Language and Gesture Descriptions Affect Memory: A Nonverbal Overshadowing Effect

Mark Koranda (mjkoranda@wisc.edu)
Department of Psychology, University of Wisconsin-Madison,
1202 W. Johnson Street, Madison, WI 53706 USA

Maryellen MacDonald (mcmacdonald@wisc.edu)
Department of Psychology, University of Wisconsin-Madison,
1202 W. Johnson Street, Madison, WI 53706 USA

Abstract

People's memory for an event is known to be affected by their verbal descriptions prior to memory assessment. The present experiment investigated whether the computational difficulty of production itself, which is known to affect what people say, can shape descriptions and subsequent event memory. Participants viewed simple scenes and were asked to describe them using either speech or silent gesture, the latter being a much more difficult task. We hypothesized that gesturing participants would over-use action pantomimes, which would yield poorer Inaction (vs. Action) scene memory. Following scene descriptions, participants were given a forced-choice recognition task to discriminate previously presented scenes from foils. Patterns of gesturing showed that gesturers used action pantomimes for Inaction scenes, and they performed reliably worse on Inaction scene memory. Increased production of action pantomimes predicted increased guesses for Action scenes at test, independent of the correct response. Implications for memory and production are discussed.

Keywords: memory; eyewitness memory; language production; gesture analysis; experimental research with adult humans

Introduction

Everyone knows (that is, we all remember) that episodic memory is fallible, and there appear to be a number of sources of memory disruption in the interval between the initial perception of an event or stimulus and its later recall. One potential source of disruption is talking about a recently perceived event, which appears to influence recall of details of the event or its participants at a later test. For example, Schooler and Engster-Schooler (1990) found that describing a person's face led participants to poorer forced choice recognition of that face (as in selecting someone from a police lineup). This *verbal overshadowing* phenomenon appears to reflect the linguistic description of a nonlinguistic percept (the face) interfering with the fine detail of the visual memory that would support later recognition of the face. These effects of description on later memory suggest that any instructions or contextual cues that affect the nature of a speaker's description could also affect that person's later memory. These effects appear to be examples of the role of verbal labels on perception and categorization, in which the label makes representations more categorical (e.g., Lupyan, 2012). On this view, the labeling of events, facial features, or other percepts that occurs when someone reports an experience could have downstream consequences for the accuracy of later memory of the experience. These effects might improve memory in some cases and impair it in others. For example, Marsh, Tversky & Hutson (2005) instructed participants to describe either the emotional content or the events of a just-watched video. They found that in comparison to the event-description group, the emotion-description group later had better recall of the emotional content of the film but poorer recall of the events (see also Soleti, Curci, Bianco & Lanciano, 2012).

These emotion- vs. fact-reporting results arise from explicit instruction to participants to emphasize certain aspects of an event, but conversational contexts, which more implicitly bias speakers to tailor the nature of their spoken description, also appear to affect what is later retained about the event. For example, Hellmann Echterhoff, Kopietz, Niemeier and Memon (2011) showed participants a video depicting the actions of a person and had participants describe the video. There was no explicit instruction for what to describe, but the participant was told that the audience for their description either did or did not like the character in the video. Hellmann et al. found that the more participants tuned their descriptions to the audience's perspective, the more their later recall was predicted by what they had included in their event descriptions.

These effects of description on memory are interesting in light of language production research suggesting that speakers' implicit choices of how to convey their messages are influenced by inherent computational demands of language production itself. Producers appear to apply several implicit effort-minimization strategies during the process of settling on a plan for what to say. These strategies include placing highly salient words early in the sentence (termed Accessibility or Easy First, e.g. Bock 1982; MacDonald, 2013). A second implicit strategy is the repetition of recently produced or comprehended sentence structures (Syntactic priming or Plan Reuse, Bock 1982; MacDonald, 2013). For example, Fausey, Snider & Boroditsky (2008) primed participants with either an agentive sentence structure (*He opened the umbrella*) or a non-agentive structure (*The umbrella opened*) and then had

participants describe pictures showing an object's change of state (e.g. a picture of an intact vase followed by a picture of its broken state). Fausey et al. measured the rate of agentive descriptions (*Someone broke the vase*) vs. non-agentive descriptions (*the vase broke*) and found that participants were more likely to use the primed sentence structure, compared to no-prime control group. Fausey et al. did not test later memory, but the descriptions they investigated—whether a person was responsible for an adverse event—clearly aligns with issues of interest in the eyewitness memory literature.

As a first step in investigating the role of communication difficulty on what is said and potentially what is remembered, we designed a very robust manipulation of production difficulty, whether descriptions should be spoken or conveyed via gesture, with no speaking allowed. There is a tradition of comparison between speech and gesture in the psycholinguistic literature (e.g. Goldin-Meadow, McNeill & Singleton, 1996), although not typically with attention to consequences for memory.

Using gesture to convey the contents of a video or picture is clearly much harder than producing a spoken description, and the gestures differ from language in several interesting ways that could have downstream effects on memory. First, the order of gestures is often different than in speech: Goldin-Meadow et al. (1996) found that English-speaking gesturers tended to sequence their gestures in a Subject Object Verb order rather than the typical English order Subject Verb Object (see also Gibson, Piantadosi, Brink, Bergen, Lim & Sacks, 2013; Hall, Mayberry & Ferreira, 2013).

Second, unlike natural language, gesturing affords participants the opportunity to depict information *synthetically* (McNeill, 1992), in that a single gesture can depict multiple semantic meanings. For example, to convey a scene where a man is throwing a football, a gesturer might pantomime the motion of throwing a football with a football-holding handshape, simultaneously presenting information about the action (throwing) and object (football). Hostetter and Alibali (2008) propose that part of how we choose to gesture while speaking comes directly from prior manual experience with an object. In line with this view, we predict that synthetic or "fused" action-object gestures will be used frequently to convey the stereotypical use of an object (throwing a football, talking on a phone). That is, an excellent way to convey a football in gestures is to pantomime throwing it. That gesture also conveys a throwing event, which is appropriate if the actual event to be described contains throwing. However, if an event has holding (but not throwing) a football, the action-relevant pantomime of football-throwing is inappropriate for a scene that does not contain any throwing.

We hypothesized that in this situation containing an object but not its characteristic action, gesturers will often resort to pantomiming an action, and that this inclusion of an action gesture will have consequences for later memory of the event. In other words, we expect a sort of non-verbal overshadowing, on analogy to Schooler and Engster-Schooler's (1990) verbal overshadowing, in which verbal descriptions impaired later forced choice recognition. Here, we predict that nonverbal gestures will, for one type of scene, impair later recognition.

To test this hypothesis, we showed participants two types of scenes: ones with agents using objects in canonical ways presumably easy to pantomime (*Action* scenes), and scenes with agents engaging objects in ways less clearly communicable via pantomime (*Inaction* scenes). We predicted that for both types of scenes, gesturers would frequently pantomime a canonical action in an attempt to convey the object. We further predicted that as a result of this use of the fused action-object gestures, gesturers would tend to mis-remember the Inaction scenes as containing canonical actions. To test this prediction, a surprise recognition task required participants to identify the previously-viewed scene out of two choices. The memory performance of gesturers was compared to that of a group who provided spoken descriptions of the scenes. Action and Inaction scenes are predicted to both be easily described in speech, and so we predicted no effect of scene type on recognition performance for the spoken description group.

Method

Participants

Thirty-nine University of Wisconsin-Madison psychology undergraduate students participated for course credit. Six subjects' production (fused gesture) data were incomplete or missing due to technical difficulties (n=6), however data were retained and included in analyses as available.

Materials

An online cartoon generator (pixton.com) was used to create 200 single-panel cartoon scenes. All scenes depicted one human agent engaging in an activity with one object, and contained minimal additional background (e.g. chairs or tables in indoor scenes, trees or bench in outdoor scenes). Four examples are shown in Figure 1. Ten agents (5 male, 5 female) engaged with 10 inanimate objects each of two ways. In the Action scene, the agent interacted with the object in a highly common activity for that object, as in scene A (drinking water) or C (eating an apple) in Figure 1. The paired Inaction scenes showed the same character and object but without the common activity. In these scenes, (B and D in Figure 1), the character simply held or looked at the object. Each event (e.g. throwing a football, holding an apple) was created with all 10 agents, yielding 10 different scenes showing the same action, differing in the agent and subtle aspects of the background.

Eight counterbalanced lists were prepared so that each list contained 20 of the 200 scenes, chosen to satisfy several criteria. First, across the 20 pictures, each human agent appeared twice, once in an Action scene and once in an Inaction scene. Second, each object (e.g. a football)

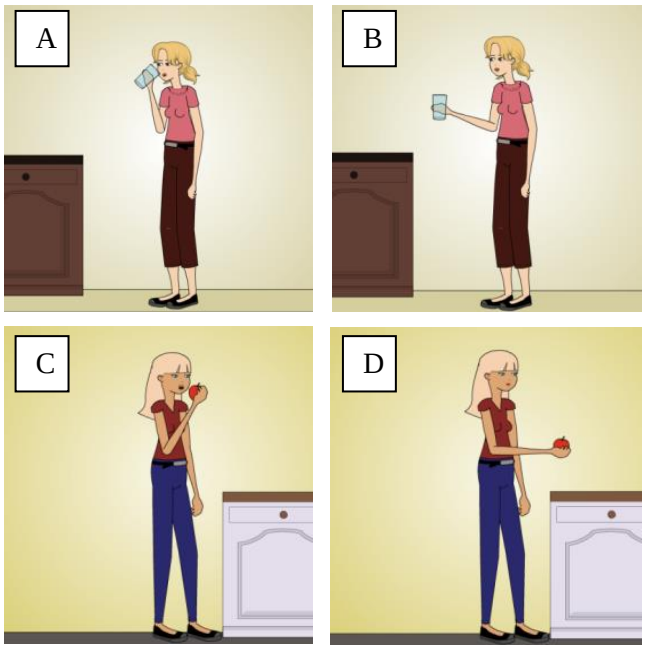

Figure 1: Example action and Inaction scenes. A and C are example scenes participants might see during trial. At test, A and C would be paired with foils B and D respectively. Successful recognition would require identifying the correct action associated with the object and agent.

appeared twice, once in an Action scene, once in an Inaction scene and that these two paired pictures would have two agents of the same gender. For example, List 1 contained scene A (drinking water) from Figure 1 and a second scene (not shown) of another female agent holding a glass. The eight counterbalanced lists covered many but not all possible agent-object-action combinations. An additional six practice scenes, similar in format to the experimental items, were placed at the start of each list.

Procedure

Participants were randomly assigned to the speaking or the gesture condition. Participants were instructed that they would see cartoon scenes on a computer screen, and that their task was to convey the event depicted in the scene, speaking or using gestures, as appropriate to the assigned condition. Participants were informed that their gestures or spoken descriptions would be videotaped.

After the instructions were given, participants completed six practice trials. The practice scenes were designed to familiarize them with the drawing style of the scenes and type of characters, objects, and actions that occur in the experimental items. For the participants in the gesture condition, the practice items also provided an opportunity for experimenter feedback and additional instruction on amount of information to report on the scene, as pilot testing revealed that some participants were reluctant to gesture. During practice trials, both groups of participants received feedback from the experimenter if gesture or verbal production was fewer than two words or gestures, or more than ten. Clarification was allowed if the participant was still unclear as to what the appropriate amount of information was required.

Each trial began with a scene appearing on screen for three seconds, after which a blank screen with a question mark appeared, cueing the participant's production. For practice trials, the experimenter advanced the scene to the next trial, and in experimental trials, the participant advanced to the next scene upon completing the spoken or gestured description of the scene. Scenes were presented in random order.

After the description task was completed, participants completed a two minute distraction task consisting of two-digit multiplication problems.

Participants next completed a forced-choice recognition task, where they were presented with one picture that they had seen in the earlier description task together with its agent-object pairmate—the same pictured agent and same object, but different action. For example, a participant who had seen Picture A from Figure 1 in the description task would see Pictures A and B side by side in the recognition task, and they had to choose which picture they had seen previously. The left vs. right position of the previously-seen picture and foil and the order in which scenes were tested was randomized.

Results

Coding Transcripts of spoken production were made for participants in the spoken condition, and gestures were coded for the presence of a fused gesture that combined an action and object in pantomime, such as throwing a football, eating an apple, or drinking water. For the purposes of coding, video clips of each gesture production were cropped to remove the portion showing the to-be-described scene to allow for coders to be blind to the scene condition in which the gestured description was given. Clips and sentences were then randomized to reduce bias in coders' ratings.

Two coders were trained to identify fused gestures, defined as a gesture that simultaneously conveys information about both an action and object being acted on. A fused gesture is typically one in which the hand shape gives information about the object's shape and the motion of gesture provides information about its use. To avoid subjective judgments of action-object fusion, one fused gesture template for each object was identified *a priori* and only gestures that conformed to this template were coded as fused gestures (e.g., for *apple* scenes, only a biting pantomime with hand at the mouth was counted), while other fused gestures were excluded (e.g., pantomiming picking up, or polishing an *apple* against a shirt). For one object, book, the *a priori* gesture template was recoded to reflect the more common gesture in participants' productions, but in all cases only one template was used. In contrast, a non-Fused gesture, for example, could be a single finger moving in an arc showing the path of a football, or

finger tracing to denote the outline of a football. Intercoder agreement was 91% in identifying fused gestures.

Fused gesture rates Figure 2 shows the proportion of fused gesture for the Action and Inaction scenes. As can be seen from the figure, the rate of fused gestures is high in both the Action condition, where the gesture is appropriate (a football-throwing gesture is appropriate for a scene showing someone throwing a football) and in the Inaction scenes (e.g., holding a football), where indicating a throwing action is not accurate. Participants who made fused gestures for Inaction scenes typically added extra gestures afterwards to convey that there was no canonical action (such as throwing) depicted in the scene. We did not compare fused gesture rates in the two scene conditions; the important point is that in Inaction scenes, where an action yields a misleading description, gesturers nonetheless produced action pantomimes more than ¾ of the time. We next test whether the use of fused gestures leads to errors in memory.

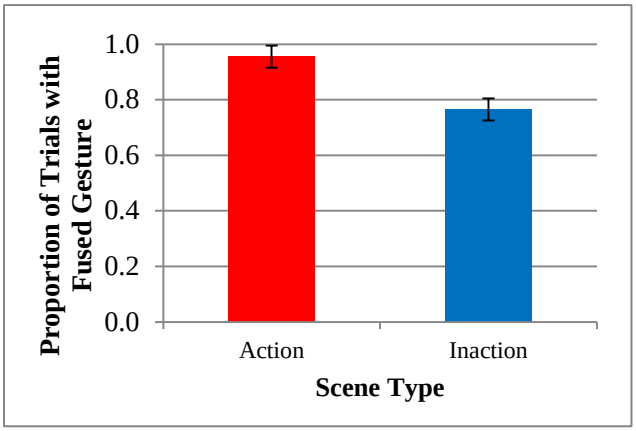

Figure 2: Proportion of gesture productions that included the target fused gesture, by scene type. No statistical analyses were conducted as this information served a descriptive purpose.

Forced-Choice Recognition The proportion of correct recognition can be seen in Figure 3. To evaluate recognition performance and differences from chance, we used a multi-level model analyzed with REML estimation procedure as per the recommendation of Judd (2013). Accuracy of forced-choice was evaluated with a generalized linear mixed effects model using the binomial family with the logit link function (Jaeger, 2008), and reported in *z* scores. Mixed effects models to determine difference from chance were adapted to simulate one-sample *t*-tests compared against chance, in this case .5 accuracy.

As can be seen in Figure 3, all four combinations of Scene Type and Production Modality yielded above chance performance (chance = .5, all $t > 2.37$, $p < .05$). We predicted that scene type (Action vs. Inaction) would

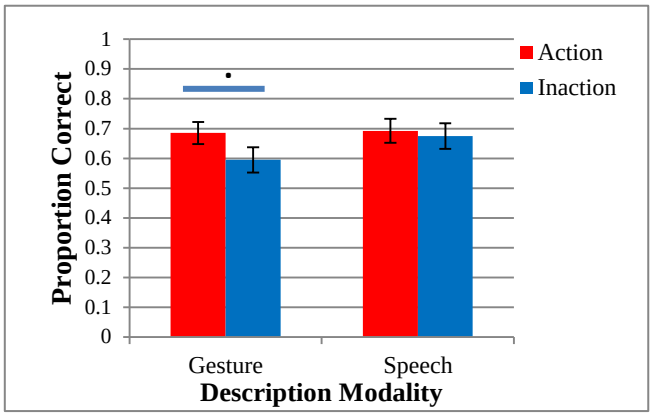

Figure 3: Effect scene type and production modality on forced-choice accuracy (all values reflect raw data). Log-odds were calculated using a generalized liner mixed effects model and transformed to proportion correct. Error bars reflect standard error.

interact with production modality (Gesture vs. Speech), such that gesturers would be particularly inaccurate on Inaction scenes. As can be seen in Figure 3, the numerical values are consistent with this prediction, but the interaction was not reliable. Our hypothesis led us to investigate accuracy in the gesture condition, where memory for Inaction scenes trended lower relative to Action scenes ($\beta = -.41$, $SE = .22$, $z = -1.87$, $p = .06$).

Fused gesture and Scene type In order to determine whether recognition accuracy was related to gesturers' use of fused gestures (which are accurate for Action scenes and misleading for Inaction scenes), we used a regression model of recognition accuracy with fused gesture use and scene type as predictors. A regression analysis shows that controlling for the use of fused gesture scene type significantly predicts accuracy in the direction hypothesized ($\beta = -.48$, $SE = .24$, $z = -1.97$, $p < .05$; see Figure 4). Memory for Action scenes were significantly above chance ($t > 2.37$, $p < .05$) while memory for Inaction scenes were not significantly different from chance. The proportion of trials in which fused gestures were not used was marginal (Action, n=5, Inaction, n=40), and Fused production did not significantly predict accuracy.

However, accuracy is somewhat independent of whether fused production motivates selection of Action scenes on test (for 50% it is the correct choice). A simple regression analysis was performed between subjects comparing the proportion of fused gestures produced and proportion of guesses for the Action scene on the forced-choice recognition task. Participants who produced fused gestures more frequently also more frequently selected action scenes at test, $\beta = .48$, $SE = .20$, $t = 2.45$, $p < .05$.

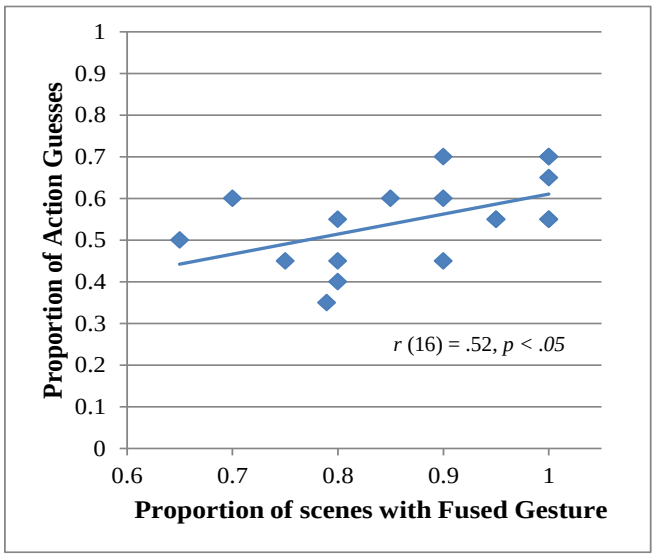

Figure 4: Relationship between participants' proportion of fused gesture productions (out of 20 trials) and their guesses for Action scenes at test. The more frequently a participant produced a fused gesture, the more likely they chose the action scene in the recognition test.

Discussion

This study aimed to investigate the effect of scene descriptions on later memory for the scene. In particular, we investigated whether providing a description in gesture rather than speech would have particular effects on scene memory. Gesturers frequently use pantomime to simultaneously convey an action and object (fused gestures such as throwing a football), and we found that even in the absence of extensive action in a scene, gesturers often resorted to using fused gestures to convey the existence of an object, such as pantomiming throwing a football to indicate that a football exists in a scene. We hypothesized that this use of fused gestures in Inaction scenes would disrupt later memory and found tentative support for that view—gesturers were not above chance in recognition accuracy for Inaction scenes, but all other conditions did yield above-chance recognition. One cautionary note is that the interaction of scene type and modality was not reliable, however our predictions concerning fused gestures and accuracy were supported. Future work is needed to better understand the role of gesture form in memory across scene types.

The use of fused gestures does not reliably predict misremembering of the respective scene, however it does reliably predict an increase in guesses for Action scenes. For approximately every two fused gestures produced, participants guessed an additional Action scene at test. This provides preliminary evidence that Fused gesture productions play a role in the consistent misremembering of gesture but not speech participants.

This work is consistent with previous results on the effect of description on subsequent eyewitness memory (e.g., Hellman et al., 2011; Marsh et al., 2005; Schooler & Engstler-Schooler, 1990; Soleti et al., 2012) and may extend those results in several ways. First, the effects in the current experiment were obtained without any instruction on what participants should convey or any information about the audience that could shape the content of their description; instead both gesturing and speaking participants were told simply to depict what was in the scene. Second, these results show that the modality of the description and more specifically, the ease of conveying information affects the nature of the description (the use of fused gestures) and consequently the memory for the information. This work is consistent with prior research suggesting that the difficulty of conveying a message affects exactly how the message is conveyed. Whereas previous research has examined how language production difficulty affects the sentence structure that is used in spoken or written production (e.g. Bock, 1982; MacDonald, 2013), the present results may show that the same effect holds in nonlinguistic communicative gestures, where the difficulty of conveying an object such as a football causes gesturers to produce a fused action-object gesture even when no action is present, and with downstream consequences for event memory.

The effects of gesture on memory also extend Schooler & Engstler-Schooler's (1990) verbal overshadowing effect, in which describing a visual percept (in their case a face) reduced force choice recognition for the face at a later time. The results here show that the overshadowing effect need not be verbal, and merely conveying the information through action gestures can yield more tendency to assume that an action was presented (i.e. a higher rate of choosing action scenes in the recognition test)—a sort of nonverbal overshadowing. Also note that the memory effect here is not that gesturers were simply worse than speakers; rather, their lower performance was limited to the Inaction scenes, where fused gestures are both common and misleading. This result suggests that the "overshadowing" effects may come from the fact that communicating a message is inherently selective. If we say, for example, "A girl is drinking water", this description is a useful description of Scene A in Figure 1, but it leaves out extensive information from even this simple scene, including more about the girl (clothes, hairstyle age, the angle of her arm), the background, etc.. The overshadowing effect may be that, for speech and gesture, the communicative elements that are produced are reinforced at the expense of unmentioned elements. Note, however, that this study was not designed to test whether spoken descriptions also reduced recognition memory, in that there was not a no-description control group. It is not expected that the participants' short spoken descriptions of these simple scenes would have had substantial effects on memory given the foils in the present study, but it is possible that with more challenging foils, such as differing from the original only by subtle changes to angle of an arm,

that spoken descriptions would impair visual memory, as in Schooler & Engstler-Schooler's (1990) study.

The pattern of results in the present experiment also raises the interesting question of whether these effects of gestured-retelling are or are not examples of labeling making representations more categorical (e.g. Lupyan, 2012). Lupyan's hypothesis is about the effect of verbal labels, whereas the present results arise from nonverbal gestures, and moreover, the fused gestures are produced in the Inaction scenes not with the intent to label the action in a scene but with the goal of trying to convey the object in an Inaction scene. The present study was not designed to identify whether gesturere's representations became more categorical, and future research is necessary to link the memory results here with effects of verbal or non verbal labeling.

The linkage between production difficulty, producers' implicit choices in how to convey a message, and later memory offers several avenues for future research. For example, in picture descriptions, Fausey et al. (2008) found that sentence structure choices like the agentive (*Someone broke the vase*) vs. non-agentive forms (*The vase broke*) were modulated by the presentation of a prime sentence structure, even though the content of the prime sentence was unrelated to the picture. Although Fausey et al. did not test later memory, their result offers a potential mechanism for at least some effect of leading questions on later memory (e.g. Loftus, 1975), in that a leading question (e.g. *Who broke the vase?*) offers not only a potentially misleading semantic interpretation of the event, but it also introduces a sentence structure (the agentive form) that could promote retelling of the event in the agentive form.

Second, the present work may have relevance to children's eyewitness memory. Compared to adults, children have poorer eyewitness memory (e.g. Bruck & Ceci, 1999), and children also are still developing language production skills, meaning that language production is harder for them than it is for adults. Children and adults do make different implicit description choices for how to describe pictures, including different sentence structures, and different rates of mention of event participants (Montag & MacDonald, 2015). More work is necessary to address these possibilities, but it may be that these different description patterns in children and adults could be another influence on the different memory abilities in these groups.

References

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.

Bruck, M. & Ceci, S.J. (1999). The suggestibility of children's memory. *Annual Review of Psychology, 50*, 419-439.

Fausey, C. M., Snider, N., & Boroditsky, L. (2008). Causal priming: How a language production mechanism guides representation. In *Proceedings of the 30th annual meeting of the Cognitive Science Society,* (pp.1130-1135). Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2008/pdfs/p1130.pdf

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science, 24*, 1079–1088.

Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychological Review, 103*, 34-55.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition, 129*, 1–17.

Hellmann, J. H., Echterhoff, G., Kopietz, R., Niemeier, S., & Memon, A. (2011). Talking about visually perceived events: Communication effects on eyewitness memory. *European Journal of Social Psychology, 41*, 658–671.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*, 495–514.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language 59*, 434–446.

Judd, C.M., Westfall, J. & Kenny, D.A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*, 54–69.

Loftus, E.L. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*, 560-572.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00054

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4:226.* doi: 10.3389/fpsyg.2013.00226.

Marsh, E.J., Tversky, B. & Hutson, M. (2005). How eyewitnesses talk about events: implications for memory. *Applied Cognitive Psychology 19*, 531-544.

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

Montag, J.L. & MacDonald, M.C. (2015). Text exposure predicts spoken production of complex sentences in eight and twelve year old children and adults. *Journal of Experimental Psychology: General, 144*, 447-468.

Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36–71.

Soleti, E., Curci, A., Bianco, A., & Lanciano, T. (2012). Does talking about emotions influence eyewitness memory? The role of emotional vs. factual retelling on memory accuracy. *Europe's Journal of Psychology, 8*