Gesture Production during Spatial Tasks: Its Not All About Difficulty

Autumn B. Hostetter (Autumn.Hostetter@kzoo.edu)
Kalamazoo College, Department of Psychology, 1200 Academy Street
Kalamazoo, MI 49006 USA

Erin L. Sullivan (Erin.Sullivan08@kzoo.edu)
Kalamazoo College, Department of Psychology, 1200 Academy Street
Kalamazoo, MI 49006 USA

Abstract

Previous research has shown that speakers gesture more when describing imagistic information than when describing non-imagistic information. One explanation for this finding is that spatial information is more difficult to describe verbally than non-spatial information. In the present study, we designed two novel tasks, one verbal and one spatial, in which difficulty and spatiality were not confounded. In Experiment 1, we demonstrated that the spatial task is actually less difficult and leads to fewer errors than the verbal task. In Experiment 2, we found that speakers produced more representational gestures on the spatial task than the verbal task, even though it was not more difficult. Results suggest that speakers do not gesture on spatial tasks simply because they are more difficult.

Keywords: gesture; mental imagery; task difficulty

Introduction

Representational gestures are movements of the hands and arms that represent motor or spatial concepts (e.g., moving a hand in an arc to show the trajectory of a falling object or drawing the shape of a triangle in the air to indicate a triangle). There is accumulating evidence that speakers are particularly likely to produce representational gestures when they are describing spatial information. For example, Beattie and Shovelton (2002) found that a property they termed "imageability" affected the probability that a clause of speech would be accompanied by gesture. Essentially, clauses that were most frequently accompanied by gesture were also most evocative of a mental image. Similarly, Feyereisen and Havard (1999) found that participants gestured significantly more when asked questions that would presumably activate mental images (e.g., Could you describe the room in which you live most often?) than questions that focused on abstract concepts (e.g., What do you think about the death penalty in the United States?).

There are a variety of explanations for the co-occurrence of gestures with speech about spatial ideas. First, speakers may gesture with spatial concepts because gestures about spatial concepts are particularly likely to benefit a listener's comprehension. However, Pine, Gurney, and Fletcher (2010) found that speakers gestured primarily with spatial concepts even when a listener could not see them, suggesting that the co-occurrence of gestures with spatial information is likely more than communicative.

Second, gestures may occur with spatial information because spatial information affords gesture in a way that non-spatial information does not. For example, if a speaker is describing how big his fish was, he can easily display the size of the fish in gesture. If the speaker is describing how the fish tasted, on the other hand, it is much more difficult for him to represent the taste of the fish in his gesture. Indeed, Krauss, Dushay, Chen, and Rauscher (1995) found that very few of the speakers in their study gestured when describing the taste of a particular tea. While this explanation is often ignored by gesture researchers, it is likely part of the reason for the frequent occurrence of gesture with speech about spatial information.

However, in addition to the fact that spatial information likely affords more gestures than non-spatial information, spatial representations in the mind of the speaker may also lead more naturally to gesture than non-spatial representations. When speakers think about spatial information, areas of visual and motor cortex are activated in much the same way as when the speakers actually interact with the physical world. When speakers imagine themselves performing a particular action, for example, the same areas of their motor cortex are activated as when they actually perform the action (Willems, Hagoort, & Cassasanto, 2010). Further, imagining how something rotates relies on the same motor areas involved in physically rotating an object (Wexler, Kosslyn, & Berthoz, 1998). Such evidence suggests that spatial cognition is embodied, or rooted in the way our bodies interact with the physical world (see Wilson, 2002).

A third possibility is thus that the gestures people produce when they speak about spatial information are a manifestation of the embodied cognitive processes that are involved in thinking about spatial information. Several current theories about the cognitive origin of gestures propose that gestures arise from imagistic representations in the mind of the speaker (e.g., Hostetter & Alibali, 2008; Kita & Ozyurek, 2003). One such view, termed the Gesture as Simulated Action (GSA) framework, suggests that people produce representational gestures when their motor cortex is activated as the result of visual and motor simulations during thinking and speaking. When this motor activation is strong enough to cross a certain threshold (Hostetter & Alibali, 2008), the speaker produces a gesture. This threshold varies between individuals and situations, so what may be strong enough activation to produce a gesture in one individual or one situation may not be in another. Still, the hypothesis under the GSA framework is that the higher the degree of motor activation, the higher the likelihood that a gesture will be produced.

Thus, the GSA framework and other views that propose an imagistic source of gesture lead to the claim that the occurrence of gesture during speaking should be more dependent on the way in which the information is represented than on the speaking topic. Under this view, speakers gesture with spatial information not just because they can or just because they are trying to facilitate communication; rather, they gesture with spatial information because they routinely think about spatial information in a way that naturally evokes gesture. Indeed, Hegarty, Mayer, Kriz, and Keehner (2005) showed that 90% of descriptions of mental animation problems are accompanied by gesture.

Finally, a fourth possibility is that gestures occur with spatial information because spatial information is difficult to describe. Much research has suggested that speakers gesture more when they are having speech production difficulty, either at the formulation (Krauss, 1998) or conceptualization stage (Kita, 2000). For example, Chawla and Krauss (1994) showed that speakers gestured at higher rates when they had not had an opportunity to rehearse their speech than when they had. When speakers describe spatial information, it may be more difficult to think of or find the words needed to capture the meaning. As a result, speakers might gesture as a means of signaling their difficulty to the listener or priming the appropriate word in their lexicon. For example, some research has shown that speakers are more fluent when they gesture than when they do not gesture (Rauscher, Krauss, & Chen, 1996).

In the present study, we pit these latter two possibilities against one another. Specifically, we designed two versions of a task, one spatial and one verbal, that we expect to differ in terms of difficulty. In both versions of the task, participants describe how to transform smaller parts into a meaningful whole. In the verbal version of the task, participants describe how to create a given sentence from words and phrases. In the spatial version, participants describe how to create a given object from smaller shapes. Both conditions require participants to use words that are familiar to the participant but not frequently used on a daily basis (e.g., *trapezoid* and *parallelogram* in the spatial condition and *direct object* and *subject* in the verbal condition). Further, in both tasks, participants are describing transformations, either spatial transformations (e.g., rotate the trapezoid 180 degrees) or propositional transformations (e.g., change the verb ‘chase’ to the past tense). Finally, in both tasks, participants are describing how to manipulate things, either objects or words on the page, and where to relocate them in space. Thus, in both conditions, speakers could theoretically represent the location of the word or the object in gesture. However, in addition to relying more on spatial resources, it is predicted that the spatial task will also be easier for participants than the verbal task because the objects and words involved are more familiar to the speakers.

In Experiment 1, we directly tested this claim in a dual task paradigm. We compared performance on the two description tasks when participants engaged simultaneously in a secondary memory task. In one condition, the memory task required participants to remember verbal information (i.e., letter pairs). In the other condition, participants remembered spatial information (i.e., the locations of dots in a grid). We predicted that speakers would make more errors on the verbal task than on the spatial task, particularly when they are simultaneously remembering verbal information.

In Experiment 2, we compared the gesture rates produced by speakers on the two versions of the task. Do gestures occur more frequently on the spatial task even when that task is not more difficult? Or do gestures occur more frequently on the difficult task even when that task is not spatial?

Experiment 1 Method

Participants

Participants in Experiment 1 were 13 undergraduate students at Kalamazoo College (4 male). Their average age was 18.4 years. They were recruited from introductory psychology classes and received extra credit for their participation. All participants were native English speakers, and all were Caucasian except for one who was African American.

Materials

For the verbal stimuli, four words and short phrases were printed at the top of a piece of paper. A target sentence that could be made from these words was printed across the bottom of the page. See Figure 1. For the spatial stimuli, three shapes were located at the top of a piece of paper in a row. The name of the target object and a picture of the object made from the shapes on the top was printed at the bottom of the page. See Figure 2. Nine stimuli were made of each type, six to be used in the trials of the experiment, one to be used as an example, and two to be used as practice.

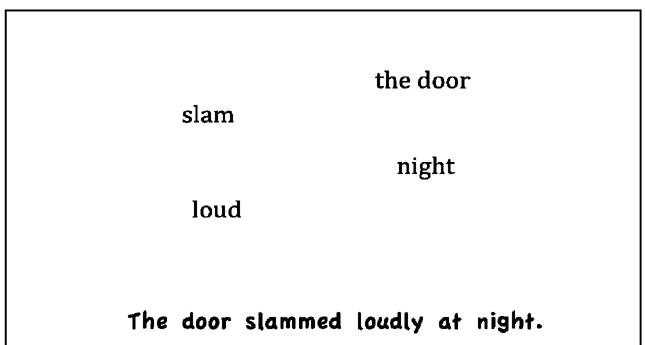

Figure 1: Example Verbal Stimuli

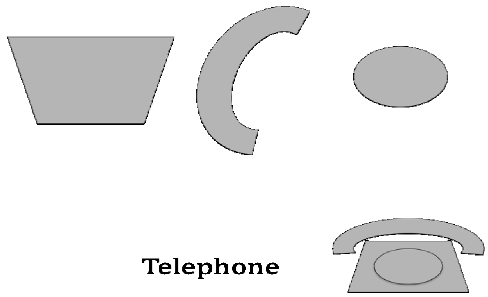

Figure 2. Example Spatial Stimuli

Each stimulus was printed on a separate sheet of 8.5 x 11 inch paper and laminated.

For the secondary task, participants remembered a series of letters (in the verbal secondary task) and a series of dots in a 5 x 5 grid (in the spatial secondary task) that were presented on a MacBook laptop with a 14 in. screen running PsyScope software (Cohen. MacWhinney, Flatt, & Provost, 1993). The letters for each verbal trial were presented in three pairs (e.g., DX, VY, RP). All letter pairs were chosen by the experimenters, with care to avoid letter combinations that evoked common abbreviations or words. The dots for each spatial trial were also presented in three pairs. Each pair of letters and dots remained on the computer screen for 2 seconds, and the three pairs were shown in rapid succession. Following each description, participants recorded the dots or letters they remembered on an answer sheet.

Procedure

When participants arrived at the testing site, the experimenter led them to the testing room. Participants were told that the purpose of the experiment was to investigate how people complete multiple tasks at the same time. Participants signed a consent form and were given the opportunity to ask questions.

There were four conditions in the experiment, as each primary description task (spatial vs. verbal) was paired with each secondary memory task (dots vs. letters). Participants completed one practice and three experimental trials in all four conditions in a within subjects, blocked design. Participants completed all six of the description trials of one type, first paired with one secondary memory task (e.g., dots) and then paired with the other secondary task (e.g., letters), before completing the six description trials of the other type. Thus, there were four possible orders to which participants were randomly assigned.

Before beginning the first trial, participants received detailed instructions from the experimenter. The experimenter first showed the participant how the dots or letters (depending on the order to which the participant had been assigned) would appear on the computer screen. Participants were then given instructions for the description task. The experimenter instructed them to describe how to manipulate and reposition the items at the top of the page to create the object or sentence at the bottom. They were told that their descriptions were being audio-taped and that someone would later listen to their descriptions and attempt to recreate the objects and sentences that were being described.

In the verbal condition, the experimenter showed the participants an example verbal stimulus (see Figure 1) and told the participants that they were not allowed to state the sentence, only what grammatical role each word played in the final sentence. For example, an appropriate description of the sentence shown in Figure 1 might be "Slam is the verb in past tense. The door is the subject. Loud becomes an adverb, and night is part of a prepositional phrase beginning with at." In the spatial condition, the experimenter showed the participants an example of a spatial stimulus (see Figure 2) and explained that participants were not allowed to state what the object was, only how to manipulate the given shapes to create that object. For example, the spatial stimulus depicted in Figure 2 might be described by saying "rotate the trapezoid 180 degrees so that the longer side is on the bottom. Take the arc and rotate it to place it on top of the shorter side of the trapezoid. Take the circle and place it in the center of the trapezoid." In both conditions, participants were to describe the role or position of each word or shape in the order that they were given at the top of the page.

The experimenter then showed the participant how to complete the memory task by reproducing as many of the dots or letters as possible on the answer sheet. Participants recalled six letters or dots for each trial; if they could not remember all six, they were encouraged to guess. Order of recall did not matter.

After the example and instructions, participants were given a quick geometry or grammar lesson, which reviewed the geometrical or grammatical terms that would be needed in the trials. Participants then completed a practice trial and received feedback from the experimenter as appropriate.

In each trial, the participant pressed the space bar on the computer to view the first set of memory stimuli. Following the third pair of dots or letters, the experimenter gave the participant the stimulus to be described. Once the participant completed the description, the experimenter took away the stimulus and gave the participant the answer sheet. The participant then reproduced as many of the dots or letters as he or she could on the answer sheet. Letters were recorded on lines, and participants filled in a blank grid with dots. Order of recall was not important.

After completing the first three trials, the experimenter explained that the type of information being remembered was going to change. The experimenter then demonstrated the other type of memory stimuli on the computer, and the participant completed a second practice trial while remembering the new type of information. The participant then completed the three trials as they had in the first block.

After completing the six trials with one description task, the experimenter introduced the second description task. As with the first set of trials, the participant received an example, a geometry or grammar lesson, and a practice trial before beginning the first experimental trial. After completing the first three trials with memory stimuli of the first type, they each did a practice trial and three experimental trials with the other memory task.

A hidden camera recorded the participants' descriptions. All participants were made aware of the presence of the camera at the end of the study, and they all consented for their data to be used in the study.

Coding

The descriptions were transcribed by the experimenter, and the errors produced by the participants were coded. Mislabeling an item (calling a trapezoid a parallelogram or a direct object an indirect object) and inaccurate descriptions of transformations (saying that a shape should be rotated 90 degrees instead of 180 or that a verb should be made past tense instead of present) were coded as errors.

Experiment 1 Results

Analysis of Primary Task Performance

The number of errors produced by speakers per 100 words of speech on each description task were analyzed with a 2 (type of task: verbal or spatial) x 2 (type of information remembered: letters or dots) repeated measures analysis of variance (ANOVA). As predicted, there was a main effect of description task, $F(1, 12) = 9.57$, $p < .01$. Participants made more errors in the verbal description task than in the spatial description task. There was also a significant interaction between the type of task (verbal vs. spatial) and type of information remembered (letters vs. dots) on the number of errors made in the description task, $F(1, 12) = 4.68$, $p = .05$. When participants described how to make a sentence from component words, they made more errors when simultaneously remembering verbal information than when remembering spatial information ($p = .06$). Participants showed the reverse pattern when describing how to make an object from component shapes, although the difference is not significant ($p = .41$). There was no effect of memory task, $F(1, 12) = 3.50$, $p = .09$. See Table 1.

Analysis of Secondary Task Performance

The average number of memory stimuli remembered in each condition was compared in a 2 (description task: spatial vs. verbal) x 2 (memory task: letters vs. dots) repeated measures ANOVA. There was a main effect of memory task, such that participants always remembered more letters

Table 1. Average errors (and standard deviations) made per 100 words in the Description tasks in Experiment 1

	Dots	Letters
Spatial	0.1 (.26)	0.04 (0.13)
Verbal	1.57 (2.57)	2.77 (2.72)

($M = 4.62$) than dot locations ($M = 2.96$), $F(1, 12) = 28.93$, $p < .001$. There was no effect of description task, $F(1, 12) = 1.89$, $p = .19$ and no interaction, $F(1, 12) = 0.67$, $p = .43$. Participants did not remember more dot locations when performing the verbal task ($M = 3.15$, $SD = 1.14$) than when performing the spatial task ($M = 2.77$, $SD = 0.89$) and did not remember more letters when performing the spatial task ($M = 4.62$. $SD = .89$) than when performing the verbal task ($M = 4.62$, $SD = 1.03$).

Experiment 1 Discussion

As predicted, the verbal task appears to be more difficult for participants than the spatial task. Participants made very few errors on the spatial task. Further, the verbal task appears to rely more strongly on verbal cognitive resources than does the spatial task. Participants made more errors on the verbal task when remembering verbal information than when remembering spatial information, but they did not show this pattern on the spatial task.

In designing the spatial task, we had hoped that the spatial task would primarily rely on spatial resources, just as the verbal task primarily relies on verbal resources. Unfortunately, the results do not strongly support this conclusion, as the secondary spatial task did not interfere with performance on the spatial task more than the secondary verbal task did. However, the error rates were so low on the spatial task overall that a floor effect may have prevented the predicted pattern from emerging.

Despite the lack of clear evidence that the spatial task relies on spatial resources, the results do suggest that our novel tasks are different in two fundamental ways. First, the verbal task is more difficult than the spatial task. This suggests that speakers might gesture more frequently on this task than on the spatial task as they struggle to formulate their thoughts and choose the best words needed to convey their meaning (e.g., Chawla & Krauss, 1994). Second, the verbal task relies more on verbal resources than the spatial task does. If gestures occur primarily with spatial ideas, then speakers should gesture at low rates on the verbal task because it appears to primarily involve verbal rather than spatial resources. Thus, there are two alternative hypotheses about which task should be accompanied by the most gesture. If gestures occur when ideas are difficult to describe, then gestures should be more prevalent on the verbal task. In contrast, if gestures occur when ideas are spatial in nature, then gestures should be less prevalent on the verbal task. Experiment 2 will test these two alternatives.

Experiment 2 Method

Participants

The 11 participants (1 male) in Experiment 2 were recruited from the same population as those in Experiment 1. Their average age was 18.5 years. All participants were native English speakers. Participants were Caucasian, except for one who was Hispanic.

Materials and Procedure

The procedure and stimuli were identical to those used for the description task in Experiment 1, except that participants did not perform the memory task simultaneously. Participants were told that the purpose of the experiment was to investigate how people remember and communicate information. Like in Experiment 1, participants were told that someone would later listen to their descriptions and have to recreate the object or sentence.

A video camera recorded the participants' descriptions. All participants learned of the presence of the hidden camera and the interest in gesture at the end of the study. All participants declined the opportunity to have their data removed from the study.

Coding

The experimenter watched each video and transcribed the participants' speech. Participants' representational gestures were coded. Movements that represented concepts discussed in speech or matching the stimuli being described were coded as representational gestures. For example, a movement made along with the words "take the trapezoid and turn it upside down" in which the speaker turned her palm over was coded as a representational gesture. Deictic gestures were also coded. Deictic gestures are movements that indicate something in physical space by pointing to it. For example, a speaker who said "the dog is going to be the subject" while pointing to the words "the dog" on the page is producing a deictic gesture. Beat gestures, or simple up and down movements that add emphasis without representing content, were also coded.

Experiment 2 Results

Gestures rates per 100 words were calculated and averaged across the six trials of each type for each speaker. We then compared gesture rates across the conditions with paired-samples t-tests. Speakers produced significantly more representational gestures per 100 words when describing how to make an object from component shapes (M = 3.97, SD = 4.21) than when describing how to make a sentence from component words and phrases (M = 0.26, SD = 0.50), $t(10) = 3.11$, $p = .01$. There was no difference in the rate of deictic points produced across conditions, $t(10) = 0.07$, $p = .94$, or in the rate of beats produced across conditions, $t(10) = 0.24$, $p = .82$. See Figure 3.

General Discussion

The results of Experiment 2 suggest that participants produce more representational gestures when describing the spatial task than when describing the verbal task. This effect emerged even though the error rates produced in Experiment 1 suggest that the spatial task is not more difficult than the verbal task. Thus, it does not appear that speakers simply gesture more whenever they are having more difficulty describing something.

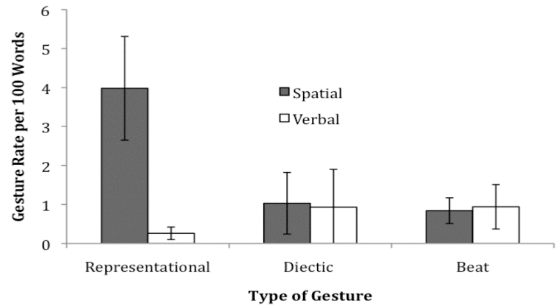

Figure 3. Average gesture rates in the spatial and verbal trials of Experiment 2. Error bars represent standard errors of the means.

If the increased use of gesture in the spatial task is not due to the difficulty of the task, how can it be explained? One explanation is that the spatial thinking that underlies speakers' descriptions of the spatial task is naturally evocative of gesture. According to the GSA framework (Hostetter & Alibali, 2008), there is overlap in the mental resources that are involved in thinking about spatial information and those that are involved in producing gestures. When speakers are thinking about spatial information, their mental representations are in a format that naturally results in gesture.

However, these results do not rule out the possibility that difficulty does play a role in representational gesture production, particularly in spatial tasks. When speakers think about spatial information, they may be more likely to gesture when that spatial information is difficult to describe than when it is easy to describe (e.g., Hostetter et al., 2007; Kita & Davies, 2009). However, the present findings suggest that whether the task is spatial or not is a more important determinant of representational gesture rates than whether the task is difficult or not.

It is possible that the spatial version of the task simply affords more representational gesture than the verbal version of the task. That is, regardless of what type of mental representation underlies the description, there may simply be more things to gesture about in the spatial task than in the verbal task. However, it is possible to gesture on the verbal task; in fact, some of the participants did. The representational gestures that they produced involved showing in gesture how the words would "move" to their correct location in the sentence. For example, to make "the door" the subject of the sentence, it needs to move to the left of the page, and this position can be shown in gesture just as readily as the position of the arc on the top of the telephone can be shown in gesture. The fact that participants rarely chose to gesture in this way in the verbal description seems to have more to do with the fact that such a gesture is not naturally evoked by the description than by the fact that it is not possible to do so.

Indeed, in future work, we will actively encourage speakers to gesture during their descriptions of both tasks. If gestures arise naturally from the spatial representations that are evoked during the spatial task, then gesturing should come with very little cognitive cost on this task. In contrast, gesturing during the verbal description task should actually be somewhat difficult for participants, because they will have to first think about the task in a way that affords gesture. This increased difficulty should be manifested in poorer performance on a simultaneous memory task.

Previous work (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner, Nusbaum, & Goldin-Meadow, 2004) has shown that encouraging speakers to gesture typically reduces cognitive resources and improves performance on a secondary memory task. The tasks used in these previous studies, however, were arguably spatial in nature, and as a result presumably evoke gesture quite naturally and effortlessly (see Hostetter & Alibali, 2008). The proposal made by the GSA framework and other views that propose that gestures arise from mental images activated during speaking is that gestures should only be effortless when they occur during descriptions that are based on mental images.

In conclusion, the data described here provide an important first step toward identifying the type of cognitive representations and cognitive effort that are involved in gesture production. This is the first study in which speaking difficulty and task spatiality have been intentionally separated. It seems that speakers do not gesture about spatial information simply because that information is more difficult to verbally describe.

Acknowledgments

We thank Chelsea Baumgarten, Cierra Gillard, Keith Moreno, and Andrea Potthoff for input on the manuscript.

References

Beattie, G., & Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology, 41,* 403-417.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*, 580-601.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

Goldin-Meadow, S., Nusbaum, H., Kelly, S., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516-522.

Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior, 23,* 153-171.

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation, 5*, 333-356.

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

Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain & Language, 82*, 22-29.

Hostetter, A. B., & Skirving, C. J. (in press). The effect of visual vs. verbal stimuli on gesture production. *Journal of Nonverbal Behavior*.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162-185). Cambridge, UK: Cambridge University Press.

Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes, 24*, 761-775.

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.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54-60.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of communicative hand gestures. *Journal of Experimental Social Psychology, 31*, 533-552.

Pine, K. J., Gurney, D. J., & Fletcher, B. (2010). The semantic specificity hypothesis: When gestures do not depend upon the presence of a listener. *Journal of Nonverbal Behavior, 34*, 169-178.

Rauscher, F., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226-231

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.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68*, 77-94.

Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-specific representations of action verbs: Neural evidence from right- and left-handers. *Psychological Science, 21*, 67-74.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625-636.