeill, D., & McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition*, *7*(1), 1–34.

Chan, T. T., & Bergen, B. (2005). Writing direction influences spatial cognition (pp. 412–417). Presented at the Proceedings of the 27th annual conference of the cognitive science society, Lawrence Erlbaum: Mahwah, NJ, USA.

Cowan, A. D. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory : mechanisms of active maintenance and executive control*. Cambridge University Press.

DeLiema, D., & Sweetser, E. (2016). Rethinking gestural viewpoint as multidimensional rather than a dichotomy. Presented at the ISGS7, Paris, France.

Foster, J. K., & Jelicic, M. (1999). Memory : systems, process, or function? *Debates in Psychology*, xii, 297p.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(01), 1–19. https://doi.org/null

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's Influence on Thought: The Case of Gesture. *Perspectives on Psychological Science*, *5*(6), 664–674.

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/s1532690xci0903_2

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

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334.

Kelly, S. D., Özyürek, A., & Maris, E. (2009). Two Sides of the Same Coin. *Psychological Science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Kendon, A. (2000). Language and gesture: Unity or duality. *Language and Gesture*, (2).

Lavergne, J., & Kimura, D. (1987). Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia*, *25*(4), 689–693. https://doi.org/10.1016/0028-3932(87)90060-1

Rogoff, B., & Waddell, K. J. (1982). Memory for Information Organized in a Scene by Children from Two Cultures. *Child Development*, *53*(5), 1224–1228. https://doi.org/10.2307/1129009

Shah, P., & Miyake, A. (1999). Models of working memory: an introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory : mechanisms of active maintenance and executive control* (p. 506). Cambridge University Press.

Tversky, B. (1991). Spatial mental models. *Psychology of Learning and Motivation*, *27*, 109–145.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*(4), 395–407. https://doi.org/10.1016/j.jml.2004.01.002

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, *114*(4), 591–600.

APPENDIX

	Transcript of STORY		Coding of Gesture	
1	C [[is 5 years old]]	g1:	(RH)	indexical
2	[[and wants to play.]]	g2:	(RH)	abstract
3	But playing [[is not allowed]]	g3:	(RH)	abstract
4	[[inside the house,]]	g4:	(RH)	deictic, model-world-making
5	so [[C runs down the stairs]]	g5:	(RH)	observer viewpoint
6	[[and outside]] into the yard.	g6:	(RH)	deictic, model-world-making
7	[[On one side of the yard is a ball,]]	g7:	(RH)	deictic, model-world-making
8	[[and on the other side, a puppy.]]	g8:	(LH)	deictic, model-world-making
9	[[C walks to the ball]]	g9:	(RH)	deictic, observer viewpoint
10	[[and picks it up,]]	g10:	(LH+RH)	character viewpoint
11	[[then throws it to]] the puppy.	g11:	(RH)	character viewpoint
12	[[The puppy catches the ball,]]	g12:	(LH+RH)	character viewpoint
13	and [[brings it to C.]]	g13:	(RH)	deictic, observer viewpoint
14	[[Before dropping it though]],	g14:	(RH)	abstract
15	[[the puppy runs in circles around]] C, excited.	g15:	(RH)	observer viewpoint

Figure 8: Story transcript and gesture (g) coding, one gesture per transcript line. Double brackets indicate co-speech gesture occurrence. Gestures made by the right hand (RH), left hand (LH), or both (LH+RH) are noted. Coding per Viewpoint Diagram (DeLiema & Sweetser, 2016).

Q 13 - What is the last thing the Puppy did?
***Q 14 - When C went outside, what was to C's left?**
Q 15 - Place these events in the order they occurred in the story.
Drag to position the events so that the first to occur is at the top of the list and the last to occur is at the bottom.

- C throws the ball
- the Puppy runs around C
- C picks up the ball
- the Puppy drops the ball
- C runs down the stairs
- the Puppy catches the ball

Q 16 - Which objects were NOT in the story?
Q 17 - How old is C?
Q 18 - Why did C go outside?
Q 19 - Are the stairs inside or outside of the house?
Q 20 - At the end of the story, where is C standing?
***Q 21 - Where was the Puppy standing when C entered the yard?**
Q 22 - After bringing the ball to C, what did the Puppy do?
Q 23 - What did C do outside?

*Figure 9: Story sequence and event questions. * Indicates information conveyed only through gesture.*

The Experimenter will now provide you with several objects:
A ball, a house, a puppy, and the character "C."
Using these objects, re-create the story scene as you remember it after C exited the house.
Take as much time as you need to complete this task. When finished press the button to proceed.

The Experimenter will now provide you with another object that indicates direction.
Position this object in the scene area you've created to indicate where YOU are. That is, when you imagine this scene, from what position are you viewing it in your mind's eye? Place the marker object in that spot, pointing in your direction of view.
When you have completed this, please continue.

Figure 10: SR-P task instructions

Experimenter sets up camera, records interview (with consent)

Experimenter asks the Participant the following questions, allowing Participant to elaborate and talk freely.
Question : Were you able to understand the story?
Question : Was there any part of the story that was unclear or confusing?
Question : Referring to your scene recreation, please explain why you've placed each figure where it is.
(*Ensure Participant explains each Figure location: Child, Ball, Puppy, House, Observer* .)
Question : Were you unsure about where any of the figures should be placed?
Question : Did you have a strong mental image of the scene before recreating it here, or did you piece it together while performing this task?
Question : Is there any difference between the scene you've recreated and the scene you imagined while listening to the story?
Question : What perspective did you take when listening to the story? (inside the character, an observer viewpoint)
Question : Is that the same perspective you've recreated in your scene?
Question : Do you think the video had an impact on your perspective?
Question : When you hear stories about others, do you tend to put yourself in their shoes?
Question : Please use your own words to relate the story you heard back to me.
Question : What do you think the purpose of this experiment was?

Question : Do you have any questions?

Figure 11: Interview Questions

raincoat	cabbage	key	mirror	socks
octopus	sofa	broccoli	noodle	motorcycle
banana	grapes	turtle	horse	kangaroo
pepper	building	snake	window	scorpion
bookstore	snail	chicken	hospital	carrot

Figure 12: Word list used in verbal memory test.

Table 2: Results of Auditory Memory Test

	SG	*SO*
Mean	10.16667	10
Standard Deviation	2.852874	2.768875
Observations	12	12
P(T<=t) two-tail	0.890683	

Table 3: Results of Visual Memory Test

	SG	*SO*
Mean	14.66667	14.33333
Standard Deviation	3.009245	2.808717
Observations	12	12
P(T<=t) two-tail	0.790761	

Table 4: Accuracy Analysis between ESG and OSG on verbal-only dependent questions

	ESG	*OSG*
Mean	0.666666667	0.642857
Variance	0.027777778	0.003968
Observations	5	7
P(T<=t) two-tail	0.733547288	

Table 5: Accuracy Analysis between ESO and OSO on verbal-only dependent questions

	ESO	*OSO*
Mean	0.645833333	0.541667
Variance	0.074900794	0.0625
Observations	8	4
P(T<=t) two-tail	0.538072939	

Table 6: Accuracy Analysis between SG and SO groups on verbal-only dependent questions

	SG	*SO*
Mean	0.685185185	0.611111
Variance	0.01758324	0.059484
Observations	12	12
P(T<=t) two-tail	0.365347354	

Table 7: Accuracy analysis between SG and SO groups on gesture dependent questions

	SG	*SO*
Mean	0.583	0.450
Variance	0.247	0.252
P(T<=t) two-tail	0.146	

Table 8: SR-P Analysis

SR-P					
GROUP	%BP	BP COUNT	%PB	PB COUNT	TOTAL
ESG	0.20	1	0.80	4	5
OSG	0.86	6	0.14	1	7
SG	0.58	7	0.42	5	12
ESO	0.38	3	0.63	5	8
OSO	0.50	2	0.50	2	4
SO	0.42	5	0.58	7	12

Table 9: SR-I Analysis

SR-I					
GROUP	%BP	BP COUNT	%PB	PB COUNT	TOTAL
ESG	0.40	2	0.60	3	5
OSG	0.86	6	0.14	1	7
SG	0.67	8	0.33	4	12
ESO	0.25	2	0.88	7	8
OSO	0.50	2	0.50	2	4
SO	0.33	4	0.75	9	12

Table 9: SR-I Task Object Placement in sequence left to right, house positioned behind, BCP = Ball, C, Puppy

GROUP	BCP	CPB	PBC	BPC	PCB	CBP	TOTAL
ESG	2	0	2	0	1	0	5
OSG	6	0	0	0	1	0	7
SG	8	0	2	0	2	0	12
ESO	1	0	2	0	4	1	8
OSO	2	0	0	0	2	0	4
SO	3	0	2	0	6	1	12
TOTAL	11	0	4	0	8	1	24
% OF TOTAL	0.46	0.00	0.17	0.00	0.33	0.04	
% OF SG	0.67	0.00	0.17	0.00	0.17	0.00	
% OF SO	0.25	0.00	0.17	0.00	0.50	0.08	