Does Extraneous Perception of Motion Affect Gesture Production?

Autumn B. Hostetter (ahostett@kzoo.edu) and Kira Boneff (kboneff@emich.edu)
Department of Psychology, Kalamazoo College, 1200 Academy Street
Kalamazoo, MI 49006 USA

Martha W. Alibali (mwalibali@wisc.edu)
Department of Psychology, University of Wisconsin-Madison, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

Speech-accompanying gestures vary depending on features of the communicative situation. In the present study, we examined whether they might also be affected by extraneous activity in the speaker's sensorimotor system. We asked participants to describe short animations that involved vertical motion while simultaneously watching a display that depicted vertical motion in either a congruent or an incongruent direction. Speakers produced gestures depicting vertical motion at a higher rate when describing the target motion events when they were simultaneously watching a display that depicted motion in the same direction than when watching motion in the opposite direction. These results suggest that the cognitive basis of gesture lies in the sensorimotor system.

Keywords: gesture; perceptual simulation; embodied cognition

Introduction

Gestures are ubiquitous in communication. How much speakers gesture in a particular situation depends on the speakers' cognitive load (e.g., Hostetter, Alibali, & Kita, 2007), as well as on features of the communicative situation (e.g., Alibali, Heath, & Meyers, 2001). Such effects suggest that gestures are formed from a communicative intention to describe the gestured information. The present study investigates whether gesture production is also affected by activation of the sensorimotor system more broadly—including activation that is extraneous to the speaker's communicative intent.

There is now ample evidence that language comprehension involves the activation of sensorimotor information (Glenberg & Kaschak, 2002; Speed & Vigliocco, 2013; Zwaan, Stanfield, & Yaxley, 2002). Participants' reaction time to judge a sentence as making sense or not is affected when their motor (Glenberg & Kaschak, 2002) or perceptual systems (Kaschak et al., 2005) are simultaneously engaged in a way that mirrors the meaning of the sentence. For example, Kaschak, Zwaan, Aveyard, and Yaxley (2006, Experiment 2) asked participants to listen to a sentence that implied motion in a particular direction (e.g., The rocket blasted off) while simultaneously listening to an auditory percept of motion in a particular direction (e.g., up vs. down). They found that participants were faster to say that the sentence made sense when they were listening simultaneously to an auditory percept that sounded as though it was moving in the same direction as that implied in the sentence. Other studies have found that activation of the sensorimotor system actually inhibits the ability to use the system to process language simultaneously (e.g., de Vega, Moreno, & Castillo, 2013). Whether activation of the sensorimotor system results in facilitation or inhibition of language processing seems to depend on task demands and on the relative timing of the sensorimotor stimulation and language comprehension (Diefenbach, Rieger, Massen, & Prinz, 2013; Kaschak & Borreggine, 2008). Regardless of whether the result is facilitation or inhibition of language processing, the evidence suggests that activating the sensorimotor system recruits the same processing mechanisms that are involved in language comprehension.

Although the activation of sensorimotor information in language comprehension is well documented, there is less direct evidence for the activation of sensorimotor information in language *production*. Nonetheless, it is an increasingly common assumption that speakers activate sensorimotor information pertaining to the content of what they wish to speak about, just as language comprehenders activate such information about the content of what they read or hear (e.g., Perlman, Clarke, & Johannson Falck, 2015; Pickering & Garrod, 2013). Indeed, one often cited source of evidence for this position is that speakers frequently accompany their speech with hand and arm movements that depict the sensorimotor content of what they are describing (e.g., Glenberg & Gallese, 2012). For example, a speaker who describes a balloon moving up might move her hand upwards as she talks. On the surface, such movements appear to support the idea that speakers activate their sensorimotor system in the interest of language production, because clearly, gestures are a motor act: to produce a gesture requires activation of the sensorimotor system in the interest of planning and executing the movement. However, it is less clear whether the sensorimotor system only becomes activated after the intention for a gesture has been formed, or whether the activation in the sensorimotor system is actually part of what elicits the gesture in the first place. Indeed, theories about the cognitive origin of speech-accompanying gestures disagree on this point.

Theories about the cognitive origin of gesture vary in terms of whether they view gestures as originating primarily from linguistic processes or primarily from visuospatial processes. On the side of linguistic processing, Butterworth and Hadar (1989) propose that gestures arise from lexical items. When a speaker gestures, he or she activates the words to be spoken and then plans and produces a gesture based on the semantic features of the words. Along similar lines, Wagner, Nussbaum, and Goldin-Meadow (2004) provide some evidence that gestures may be based on propositional, rather than spatial, representations. They found that producing gestures (specifically, pointing gestures) did not interfere with spatial processing on another task, as might be expected if the act of producing the gesture required spatial resources. If gestures emerge from lexical or propositional representations, then the sensorimotor system would only become activated after the intention for the gesture is formed.

In contrast, there are several theories that propose a visuospatial origin for gesture, in which gestures arise from visuospatial imagery that is formed when conceptualizing information for speaking. For example, de Ruiter's (2001) Sketch Model proposes that visuospatial images are formed alongside preverbal messages in the interest of communication. Features of these visuospatial images are selected, depending on the speaker's communicative goals, and expressed alongside speech as gesture. Similarly, Kita and Özyürek's (2003) Interface Model proposes that gestures arise from spatial features held in visuospatial working memory that are selected by an action generator for the purposes of communication. A communication planner "decides what modalities of expression should be involved" and engages gesture if deemed relevant. This view assumes that the sensorimotor system is activated very early during speech planning, and is part of the representation that gives rise to gesture.

This view that gestures arise from sensorimotor activation is articulated most explicitly in the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). The GSA framework considers gestures to be outward manifestations of simulated actions and perceptions that are formed in the interest of communicating. Under this view, speakers use their motor and perceptual systems to support conceptualization for speaking, just as they do to support language comprehension. This motor and perceptual activity is the precursor for speech-accompanying gestures, which are only expressed as gestures if they are activated with enough strength to surpass the speaker's current gesture threshold, which is the minimum level of activation that must be exceeded in order for a gesture to be produced. The gesture threshold is hypothesized to vary from speaker to speaker and to change from moment to moment during speaking depending on factors such as how helpful a gesture would be for communication, the speaker's beliefs about the utility of gesture, and whether the speaker's cognitive system is currently taxed. Thus, in the GSA framework, the speaker does not have to decide whether to gesture or not; rather, gestures arise spontaneously from activation in the sensorimotor system that is present at the moment of speaking.

The GSA framework therefore predicts that gestures should be the byproduct of more than just a communicative plan; specifically, activating motor and perceptual representations in the speaker's mind at the moment of speaking should lead to increased gesture about those representations, even if the activation is not related to communicative intent. There is some evidence for this claim. Chu and Kita (2016) found that speakers gestured more about a mug that afforded physical action than about a mug that did not afford physical action (because it was lined with spikes). Although the communicative intent was the same in both situations (to describe the orientation of the mugs), speakers gestured more when the representations they were describing allowed potential actions. Similar results were obtained by Hostetter and Alibali (2010), who found that speakers gestured more when they had motor experience making patterns they were describing than when they had only viewed the patterns. In these studies, representations that had stronger ties to previous action (Hostetter & Alibali, 2010) or possible action (Chu & Kita, 2016) were described with more gestures, even though the action itself was unrelated to the communicative intent.

The present study tested a stronger version of the claim that gestures are affected by how strongly sensorimotor representations are activated during speaking. In previous demonstrations (e.g., Chu & Kita, 2016; Hostetter & Alibali, 2010), the representation being described was manipulated so as to increase its association with imagined or performed action. However, if the representations that underlie gestures are more general sensorimotor representations that are not specific to communication, then any complementary activation in the sensorimotor system should result in increased gesture, even if the activation is extraneous to the specific representation being described.

To test this claim, we asked participants to describe events with either up or down motion while simultaneously viewing a perceptual display of either up or down motion. Our logic is that viewing a perceptual display of, for example, upward motion activates a representation of upward motion in the sensorimotor system. At the same time, speaking about something moving upwards also involves activating a representation of upward motion. If the representation of upward motion that gives rise to gesture during speaking is the same as the sensorimotor representation that is activated when upward motion is viewed, then viewing a perceptual display that is moving upward should increase activation on that representation, making it more likely that that activation surpasses the speaker's gesture threshold and yields a gesture. Thus, gesture rate should be greater when viewing a congruent perceptual display. In contrast, if gestures arise from representations that are formed only for the purposes of communication, then gesture rate should be unaffected by concurrent sensorimotor activity.

Method

Participants

Participants were 25 native English speakers (13 female) who were recruited from Introductory Psychology courses. Data from an additional seven participants were collected but discarded because of technical errors in the video recording software or because they were not native English speakers. The majority of participants (80%) identified as white or Caucasian, and smaller percentages identified as African American (8%), Hispanic or Latinx (8%) or Asian (4%).

Stimuli

Twenty stimulus animations were created for this study using the program Anime Studio Debut 9. Each animation lasted 10 seconds and depicted events that involved motion (UP, DOWN, or OTHER). Six of the animations depicted upward motion, such as a hot air balloon rising out of the screen, a cow being abducted by an alien space ship, or a tea kettle releasing steam into the air. Six of the animations depicted downward motion, such as a boat dropping an anchor to the bottom of the sea, a helicopter landing, or a curtain falling at the end of a performance. Eight animations depicted non-vertical motion and served as filler trials. The motion conveyed in the filler videos was side-to-side (e.g., a car speeding across the screen), circular (e.g., a Ferris wheel spinning around), or random (e.g., an octopus waving his legs as he dances). One of these filler animations was used as an example and one was used as a practice trial. The remaining six were intermixed with the vertical animations during the experimental trials, in order to prevent participants from detecting the specific focus on up and down motion. Table 1 describes the key motion events in the animations used in the experimental trials.

In addition, 20 distractor videos were created. Each video showed circles of various sizes filling the computer screen and moving across the screen for 15 seconds; half of the videos depicted the circles moving upwards and half depicted the circles moving downwards. In order to insure that participants were attending to the distractor video, they were asked to track the number of red circles that appeared during each trial. In each video, the majority of circles were light blue, with a few circles (either 1, 2, or 3) colored dark red. The position of the red circles on the screen and the timing of their appearance was random but was matched across the down and up distractor videos. That is, for each upward distractor video, there was a downward distractor video that showed the same number of red circles appearing in the same position on the screen at the same time intervals.

Procedure

Participants arrived individually to participate in the study. Each participant was seated in front of a Macintosh computer with a 25 inch screen that was running PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). The experimenter explained that the study was about how well people can communicate while being visually distracted. Participants were told that they would view some short cartoons and then describe them to a video camera while also engaging in a secondary visual perception task.

After signing the consent form, participants were shown an example of how the trials would proceed. First, an animation played on the computer screen that showed people, animals, or objects moving in various ways. Immediately after the animation finished, the screen filled with a distractor video, in which circles of various sizes moved vertically across the screen. As soon as the circles appeared, participants were instructed to begin describing the events that had occurred in the animation they had just seen. They were told that another participant would watch a video of each description and attempt to recreate the motion of the objects in the animation, so it was particularly important that the speakers describe how the people, animals, or objects in the animation had moved. During the example trial, the experimenter gave an example of how to describe the animation (e.g., “there is a Ferris wheel that lights up and spins around several times counterclockwise”) that included one scripted gesture (e.g., right hand index finger draws circle in air two times).

While retelling each story, participants were instructed to look straight ahead at the computer screen and watch the circles on the screen so that they could track the number of red circles that appeared in each display. Each display lasted 15 seconds, which was long enough for participants to finish their description of the previous animation. After they were done with their description, they were instructed to continue watching the display until the circles stopped moving, at which point a prompt appeared on the computer screen that asked participants to report the number of red circles they had seen in the display. The participants gave their response out loud and the experimenter, who was sitting across from the participant, made a note of it. The participant then pressed any key to proceed to the next trial and display the next animation.

After hearing the instructions and seeing the example, the participants completed a practice trial. The experimenter provided any necessary feedback to reemphasize the need to describe the motion of the events in the animation or to

Table 1. Target Motion Event from each Animation

Up Animations	Down Animations	Neutral Animations
Boy’s balloon floats away	Boat drops anchor	Octopus waves its arms
Cow is abducted	Curtain falls at end of performance	Ants march across a picnic
Kettle releases steam	Raindrops fall through ceiling	Police car races through scene
Cat is startled and jumps up	Helicopter lands on roof	A girl eats an ice cream cone
Hot air balloon floats up	Apple falls from tree	A train comes through a tunnel
A rocket takes off	Skydiver parachutes out of plane	A frog eats a fly and hops away

make sure to keep their eyes on the circle display as they were describing the events. Participants then began the experimental trials, which consisted of 6 animations depicting upward events, 6 depicting downward events, and 6 filler trials depicting events with no vertical motion.

The animations were presented in the same preset randomized order, in which no more than two animations depicting motion in the same direction occurred in a row. The 12 experimental trials were intermixed with the 6 filler trials. Half of the trials displayed a distractor video in which the circles were moving upward and half displayed a distractor in which the circles moved downward, so that over the experiment, each participant experienced 6 trials in which the direction of the motion they were describing matched the direction of the circles in the display (i.e., the Congruent condition) and 6 trials in which the direction of the motion they were describing was opposite to the direction of the circles in the display (e.g., the Incongruent condition). The order of the up versus down distractor videos was counterbalanced across participants. Thus, each animation was described in the Congruent condition by half of the participants and in the Incongruent condition by the other half of the participants. The number of red circles that appeared in each display video (1, 2, or 3) varied randomly across the trials but was held constant across the counterbalanced orders, so that only the direction of the circles' movement, and not the number of red circles, differed between the conditions.

After completing the 18 trials, participants completed a brief questionnaire on which they reported at what age they began learning English and whether they had any suspicions about the purpose of the study. None of the participants reported suspecting that gesture was the target of the study, and data from those who reported learning English after age 5 were discarded.

Coding

Participants' narrations of the up and down animations were viewed and coded from the videos. First, all speech was transcribed and all accompanying representational gestures were noted. Representational gestures were defined as movements of the hands or arms that coincided with speech and conveyed semantic meaning. Individual gestures were segmented from one another by a change in handshape, motion, or meaning.

Next, we identified the speech utterances that described the target event in each animation (e.g., the main event depicting motion either up or down; see Table 1) and identified the words produced to describe each key event. Finally, the gestures that accompanied the speech about the action in the key event were coded for whether they conveyed vertical motion. To do this, the relevant section of the video was re-watched and the motion of the hand during the gesture was described. Gestures in which the hand showed vertical displacement from the start of the gesture to the end of the gesture were coded as depicting vertical motion, even if the movement was not a straight trajectory. For example, one speaker described the key event in the parachute animation by saying "he jumps out of the plane" while moving her right hand in an arc trajectory slightly up from chest height, to the left, and then down to her lap. This gesture was coded as conveying vertical motion because the endpoint of the gesture was lower than the starting point. Because our coding was limited to the descriptions of the action in the key events, all gestures coded as depicting vertical motion were specifically about the vertical motion event (e.g., the balloon moving up, not the shape of the balloon). Further, we also made note of whether each vertical gesture depicted motion upward or downward; in all cases, the direction of motion in the gesture matched the motion being described.

Note that it was not apparent from the video of the participants' descriptions which direction the distractor video was moving during each narration; thus, the coders were blind to the experimental condition on each trial. The key events were watched and coded by a second coder. For 92.5% of the trials, the coders agreed exactly on the number of vertical gestures that occurred. In cases where there was disagreement, the counts observed by the more experienced coder were used.

Data Analysis

Our hypothesis is that speakers will gesture more about vertical direction when they are simultaneously viewing a display that shows motion in that same direction than when they are viewing motion in the opposite direction. First, we wanted to make sure that participants were actually attending to the displays. We checked this by considering how successfully participants were able to track the number of red circles on the distractor task. Participants had near perfect performance; errors were made on only four trials across the entire experiment. These four trials were removed from further analysis. We also discarded 15 trials in which the participant failed to describe the key vertical motion event in speech. For example, one speaker said "the boy let go of the balloon," but then did not describe how the balloon floated up and out of the screen.

We calculated each participant's vertical gesture rate per 100 words while describing the target event in each trial. We then analyzed these rates using multilevel modeling in order to account for the large individual variability across participants and items in gesture rates. We did not include separate fixed factors for direction (up vs. down) of the story and distractor because we had too few items per participant to make meaningful comparisons and because we did not have a prediction regarding differential effects for up versus down stories. We included condition (Congruent vs. Incongruent) as a fixed factor, and we also included Trial number as a fixed effect, because it seemed likely that gesture rates would decrease over the experiment (see Yeo & Alibali, 2017). We used the maximal random effects structure possible (see Barr, Levy Scheepers, & Tily, 2014), which included random intercepts for participant and story, as well as random slopes for condition and trial for

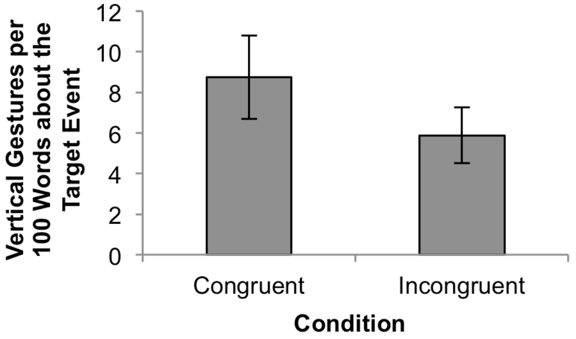

Figure 1. The vertical gesture rate per 100 words produced as participants described the target motion events. Error bars represent standard errors of the means.

both participant and story. The analysis was conducted in R using the lme4 and lmerTest packages. For significance testing, the lmerTest package uses Satterthwaite approximations to degrees of freedom.

Results

The Results are shown in Figure 1. As predicted, condition (congruent vs. incongruent) reliably affected vertical gesture rate during the descriptions of the key events, $\beta = 2.87$, $SE = 1.34$, $t = 2.14$, $p = .04$. Speakers gestured about vertical motion more while describing the target events when they were watching motion in the same direction ($M = 8.75$ gestures per 100 words, $SE = 2.06$) vs. in the opposite direction ($M = 5.88$ gestures per 100 words, $SE = 1.38$). There was no significant effect of trial number, $\beta = -.07$, $SE = 0.18$, $t = 0.40$, $p = .70$.

We next considered whether there were differences in how much participants talked across the conditions. Specifically, did speakers also talk about vertical motion more when they were watching a congruent vs. incongruent display? We considered the number of words participants used to describe the key event in each trial in a mixed linear model with the same fixed effects and random effects structure as the model predicting gesture rate. There was no effect of condition, $\beta = 0.31$, $SE = 0.54$, $t = 0.58$, $p > .05$, or trial number, $\beta = 0.21$, $SE = 0.11$, $t = 1.85$, $p = .08$. Thus, there was no evidence that perceptual input affected how many words participants used to describe the key motion events, but it did affect how much they gestured about them.

Discussion

Speakers gestured more about vertical motion when they were watching a display that moved in the same direction as the motion they were describing than when they were watching a display that moved in the opposite direction. Notably, there was no evidence for differences across conditions in how much speakers talked about the key events, suggesting that watching the congruent display did not focus participants' attention on the vertical motion and make it a more prominent part of their communicative intent. These results suggest that gestures can originate from activity in the sensorimotor system, even when that activity is unrelated to the speaker's communicative intent. These results are consistent with the claims made in the GSA framework (Hostetter & Alibali, 2008) that gestures arise from perceptual and motor simulations that underlie speaking.

Of course, our claim is not that gestures are completely unrelated to communicative intent. Indeed, in communicative situations, the speaker's sensorimotor system is engaged in simulation for the purposes of communication; thus, the communicative intent is typically what motivates the speaker to form a particular sensorimotor simulation in the first place. However, in the present paradigm, we boosted activation of this simulation by exposing participants to simultaneous motion in an unrelated task. We contend that this extra activation made it more likely that the target simulation (i.e., the vertical motion event) would be activated with enough strength to surpass the speaker's current gesture threshold. As a result, speakers expressed more of their vertical motion simulations in gesture when they were viewing congruent motion than when they were viewing incongruent motion.

Under this view, we doubt that simply viewing vertical motion without also being engaged in the communicative task of describing a motion event would elicit gesture (although it is not completely out of the question, as co-thought gestures can and do occur—see Chu & Kita, 2016). Rather, in our view, the effect observed here is the result of a sensorimotor simulation being formed in the interest of communication, and then that simulation receiving additional activation from extraneous engagement of the perceptual system. Indeed, we never observed the direction of motion in the perceptual system overriding the direction involved in the simulation being described. For example, speakers never gestured with a downward motion when describing a motion event that went up. In a communicative task like the one used here, the simulation being described appears to take precedence.

This work was modeled after work investigating the embodied nature of language comprehension (see Kaschak et al., 2005; 2006). However, in that work, there are mixed results regarding whether simultaneous engagement of the perceptual or motor systems results in facilitation of language comprehension or inhibition. In prior work that has shown an inhibitory effect (e.g., Kaschak et al., 2005), participants were already engaged in the perception task when the sentence to be comprehended was introduced, thereby making it difficult for participants to use the same neural system that was already involved in the perception task for the language comprehension task. This resulted in better comprehension when the direction of the display did *not* match the direction of the sentence. In contrast, the communicative goal was introduced first in our paradigm—participants saw the vignette that they would need to describe before the perceptual display appeared. Thus, because the simulation of the vignette was already formed, we argue that viewing the congruent perceptual display lent additional activation to that simulation, rather than inhibiting it.

It is unclear from the present design whether viewing a congruent display increases activation of the simulation being described (as we contend) or whether viewing an incongruent display decreases activation. To further consider this possibility, future work should include a condition with no display or an irrelevant display.

In conclusion, the present findings are compatible with the view that the representations that give rise to gesture are the same representations that are activated during perception of a visual display. The results are not easily explained by the view that gestures are crafted by speakers solely based on communicative intent to explain or depict for a listener (e.g., Clark, 2016), because in the present study, speakers' communicative intent to describe the vignettes remained the same regardless of the direction of motion in the perceptual display. The fact that the extraneous motion in the perceptual display influenced speakers' gestures suggests that gestures are the result of sensorimotor activation in speakers' cognitive systems even if it is extraneous to their communicative task.

Acknowledgments

We thank Lyla Rothschild and Sarah Levitt for their assistance with data transcription.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44,* 169-188.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255-278.

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

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 257–270.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review, 123,* 324-347.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25,* 257-271.

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284-311). Cambridge: Cambridge University Press.

de Vega, M., Moreno, V., & Castillo, D. (2013). The comprehension of action-related sentences may cause interference rather than facilitation on matching actions. *Psychological Research, 77*, 20-30.

Diefenbach, C., Rieger, M., Massen, C., & Prinz, W. (2013). Action-sentence compatibility: The role of action effects and timing. *Frontiers in Psychology, 4*, 1-13.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex, 48,* 905-922.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558-565.

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

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language, 63,* 245–257.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes, 22,* 313-336.

Kaschak, M. P., & Borreggine, K. L. (2008). Temporal dynamics of the action-sentence compatibility effect. *The Quarterly Journal of Experimental Psychology, 61*, 883-895.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition, 94*, B79-B89.

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science, 30*, 733-744.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of thinking and speaking. *Journal of Memory and Language, 48*, 16-32.

Perlman, M., Clark, N., & Johansson Falck, M. (2015). Iconic prosody in story reading. *Cognitive Science, 39,* 1348-1368.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36,* 329-392.

Speed, L. J., & Vigliocco, G. (2013). Eye movements reveal the dynamic simulation of speed in language. *Cognitive Science, 38*, 367-382.

Wagner, S., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is hand-waving spatial? *Journal of Memory and Language, 50*, 395-407.

Yeo, A., & Alibali, M. W. (2017). Does visual salience of action affect gesture production? *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*, 168-171.