Journal of Memory and Language 63 (2010) 245–257

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Language, gesture, action! A test of the Gesture as Simulated Action framework

Autumn B. Hostetter [a,*], Martha W. Alibali [b]

[a] Department of Psychology, Kalamazoo College, 1200 Academy Street, Kalamazoo, MI 49006, United States
[b] Department of Psychology, University of Wisconsin-Madison, 1202 W. Johnson Street, Madison, WI, 53706, United States

ARTICLE INFO

Article history:
Received 1 February 2010
revision received 1 April 2010
Available online 10 May 2010

Keywords:
Gesture
Language
Embodied cognition

ABSTRACT

The Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008) holds that representational gestures are produced when actions are simulated as part of thinking and speaking. Accordingly, speakers should gesture more when describing images with which they have specific physical experience than when describing images that are less closely tied to action. Experiment 1 supported this hypothesis by showing that speakers produced more representational gestures when describing patterns they had physically made than when describing patterns they had only viewed. Experiment 2 replicated this finding and ruled out the possibility that the effect is due to decreased opportunity for verbal rehearsal when speakers physically made the patterns. Experiment 3 ruled out the possibility that the effect in Experiments 1 and 2 was due to motor priming from making the patterns. Taken together, these experiments support the central claim of the GSA framework by suggesting that speakers gesture when they express thoughts that involve simulations of actions.

Introduction

Speech production is not a stationary endeavor; on the contrary, we utilize our motor systems in many ways while speaking. Most evidently, we move our tongue, teeth, lips, and vocal cords in order to articulate comprehensible speech. Moreover, we also often move our entire bodies, and especially our hands and arms, to depict the objects, events, and experiences we are describing. These manual movements, referred to here as *representational gestures*, are pervasive in human communication; they occur in cultures around the world (Kendon, 2004) and are even produced by congenitally blind individuals (Iverson & Goldin-Meadow, 1998). Representational gestures play important roles both in facilitating speech production (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Hostetter, Alibali, & Kita, 2007; Krauss, 1998; Melinger & Kita, 2007; Wesp, Hesse, Keutmann, & Wheaton, 2001) and in enhancing comprehension (e.g., Kelly, Barr, Church, & Lynch, 1999; Kendon, 1994). However, despite our growing knowledge about the *functions* served by gestures, little is understood about the *source* of gestures. What kinds of mental processes give rise to gestures?

Gesture as Simulated Action

The Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008) proposes an answer to this question. According to the GSA framework, gestures arise when speakers simulate actions in the interest of speaking. A *simulation* is a recreation of the neural states involved in performing or witnessing a particular action, in the absence of actually doing so. During a simulation, motor and premotor areas of the brain are activated in action-appropriate ways. Imagining oneself performing a particular action activates motor areas of the brain, such as motor cortex, the cerebellum, and basal ganglia (Jeannerod, 2001). According to the GSA framework, this motor activation is sometimes realized in overt movement during speech production, and when this occurs, a representational

* Corresponding author.
E-mail addresses: Autumn.Hostetter@kzoo.edu (A.B. Hostetter), mwalibali@wisc.edu (M.W. Alibali).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2010.04.003

gesture is "born". One implication of this framework is that speakers should gesture more when describing representations that involve highly activated action simulations than when describing representations that involve less highly activated action simulations. The purpose of the present experiments is to empirically test this claim.

The idea that action simulations are involved in language comprehension and production is gaining empirical support (e.g., Glenberg & Kaschak, 2002; Pulvermüller, 2005; Spivey, Richardson, & Gonzales-Marquez, 2005; Willems & Hagoort, 2007). Neuroimaging studies have documented that areas of premotor cortex are involved in comprehending language about action. For example, Hauk, Johnsrude, and Pulvermüller (2004) used fMRI to measure activation in premotor cortex while participants read words describing actions (*lick*, *pick*, and *kick*) that would be performed with specific effectors (mouth, hand, and foot). They found that the areas of premotor cortex that are activated when participants read these words are the same as the areas activated when participants physically produce the corresponding actions. For instance, when participants produce a *pick* action, hand and finger areas of premotor cortex are activated; similarly, when participants read the word *pick*, these same manual areas of premotor cortex are activated. A similar somatotopic mapping in premotor cortex is present when participants read entire sentences that relate to specific motor effectors (i.e., *I bite the apple*; *I grasp the knife*; *I kick the ball*) (Tettamanti et al., 2005).

In addition to this neuroimaging data, behavioral data also suggest that the motor system is involved in language processing. Participants are generally faster to make semantic and lexical judgments when their motor systems are engaged in semantically relevant ways. For example, Pulvermüller, Hauk, Nikulin, and Ilmoniemi (2005) activated specific areas of premotor cortex with transcranial magnetic stimulation and observed effects on a subsequent lexical decision task. They found that briefly stimulating the hand area of premotor cortex expedited participants' decision that PICK was a word. In contrast, stimulating the foot area of premotor cortex facilitated participants' ability to recognize KICK as a word. Similarly, at the sentence level, Glenberg and Kaschak (2002) report an Action Compatibility Effect, in which readers are faster to judge a sentence as comprehensible when they respond with a movement that is in the same direction as the movement described in the sentence. For instance, participants are faster to respond to sentences that imply motion away from the body, such as *Close the drawer*, by pushing a lever away from their body than by pulling a lever toward their body.

The evidence reviewed thus far indicates that motor areas of the brain are activated in action-specific ways during language comprehension.[1] Far fewer studies have addressed the involvement of the motor system during language *production*. In one of the few studies addressing production, Morsella and Krauss (2005) used EMG to measure electro-muscular activity in participants' hands and arms while they retrieved difficult lexical items. They found more activity when participants retrieved concrete words than when they retrieved abstract words. Morsella and Krauss interpreted this as evidence that the motor effectors involved in interacting with concrete objects are activated when their names are retrieved from the lexicon.

Although Morsella and Krauss's (2005) findings are suggestive, one does not need an EMG to detect activity in speakers' hands and arms. Rather, manual movement is often directly observable in the form of representational gestures, or movements that illustrate some aspect of a speaker's meaning, perhaps by miming a particular action, by depicting a particular spatial property, or by pointing to an object in the environment. For example, a speaker might say, "I picked the pepperoni off the pizza" while making a picking motion with her thumb and index finger. Representational gestures of this sort can be distinguished from *beat gestures*, which are simple up-and-down movements that coincide with the rhythm or prosody of speech without carrying direct semantic information (McNeill, 1992). Although both types of gestures engage the motor system by requiring the planning and execution of a physical movement, representational gestures are of particular interest because they use action to convey semantic information. Such movements therefore suggest a parallel with the way we use simulations of action to comprehend semantic information.

The Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008) proposes that activation of the motor system during language production (in the form of representational gestures) and language comprehension (in the form of action simulations) is more than coincidental. Specifically, the GSA framework claims that the motor simulations that are involved in thinking about actions are the source of representational gestures. According to the GSA framework, thinking about a to-be-mentioned action involves a neural simulation of the action event; this neural simulation activates motor and premotor areas of the brain that are capable of producing corresponding overt movements. These overt movements are what we recognize as representational gestures. Consider the "I picked the pepperoni off the pizza" example once again. The GSA framework contends that the speaker's representation of the event includes a simulation of the picking motion that she produced when she actually picked the pepperoni off the pizza. When she thinks about the picking event to describe it, her motor system simulates the picking action, just as her motor system would simulate a picking action if she had read the word *pick*. The GSA framework contends that this neural simulation of the picking action has the potential to become realized as overt movement in the form of a representational gesture.

Empirical predictions derived from the GSA framework

The GSA framework holds that the likelihood of a speaker gesturing during a particular utterance increases with the likelihood that the utterance relies on representations that

[1] Note that we are not interpreting this evidence to imply that motor simulation is *necessary* for semantic comprehension. Whether motor simulation is a necessary component of language comprehension or simply an epiphenomenon of successful image formation remains to be determined (but see Havas, Glenberg, Gutowski, Lucarelli and Davidson [in press] for some provocative experimental evidence involving Botox and comprehension of emotional language). Regardless of the necessity of action simulations, evidence suggests that motor areas of the brain are activated in action-specific ways during language processing.

involve simulations of actions. Further, for utterances that involve action simulations, speakers are more likely to produce gestures when those simulations are more highly activated. Thus, descriptions of events that are based on action or perceptual experience should be more likely to elicit gestures than descriptions that are based on verbal or propositional representations. This idea has received some support from Hostetter and Hopkins (2002), who found that speakers who learned about events from a written description gestured less than speakers who learned about the same events through an animated, spatial cartoon.

However, the strength of activation of action simulations may vary even among descriptions of perceptual experiences. Images of a visual scene, or visual images, may involve simulations of action, because speakers imagine how they would interact with the components of the scene. Indeed, a long tradition in cognitive psychology views perception and action as linked in a cycle of coordination (Dewey, 1896; Gibson, 1979; Sperry, 1952). All perception is, at some level, tied to action, because in order to perceive, we must move our eyes, heads, necks, and bodies (Campos et al., 2000; O'Regan & Noë, 2001). Moreover, we perceive in order to act; when we perceive patterns and objects in the world, we automatically activate appropriate actions (Ellis & Tucker, 2000; Tucker & Ellis, 1998). According to embodied theories of cognition (Barsalou, 1999; Glenberg, 1997), the relation between action and perception that exists in our physical dealings with the world also exists in our mental representations of the world. Thus, thinking about a particular visuo-spatial image can involve simulations of the actions associated with forming or manipulating the image or with interacting with the object of the image. According to the GSA framework, describing the image can yield corresponding representational gestures that reflect those simulated actions.

However, while visuo-spatial images may lead to action simulations, these simulations may be less highly activated than those that occur when speakers are thinking more directly about actions, as is the case with motor imagery. Motor images, or images of the body in motion, necessarily involve highly activated simulations of action, as motor images require a representation of the body's interactions with the environment (Jeannerod, 2001). Thus, action simulations are highly activated when thinking about a particular action, about how to execute it, or about its effects on the environment (Wexler, Kosslyn, & Berthoz, 1998; Wohlschläger & Wohlschläger, 1998). The GSA framework (Hostetter & Alibali, 2008) holds that these highly activated action simulations are particularly likely to result in representational gestures. Accordingly, speakers should gesture more when describing motor images than when describing visual images, because motor images involve action simulations that are more highly activated. There is some evidence to support this claim, as speakers gesture more when describing motor information than when describing visuo-spatial information (Feyereisen & Havard, 1999; Hostetter & Alibali, 2007). For example, Hostetter and Alibali (2007) found that speakers gestured almost three times more when describing how they would wrap a package than when describing events they had viewed in an animated cartoon.

Although suggestive, past studies have not been specifically designed to test the hypothesis that gestures arise from representations that include highly activated simulations of action. As a result, the differences in gesture rates between propositional and imagistic representations (Hostetter & Hopkins, 2002) or between motor and visuo-spatial tasks (Feyereisen & Havard, 1999; Hostetter & Alibali, 2007) could be traced to a number of sources, including differences in the communicative or cognitive demands of the tasks. For example, perhaps participants who are describing how to do something (like wrap a package) gesture frequently because they are attempting to describe the process clearly for their listeners; in contrast, describing what happened in a cartoon may not carry the same communicative demands. In support of this possibility, speakers do alter their gestures depending on aspects of the communicative situation (e.g., Alibali, Heath, & Myers, 2001; Gerwing & Bavelas, 2004; Holler & Stevens, 2007; Özyürek, 2002). Alternatively, perhaps describing how to wrap a package is simply harder than describing the events that occurred in a cartoon. Previous research suggests that speakers gesture more when describing complex information than when describing simple information (Hostetter et al., 2007), and that producing gestures can alleviate cognitive demands on working memory (Goldin-Meadow et al., 2001; Wagner, Nusbaum, & Goldin-Meadow, 2004). Thus, it is possible that speakers gesture more when describing motor information than when describing visuo-spatial information because the motor information is more difficult to describe, rather than because the motor information involves more highly activated simulations of action.

Rationale for the present study

Thus, a direct test is needed for the GSA framework's claim that gestures arise from representations that involve highly activated simulations of action. An ideal test would involve a clean manipulation of how highly action simulations are activated that does not also affect communicative or task demands. This is the goal of Experiment 1. In this experiment, participants describe patterns that they have physically recreated as well as patterns that they have only viewed. If gestures arise when actions are simulated in the interest of thinking and speaking, then speakers should gesture more when describing a representation that has been partially formed through motor experience than when describing a representation that has been formed primarily through visual experience. This is because the motor experience of making the pattern should result in more highly activated action simulations than the visual experience of viewing the pattern.

To preview the findings, Experiment 1 yielded the predicted pattern. However, some alternative explanations for the results of Experiment 1 remained tenable. Two additional experiments were conducted in order to strengthen the argument presented based on Experiment 1. Experiment 2 sought to equate opportunities for verbal rehearsal across conditions, and Experiment 3 sought to test whether a general prime to move in the action conditions might be responsible for the results.

Experiment 1

The purpose of Experiment 1 was to test the claim that gestures arise from highly activated simulations of action (Hostetter & Alibali, 2008). To test this claim, we manipulated the strength of activation of speakers' action simulations, and we examined whether this manipulation affected their gesture rates.

Method

Participants

Thirty-six students enrolled in Introductory Psychology at the University of Wisconsin-Madison volunteered to participate in exchange for extra course credit. All were native speakers of English. Data from two participants were discarded because they reported that they suspected the presence of the video camera or that gesture was the purpose of the experiment. Data from one additional participant were discarded because her posture made it difficult to see her gestures. Thus, the final sample was comprised of data from 33 individuals (26 female). Eighty-five percent of the participants self-identified as Caucasian, while the remaining participants self-identified as African American (9%), Asian (3%) or Hispanic (3%).

Stimuli

Six dot patterns from Hostetter et al. (2007) were used. Each pattern consisted of 6–9 black dots on a white background with lines connecting the dots to form a geometric configuration (see Fig. 1). All patterns were presented in black and white on an Apple laptop computer using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Procedure

Participants arrived individually for an experiment that they believed was about how different types of information are perceived, remembered, and described. They were told that their descriptions of various dot patterns would be audio taped so that another participant could later attempt to recreate the patterns.

Participants were randomly assigned to one of two condition orders. Half of the participants received the visual trials followed by the action trials. The other half received the action trials followed by the visual trials. For each set of trials, participants were given instructions, a practice trial, and three test trials, so that participants were not aware of the second trial type while completing the first. In both orders, the six individual patterns were presented in the same order. In one order, the first three were presented as visual trials and the last three were presented as action trials. In the second order, the conditions were reversed. Thus, each pattern was presented equally often in each condition across participants, but each participant saw each pattern only once, in either the visual or action condition, depending on the condition order to which he or she was assigned.

For the three visual trials, the computer displayed a pattern of dots and shapes for 10 s. Participants were asked to study the pattern so that they could remember it once it disappeared. Once the pattern disappeared from the screen, participants were prompted to describe the location of the dots they had seen in terms of the geometric shapes and figures. For example, when describing the first pattern shown in Fig. 1, participants often said something like: "There's a diamond with a dot on each of its corners. The bottom dot of the diamond is connected to the top dot of a triangle that also has dots on each of its corners".

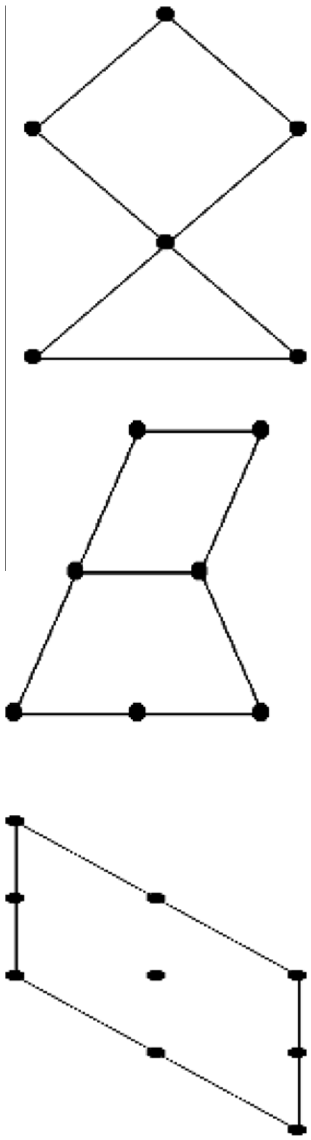

Fig. 1. Three of the six patterns used in the experiments.

For the action trials of the experiment, the computer displayed a pattern of dots and shapes for 3 s. Participants were then given round wooden pieces measuring 1/2 in. in diameter to place on the table in the configuration of the dots they had seen in the pattern. Once the pattern was made, the experimenter swept the wooden pieces off to the side and prompted the participant to describe the location of the dots they had seen in terms of the geometric figures.

At the outset of each set of trials (visual or action), the experimenter gave the participants an example of how to complete trials of that type. The example pattern description for each set included several scripted gestures. After the experimenter's demonstration, participants then completed a practice trial and received positive feedback from the experimenter. Participants then completed the three trials in that set at their own pace. Participants could take as long as they wanted to describe each pattern. They initiated each trial by pressing any key on the keyboard.

The design incorporated several features to equate communicative, lexical, conceptual, and memory demands across conditions. First, to equate communicative demands, speakers described the patterns in both conditions to an audio recorder with the belief that future participants would listen to the tape and try to reproduce the pattern. Although the experimenter was visibly present in both conditions, participants did not direct their descriptions to the experimenter in either condition. Thus, there was no apparent communicative reason for speakers to gesture more in one condition than in the other. Second, to equate lexical demands, participants described similar patterns that required access to similar lexical items (e.g., geometric terms, spatial prepositions, etc.) in both conditions. Third, to equate conceptual demands, the patterns in both conditions were presented with lines drawn to guide conceptualization, to rule out the possibility that the act of making the patterns in the action condition would aid conceptual processes and reduce gesture rates (see Hostetter et al., 2007). Finally, the memory demands of the two conditions were held constant by allowing participants in the action condition to look at the patterns for 3 s, which was shown in previous research to be enough time for participants to accurately encode these patterns (Hostetter et al., 2007). The pattern presentation time in the visual condition (10 s) was then chosen by adding the presentation time in the action condition (3 s) with the average amount of time taken by participants during pilot testing to make each pattern in the action condition (7 s). Thus, the total amount of time between originally viewing the pattern and describing it was approximately 10 s in both conditions. Note also that in neither condition was the pattern visually present as the participant described it.

Participants' descriptions were recorded with a hidden video camera. At the end of the experiment, participants were debriefed about the presence of the camera and the focus on gesture production. Participants were offered the opportunity to have their video data discarded; all declined.

Coding

Each participant's descriptions were transcribed and coded for accompanying manual gestures. Gestures were classified into two categories. *Representational gestures* are movements that depict the semantic content of a speaker's description. Individual representational gestures were segmented from the stream of manual activity based on changes in hand shape, motion, or meaning. For example, a motion made with the right index finger towards the left, diagonally up and then diagonally down while the speaker said, "There's a triangle" was coded as one representational gesture. A similar motion made diagonally upward with the words "there's a line that goes up," followed by a slight pause and then a motion diagonally downward with the words "and then a line that goes down" was coded as two representational gestures. *Beat gestures* are motorically simple, up-and-down movements that coincide with the rhythm or prosody of speech without conveying semantic content. For example, a small motion made up and down with both hands as a speaker said the word "house" in "it looks sort of like a house" was coded as one beat gesture. Other types of movements, including those not made with the hands or those that served a self-adapting function (such as fidgeting or touching clothing or hair), were not counted as gestures. Representational and beat gesture rates per 100 words were calculated for each participant's description of each pattern.

Reliability

One person initially coded the data from all speakers. To establish reliability, a second coder who was blind to the hypothesis under investigation coded one pattern from each participant (~16% of the data). Reliability between this blind coder and the primary coder was 84% for segmenting gestures from the stream of manual activity ($N = 149$). Both coders classified approximately 90% of all gestures as representational gestures. Because the estimated base rates of representational and beat gestures were unequal, Cohen's kappa could be used but its interpretation must be adjusted (see Bruckner & Yoder, 2006). Kappa was .55 for classifying each gesture as representational or beat; this should be interpreted as approximately 93% accuracy between coders, as estimated from Table 1 in Bruckner and Yoder (2006).

Results

According to the GSA framework, participants should produce representational gestures at a higher rate when they have physically constructed patterns than when they have only viewed patterns. A 2 (condition: action vs. visual) × 2 (order) mixed analysis of variance was conducted with condition as a repeated measure and representational gesture rate per 100 words as the dependent variable. There was a significant main effect of condition, $F(1, 31) = 9.69$, $p < .01$. As depicted in Fig. 2, participants produced more representational gestures per 100 words when describing patterns they had made ($M = 8.18$, $SD = 5.49$) than when describing patterns they had viewed ($M = 6.49$, $SD = 5.31$) ($M_D = 1.69$, 95% CI = 0.92, 2.46). There was no effect of condition order.

The prediction made by the GSA framework is specific to representational gestures. According to the GSA framework, there should not be an increase in participants' rate of beat gestures when they have physical experience making a pattern. If beat gestures are also affected, the increase in representational gesture rates may be due to a mechanism other than that proposed by the GSA framework. Thus, a second 2 (condition: action vs. visual) × 2 (order) mixed analysis of variance was done with beat gesture rate per 100 words as the dependent variable. There was no effect of condition on beat gesture rates, $F(1, 31) = 0.12$, $p = .73$. Participants produced comparable rates of beats

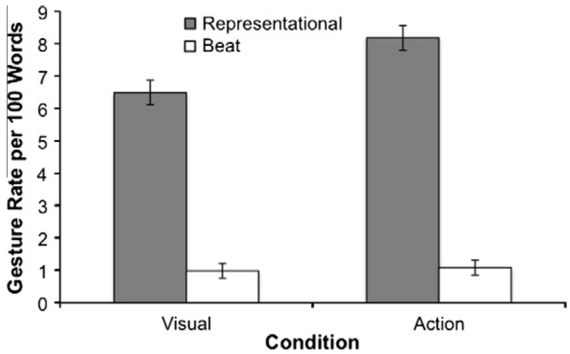

Fig. 2. Representational and beat gesture rates per 100 words in the visual and action conditions of Experiment 1. The error bars represent standard errors, using the MS_{error} for the condition × order interaction as the estimate of population variance.

in both conditions (action: $M = 0.98$, $SD = 1.59$; visual: $M = 1.08$, $SD = 2.17$, $M_D = 0.1$, 95% CI = −0.36, 0.56). There was, however, a main effect of order on beat gesture rates, $F(1, 31) = 4.7$, $p = .04$, such that beat gestures were more prevalent when speakers described the visual trials first.

Discussion

The Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008) proposes that gestures arise from action simulations produced in the interest of speaking. According to this hypothesis, speakers should gesture more when they are describing representations that are based on actions, and which therefore involve highly activated simulations of action, than when they are describing representations that are not as strongly tied to actions. The present experiment supports this claim by demonstrating that participants gesture at a higher rate when they describe a pattern they have made than when describing a pattern they have only viewed.

In this experiment we sought to equate the communicative, lexical, and memory demands involved in describing the patterns in the two conditions, because such demands have been shown to affect gesture rates in other studies (e.g., Alibali et al., 2001; Hostetter et al., 2007; Morsella & Krauss, 2004). For example, we did not want participants to form a clearer memory of the patterns they had constructed than of the patterns they had only viewed, because there is evidence that speakers gesture when they need help remembering the image they are describing (Wesp et al., 2001). Thus, if participants formed a memory of the visual patterns that was somehow less clear than the memory they formed of the action patterns, then they may have gestured in the visual condition in order to help themselves remember the harder-to-remember patterns. This would potentially work against the effect predicted by the GSA framework; while speakers may gesture a lot in the action condition because of the strong action simulations associated with the patterns they have made (as suggested by the GSA framework), speakers may also gesture a lot in the visual condition because of their weaker memories of the patterns. Thus, we intentionally gave participants as much time to view the patterns in the visual condition as it took participants to make the patterns (on average) in the action condition, in order to assure that participants knew the patterns equally well in the two conditions.

Although matching the cognitive demands in the two conditions is a strength of this experiment generally, it is not possible to simultaneously control for every possible confound. By equating the memory demands of the two conditions, we may have inadvertently given participants extra time for verbal planning in the visual condition. Participants had to physically construct the pattern during part of the time they were given to think about the pattern in the action condition, which may have made it difficult for them to plan how they would describe the pattern. In contrast, participants were given 10 s in the visual condition with nothing to do other than view the pattern and think about how they would describe it. This is potentially problematic because gestures are associated with difficulties in planning and producing speech (Christenfeld, Schachter, & Bilous, 1991; Hostetter et al., 2007; Krauss, 1998), meaning that speakers may have gestured more in the action condition than in the visual condition because their descriptions were less well planned. This is a very different explanation for the effect observed in Experiment 1 than the explanation proposed by the GSA framework.

In order to rule out this alternative explanation, we conducted a second experiment in which the amount of time given to participants to view (and verbally rehearse) the patterns in the visual condition was reduced. Specifically, participants in Experiment 2 viewed patterns in both conditions for only 3 s; in the visual condition, they described each pattern immediately afterwards and in the action condition, they constructed the pattern and then described it. If the results of Experiment 1 were due to speakers' greater opportunity for verbal planning in the visual condition, rather than to their more highly activated action simulations in the action condition, then the results should not replicate in Experiment 2, where the amount of time for verbal planning is held constant across the conditions. In fact, the effect may actually reverse in Experiment 2, because participants now have more time to verbally plan their descriptions in the action condition while they are making the patterns than in the visual condition where they must begin speaking immediately after seeing the patterns. In contrast, if gestures arise from more highly activated simulations of actions, as the GSA framework proposes, then the results of Experiment 2 should replicate those of Experiment 1; speakers should gesture more when they have physically constructed the patterns than when they have only viewed those patterns.

Experiment 2

Experiment 2 provided an additional test of the claim that gestures arise from highly activated action simulations (Hostetter & Alibali, 2008). As in Experiment 1, we manipulated the strength of activation of speakers' action simulations, and examined whether this manipulation affected their gesture rates. In this experiment, time for verbal rehearsal was equated across conditions.

Method

Participants

Thirty-four participants (10 male) enrolled in Introductory Psychology at Kalamazoo College volunteered to participate in exchange for extra course credit. Sixty-two percent of the participants self-identified as Caucasian. The remaining participants self-identified as African American (9%), Hispanic (9%), African (3%), Asian (3%), Native American (3%), Middle Eastern (3%), or mixed heritage (6%). All participants had learned English before the age of 5.

Procedure

The materials and procedure were identical to Experiment 1, with only one exception. In Experiment 2, the patterns remained on the computer screen for 3 s in both conditions (compared to 10 s in the visual condition of Experiment 1). In the visual condition, participants were instructed to begin their description of each pattern as soon as it disappeared from the screen. In the action condition, participants were instructed to place the wooden dots on the table in the locations of the dots they had seen in the pattern. As in Experiment 1, the wooden pieces were then cleared away and the participants were instructed to begin their description. Descriptions were recorded with a hidden video camera. At the conclusion of the experiment, participants were debriefed about the presence of the camera. Participants were offered the opportunity to have their video data discarded; all declined.

Coding and reliability

The descriptions were transcribed and coded for accompanying gestures following the same procedure employed in Experiment 1. One description from each speaker (~16 % of the data) was coded by a second coder who was blind to the purpose and hypothesis of the experiment. Reliability was 88% for segmenting gestures from the stream of manual activity ($N = 266$). Cohen's kappa for classifying each gesture as representational or beat was .75, which when adjusted for uneven base rates of each type of gesture, corresponds to 97% accuracy, as estimated from Table 1 in Bruckner and Yoder (2006).

Results

The Gesture as Simulated Action framework predicts that the effects of Experiment 2 will replicate those of Experiment 1. The results are depicted in Fig. 3. Representational gesture rates were analyzed with a 2 (condition) × 2 (order) analysis of variance with condition as a repeated measure. As predicted, the results mirror those of Experiment 1, with a significant main effect of condition, $F(1, 32) = 5.20$, $p = .03$. Speakers produced more representational gestures per 100 words after making the patterns ($M = 12.16$, $SD = 6.68$) than after viewing the patterns without making them ($M = 10.25$, $SD = 6.21$) ($M_D = 1.91$, 95% CI = 0.72, 3.11). There were no effects involving order.

As in Experiment 1, we also examined beat gesture rates. The 2 (condition) × 2 (order) mixed analysis of variance revealed no effect of condition on beat gesture rates, $F(1, 32) = .09$, $p = .76$. Speakers produced comparable rates of beat gestures across the two conditions (action: $M = 1.33$, $SD = 1.96$; vision: $M = 1.42$, $SD = 1.87$) ($M_D = .09$, 95% CI = −.48, .30). Unlike in Experiment 1, there was no main effect of order, although there was a marginal condition × order interaction, $F(1, 32) = 4.09$, $p = .051$. Participants who described the action patterns last produced fewer beats when describing them than when describing the visual patterns, but participants who described the action patterns first showed no difference in beat rates across the two conditions.

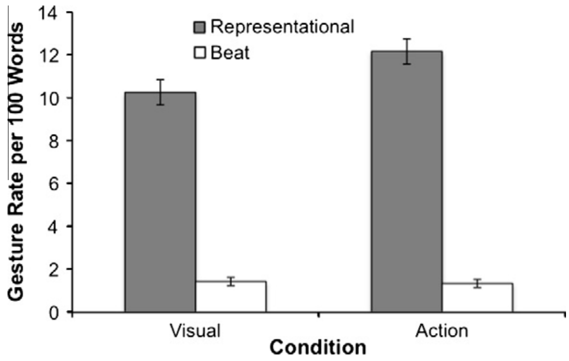

Fig. 3. Representational and beat gesture rates per 100 words in the visual and action conditions of Experiment 2. The error bars represent standard errors, using the MS_{error} for the condition × order interaction as the estimate of population variance.

Discussion

The results of Experiment 2 replicate those of Experiment 1. Speakers produce higher rates of representational gestures when describing patterns they have physically constructed than when describing patterns they have only viewed. This effect holds both when the memory demands of the two conditions are held constant by equating the amount of time participants have to think about the patterns before describing them (Experiment 1) as well as when verbal planning opportunities are held constant by equating the amount of time participants have to verbally rehearse in the two conditions (Experiment 2).

However, there is an alternative explanation for participants' increased gesture rates after making patterns that still remains tenable. Making the patterns in the action condition required participants to move their hands and arms, movement that was not required in the visual condition. Perhaps this movement "primed" participants to continue moving as they described the patterns. If participants' motor systems are activated while they are encoding a pattern, their motor systems may remain activated or become reactivated while they are describing the pattern, even if this motor system activation is completely unrelated to the representation they were forming and are now describing. We call this explanation the Primed Movement Hypothesis. There are at least two ways this priming could occur.

First, it is possible that it is difficult during retrieval to "shut off" any activation of motor cortex that occurred during encoding, particularly when the retrieval immediately follows encoding, as in the present design. As a result, participants may simply have more activity in their motor cortex when they have just been moving than when they have not been moving, and they may express this residual activation in a task relevant way, such as by producing representational gestures. According to this view, speakers should produce more representational gestures whenever they have moved during encoding, regardless of whether that movement was relevant to the representation being encoded. According to this view, the movement prime is quite general; if someone has moved during encoding, he or she will be likely to move during retrieval.

Second, it is possible that the motor priming is more specific. If a person learns something while moving, he or she may be primed to move in a similar way while recalling the learned information, in an effort to reactivate the general conditions that were involved in encoding as a retrieval cue (e.g., Dijkstra, Kaschak, & Zwaan, 2006). According to this view, the specific movement that occurs during encoding should be reproduced during retrieval, regardless of whether that movement is central to the representation that is being described. According to this view, the movement prime is more specific; if someone has moved in a particular way during encoding, he or she will be likely to move in a similar way during retrieval.

Both of these explanations could account for the effects observed in Experiments 1 and 2, and they differ from the explanation suggested by the GSA framework. Specifically, according to the GSA framework, it is not that speakers are simply expressing residual motor activation in a task relevant way when they produce representational gestures, nor is it that speakers produce representational gestures in an attempt to reactivate the conditions of encoding during retrieval. Rather, the GSA framework contends that in thinking about the representation they are describing, speakers simulate the actions that are integral to the representation. Thus, according to the GSA framework, movement during encoding should lead to increased gesture rates only if the movement is integral to the representation that is being formed. Movements that are not integral to the representation that is being formed should not be repeated as gestures during a speaker's description. According to both versions of the Primed Movement Hypothesis, on the other hand, if there is movement during encoding, there should be movement during recall. This movement may be transformed into task relevant movement, or it may more specifically mimic the movement produced during encoding.

To rule out these alternative explanations, we carried out a third experiment. In this experiment, we encouraged participants to move as they viewed the patterns in the visual condition, but to do so in a way that would not strengthen the action simulations activated when they thought about the patterns. Toward this end, we asked participants to tap their fingers on the table as they viewed the patterns in the visual condition.

If the general version of the Primed Movement Hypothesis is correct, then any movement produced during encoding, even if irrelevant to the representation being formed, could prime the motor system and be manifested as task relevant movement during retrieval. According to this view, participants should produce comparable rates of representational gestures when they have tapped as when they have constructed the patterns, because both tapping and constructing generally prime the motor system. On the other hand, according to the GSA framework, speakers produce representational gestures when the action simulations relevant to the particular imagistic representation being described are strongly activated. Constructing the patterns should increase the strength of action simulation, because it encourages speakers to construe the patterns in terms of the actions required to construct them. In contrast, the irrelevant tapping movements should not strengthen speakers' simulations of the actions required to construct the patterns. Thus the GSA framework predicts that speakers should produce representational gestures at higher rates when they have made the patterns compared to when they have tapped while viewing them.

To address the more specific version of the Primed Movement Hypothesis, we carefully considered the nature of the gestures speakers produced in the two conditions. If the specific version of the Primed Movement Hypothesis is correct, then speakers should recreate during retrieval whatever movements they engaged in during encoding. According to this view, speakers should produce more gestures that resemble a tapping movement when they have tapped during the construction of the patterns. Such tapping movements would likely be manifested as beat gestures during speech, because beat gestures are simple up-and-down movements. Thus, in contrast to the GSA framework, the specific version of the Primed Movement Hypothesis predicts that speakers should produce more beat gestures when describing patterns that they learned while tapping than when describing patterns that they constructed. Further, the rate of beat gestures produced in the tapping condition of Experiment 3 should be much greater than the rate of beat gestures produced in any of the previous experiments. The GSA framework does not make any specific predictions regarding the rate of beat gestures.

Experiment 3

The purpose of Experiment 3 was to test the claim that the findings of Experiments 1 and 2 were due to a prime to move in the action conditions. To test this claim, we compared gesture rates in an action condition (comparable to Experiments 1 and 2) and in a motor tapping condition, in which participants performed a movement that did not strengthen their action representations about how to construct the patterns.

Method

Participants

Thirty-three participants (16 male) were recruited from Introductory Psychology classes at Kalamazoo College. All received extra course credit in exchange for participating. The majority of participants (82%) self-identified as Caucasian, with a minority of the sample self-identifying as Asian (6%), Hispanic (6%), African American (3%) or Middle Eastern (3%). All participants were native speakers of English.

Procedure

The materials were identical to those used in Experiments 1 and 2. Participants completed three trials in the action condition, which were exactly the same as the action trials in Experiments 1 and 2. Participants also completed three trials in a motor tapping condition. In this condition, the pattern was displayed on the screen for 10 s, during which it disappeared from the screen for 20 ms once every second and a beep sounded. Participants were told to tap their fingers on the table every time the pattern beeped. The experimenter demonstrated how to do this. As in Experiments 1 and 2, condition order was counterbalanced across participants. Descriptions were recorded with a hidden video camera. At the conclusion of the experiment, participants were debriefed about the hidden camera and offered the opportunity to have their video data discarded; all declined.

Coding and reliability

The descriptions were transcribed and coded following the procedure used in Experiment 1. A second coder established reliability by recoding the data from one description from each participant (~16% of the data). Reliability for identifying gestures from the stream of manual activity was 89% ($N = 254$). Cohen's kappa for classifying each gesture as representational or beat was .68, which when adjusted for uneven base rates of each type of gesture (81% of gestures were coded as representational gestures), corresponds to 92% accuracy, as estimated from Table 1 in Bruckner and Yoder (2006).

Results and discussion

If the results of Experiments 1 and 2 are due to the increased activation of action simulations that comes from physical experience making the patterns in the action condition, then the results for representational gestures in Experiment 3 should replicate those of Experiments 1 and 2. However, if the increase in representational gestures is due instead to general priming of the motor system that is manifested in a task relevant way during retrieval, then speakers should produce comparable rates of representational gestures regardless of whether they constructed or tapped while learning the patterns. A 2 (condition: tapping vs. action) × 2 (order) analysis of variance was conducted with condition as a repeated measure and representational gesture rate per 100 words as the dependent variable. As in the first two experiments, a main effect of condition was found, $F(1, 31) = 7.36$, $p = .01$. Speakers produced representational gestures at a higher rate when they had physically made the patterns ($M = 9.93$, $SD = 7.95$) than when they had tapped while viewing the patterns ($M = 7.95$, $SD = 6.46$) ($M_D = 1.98$, 95% CI = 0.92, 3.04), as seen in Fig. 4. There were no effects involving order.

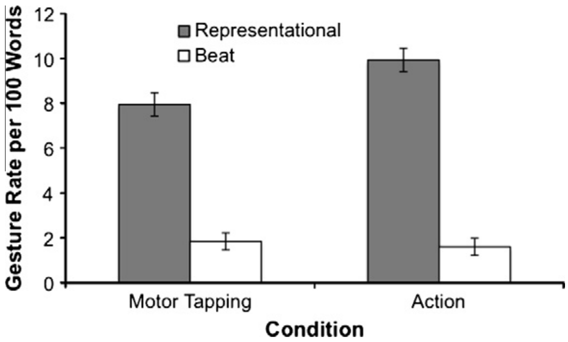

Fig. 4. Representational and beat gesture rates per 100 words in the motor tapping and action conditions of Experiment 3. The error bars represent standard errors, using the MS_{error} for the condition × order interaction as the estimate of population variance.

If the results were due to a specific prime to recreate whatever movement was enacted during encoding during retrieval, then the tapping condition in Experiment 3 should have led to an increase in the rate of beat gestures. This possibility was addressed with a 2 (condition) × 2 (order) analysis of variance with condition as a repeated measure and beat gestures per 100 words as the dependent variable. No main effect of condition was found on beat gesture rates, $F(1, 31) = 0.16$, $p = .69$, and there were no effects involving order. Speakers did not produce more beat gestures per 100 words when they had tapped during encoding ($M = 1.84$, $SD = 2.71$) than when they had constructed the patterns ($M = 1.60$, $SD = 2.64$).

Next, we directly compared the effects of condition on representational and beat gesture rates across the three experiments. First, we compared representational gesture rates in a 2 (condition) × 2 (order) × 3 (experiment) repeated measures ANOVA. Importantly, there was no experiment × condition interaction, $F(2, 94) = 0.44$, $p = .95$. The increase in gesture rates when participants made the patterns was comparable in all three experiments, regardless of the specific controls used in each. Not surprisingly, then, there was a main effect of condition across the three experiments, $F(1, 94) = 20.14$, $p < .001$. There was also a main effect of experiment, $F(2, 94) = 3.48$, $p = .04$, such that participants in Experiment 2 produced representational gestures at higher rates than participants in the other studies. We suspect that this is most likely due to extraneous factors related to the particular sample and experimenter involved in Experiment 2, rather than to anything about the specific manipulation used in that study. There were no effects involving condition order.

Finally, we compared the effect of condition on beat gesture rates across the three experiments with a 2 (condition) × 2 (order) × 3 (experiment) repeated measures ANOVA. Recall that the specific version of the Primed Movement Hypothesis predicts that speakers would produce more beat gestures when they had tapped in Experiment 3 than in any of the other conditions. Contrary to this prediction, there was no condition × experiment interaction, $F(2, 94)$, = 0.03, $p = 97$. However, there was a condition × order interaction, $F(1, 94) = 4.66$, $p = .03$. Overall, speakers produced more beat gestures per 100 words in whichever condition they completed first. There was also an order × experiment interaction, $F(2, 94) = 4.26$, $p = .02$. In Experiment 1, speakers produced more beat gestures per 100 words when they had the visual condition first; in Experiments 2 and 3, speakers produced more beat gestures per 100 words when they had the action condition first. These effects involving order are not readily interpretable or theoretically interesting, and they will not be discussed further.

General discussion

Overall summary

The Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008) holds that representational gestures arise when speakers activate representations during speech production that evoke action simulations. Experiment 1 supported this claim by demonstrating that speakers gesture more when they describe information they have learned via physical, manual actions than when they describe information they have learned via vision. Experiment 2 replicated this finding and ruled out the possibility that the effect in Experiment 1 was due to differences in opportunities for verbal planning across the two conditions. According to the GSA framework, the effect arises because action simulations, which are the precursors to representational gestures, are more highly activated by related actions than by related visual experience.

However, taken alone, the results of Experiments 1 and 2 could be interpreted in another way, referred to here as the Primed Movement Hypothesis. According to this alternative hypothesis, once speakers have been primed to move by physically constructing the patterns, they may continue to move during their descriptions by producing more representational gestures. This may be because of a general prime to their motor system that is manifested in a task relevant way (e.g., representational gestures) or because of a more specific prime to recreate the motor conditions of encoding at retrieval. There are two reasons to doubt that such an explanation can fully account for the results found in Experiments 1 and 2.

First, the increase in gesture rate when speakers made the patterns in all three experiments was specific to representational gestures. If speakers were simply primed to move more in general, it would be difficult to explain why they specifically increased their rate of representational gestures after making the patterns, rather than their rate of both representational and beat gestures. Further, in the tapping condition of Experiment 3, participants had produced beat-like movements during encoding. If speakers reproduce the specific movements they produced during encoding as a retrieval cue, then speakers should have produced beat gestures at a much higher rate in this condition than in any of the other conditions tested in these experiments. However, this was not the case; beat gesture rates in the tapping condition of Experiment 3 were comparable to those in the action condition of Experiment 3, as well as to the action and visual conditions of Experiments 1 and 2. The GSA framework clearly predicts that the effect in all three experiments will be unique to representational gestures, because representational gestures are the specific type of gesture thought to arise from simulations of actions.

The second reason to doubt that the Primed Movement Hypothesis can fully account for these results is because the effect occurred even when speakers had been primed to move in the motor tapping condition of Experiment 3. If speakers produce representational gestures because the gestures draw on residual motor activity in a task relevant way, then representational gestures should be produced at comparable rates whenever speakers move as they encode a pattern. However, Experiment 3 did not support this hypothesis; speakers gestured more when they had physically made the pattern than when they had tapped while viewing it. Thus, it seems that the effects found here are not readily explained by a general or specific movement prime. Rather, we argue that the experience of making the patterns in all three experiments increased the extent to which the speakers construed the patterns in terms of actions. These actions were simulated during speakers' thinking about the patterns and were partially recreated during their descriptions through representational gestures.

Potential alternative explanations

While the present data speak against the two versions of the Primed Movement Hypothesis considered here, there is a third version of the hypothesis that must be considered. Perhaps making the patterns primed speakers specifically to make meaningful movements, which could be expressed as representational gestures during description. This version of the primed movement hypothesis cannot be ruled out by the present data. However, this view is compatible with, and possibly even indistinguishable from, the GSA framework. In order to produce a meaningful movement, a speaker would have to think about the meaning of the information he or she is describing in terms of movement. This is identical to the claim made by the GSA framework that speakers simulate the action components of their representations. Thus, priming meaningful movement may be essentially the same as priming action simulations. It is unclear how one could prime speakers to produce meaningful movements without also priming them to think about their representations in terms of action.

It may indeed be the case that, if speakers make one pattern and then immediately afterwards describe a different pattern, they may gesture more than if they had not initially made the first pattern. At first glance, such a finding would not be predicted by the GSA framework, because the actions produced are not directly relevant to the pattern being described, and therefore should not increase the activation of the action representation of that pattern. However, it is possible that the cognitive process of action simulation can be primed; that is, perhaps making one pattern actually primes speakers to think about future patterns in terms of how they would be made. If such priming of action simulation exists, then an increase in gesture rates when describing one pattern after making a different pattern would not be incompatible with the GSA framework. More research is needed to determine whether action simulation can be primed in this general way.

There is a second alternative explanation for the present findings that should also be considered. In the action conditions in the present experiments, participants had two modes of experience with the information (vision and action), whereas in the visual conditions, participants only had one mode of experience with the information (vision). It may be the case that any additional experience with the patterns would lead to increased gesture rates, and that specific experience with action is not required. Additional experience of any type (action or non-action) may lead to deeper processing, creating a richer mental representation, and this richer mental representation may afford more gesture than a more impoverished one based on only one mode of experience. Indeed, there is some evidence to suggest that speakers gesture more when they have more detailed or robust spatial representations (e.g., Hostetter & Alibali, 2007); however, there is also some evidence to suggest that speakers may gesture more when they have less well-formed representations (e.g., Church & Goldin-Meadow, 1986; Hostetter et al., 2007; Wesp et al., 2001). Thus, it is not clear that such an explanation would predict the present findings a priori.

Further, the GSA framework does not discount the importance of additional experience with action as one reason why speakers gesture more after making the patterns. The additional experience of making the patterns, according to the GSA framework, strengthens the action representation of the pattern, thereby leading to more gestures. Future research should contrast conditions in which participants make the patterns with conditions in which participants are given additional perceptual experience with the patterns that does not involve action. At the present time, however, it seems fair to think of these two views as overlapping theoretically; speakers produce representational gestures when the action components of their representation are richer and more well-developed.

Although the purpose of the manipulations we used in all three experiments was to contrast the gestures produced while describing motor images to those produced while describing visuo-spatial images, it is also possible that the manipulations affected imagery (and thereby gesture rates) in a slightly different way. Specifically, the GSA framework follows Cornoldi and Vecchi (2003) in conceptualizing imagery as a continuum involving various amounts of action activation (see Hostetter & Alibali, 2008, Fig. 3). Visual and motor imagery are the poles of this continuum, with visual imagery relying on relatively low action simulation and motor imagery relying on relatively high action simulation. Spatial imagery lies midway between these poles. In contrast to visual images that are comprised of mental pictures resembling passive visual experience, spatial images are comprised of dynamic knowledge of spatial relationships. For example, imagining a complete triangle with a particular color, size, and orientation would involve visual imagery, whereas imagining the three corners of a triangle, where they are located and how far apart they are, would involve spatial imagery. According to Cornoldi and Vecchi, spatial imagery involves a stronger motor component than does visual imagery, and according to the GSA framework, descriptions of spatial images should therefore be accompanied by higher rates of representational gestures than descriptions of visual images.

It is possible that the construct vs. view manipulation affected the specific type of visuo-spatial images that speakers formed. In particular, constructing the patterns with the wooden pieces in the action condition may have encouraged participants to form an image of the pattern that was spatial rather than visual, as placing the dots in specific locations may have focused their attention on the spatial components of the image. It is possible, therefore, that the difference we observed in gesture rates between the conditions is due to a difference in the nature of the visuo-spatial imagery that was formed (passive visual vs. dynamic spatial) rather than a difference between motor and visuo-spatial imagery more generally. Further research is needed to disambiguate these two possibilities, but according to the GSA framework, both motor imagery and dynamic spatial imagery should lead to higher gesture rates than passive visual imagery. Thus, regardless of the way in which the manipulation altered participants' thinking about the patterns—increasing action activation or promoting spatial rather than visual imagery—the present findings provide unambiguous support for the GSA framework.

Gesture rates in the visual conditions

In these experiments, although participants gestured more when they had physically constructed the patterns than when they had only viewed the patterns, participants gestured at a fairly high rate even when they had simply viewed the patterns (6.5 representational gestures per 100 words in Experiment 1; 10.3 and 7.9 in Experiments 2 and 3). If gestures arise from simulated actions, as the GSA framework proposes, then why did speakers gesture so frequently even when they had no manual experience constructing the patterns?

According to the GSA framework, and building on the idea that perception and action are tightly linked in a cycle of coordination (e.g., Dewey, 1896), visual imagery does involve some action simulation, as it is tied to the actions undertaken to form the image as well as actions that may be planned to manipulate or interact with the imagined object. Thus, it is not the case that there were no action simulations involved when participants described the patterns in the visual condition. Rather, participants likely evoked action plans in response to the patterns they perceived, such as plans for how they might draw or manipulate the shapes in the pattern. Similarly, participants likely moved their eyes in particular ways in order to perceive the patterns.[2]

However, in the action condition, when participants had to physically recreate the patterns, the ties between

[2] At the present time it is not known how the eye and head movements involved in visual perception are translated into a motor plan that can be readily enacted by the manual system as gesture. This is an important arena for future work.

these action plans and the patterns were enhanced. While action may have been a subtle part of the way speakers encoded and remembered the patterns in the visual condition, actually recreating the patterns forced speakers to think about the specific actions involved in making the patterns. According to the GSA framework, this increased activation of action simulations after participants had experienced making a particular pattern led to the increase in representational gesture rates as they described the pattern.

Comparison to other theories about the source and function of gestures

The GSA framework makes a very specific claim about the cognitive source of representational gestures: that they stem from imagistic representations that have close ties to action. This claim differs from those made by several other theories that consider how gestures arise. For example, some theories propose that gestures arise from elementary spatial features (Krauss, Chen, & Gottesman, 2000) rather than images per se; that gestures arise from imagery that has been linguistically categorized (McNeill, 1992, 2005); or that gestures arise from propositional representations (Wagner et al., 2004).

The theory most notably in conflict with the present data is the idea that gestures arise from propositional representations (Wagner et al., 2004). It is difficult to understand how the propositional representations underlying the patterns in the action condition and those underlying the patterns in the visual condition would have been different in the present paradigm. However, the data that suggest that gestures may arise from propositional representations (see Wagner et al., 2004) are specifically relevant to deictic gestures. Although the GSA framework considers deictic gestures as a subcategory of representational gestures, it is possible that deictic gestures are less closely tied to simulations of actions than are other types of representational gestures. Very few of the gestures produced in the present experiments were deictic; rather, the majority of gestures observed here were iconic depictions of spatial properties or shapes. The present data should thus be taken as evidence that iconic gestures arise in part from action simulations; the data do not rule out the possibility that deictic gestures may arise from propositional representations that are unrelated to action simulations.

The focus of the present experiment has been on testing one specific claim made by the GSA framework about how gestures arise. The question of how gestures arise can also provide insight about what gestures do. Many researchers have suggested that representational gestures play a facilitative role in the speech production process, by priming lexical items (e.g., Krauss, 1998; Krauss et al., 2000), by organizing spatial information (e.g., Hostetter et al., 2007; Kita, 2000; Melinger & Kita, 2007), by preventing decay in visuospatial working memory (e.g., de Ruiter, 1998; Wesp et al., 2001), by offloading cognitive effort (e.g., Goldin-Meadow et al., 2001; Wagner et al., 2004), or by communicating spatial information directly (e.g., Kelly & Church, 1999; Kendon, 1994, 2004). These facilitative functions are not mutually exclusive, and indeed, there is independent evidence supporting each of them. The GSA framework does not aim to support one of these views over the other. Instead, by focusing on what gestures are, the GSA framework seeks to explain how it is that gestures have facilitative effects at all, regardless of exactly what those facilitative effects are.

The idea that doing something extra (e.g., planning and producing a gesture) can actually relieve cognitive demands and ease the burden of speaking (e.g., Goldin-Meadow et al., 2001; Wagner et al., 2004) is a bit of a conundrum. However, if gestures arise directly from the motor components of a representation, as the GSA framework claims, then realizing those components as gesture should not require any extra cognitive resources. Rather, inhibiting this motor activation from being expressed in gesture may be more difficult than simply producing gestures, as some evidence suggests (Goldin-Meadow et al., 2001). Thus, by understanding what gestures *are*, we may gain insight into what gestures *do*, and how they can be helpful in everyday communicative situations.

Conclusion

The present experiments have directly tested and supported the claim that representational gestures arise when actions are simulated in the interest of thinking and speaking. The findings provide direct empirical support for the GSA framework (Hostetter & Alibali, 2008) and as such, they suggest that the explanation for speech-accompanying gestures provided by the GSA framework is viable. Although future studies are needed to test some of the framework's other claims, the present data provide compelling evidence that gestures do arise from simulated actions. Thus, it seems that two of the ways in which our motor systems support speech production – action simulation and representational gestures – may be intimately related to one another.

Acknowledgments

This work was initiated as part of Autumn B. Hostetter's dissertation at the University of Wisconsin-Madison. We thank Arthur Glenberg, Charles Kalish, Mitchell Nathan, and Charles Snowdon for helpful discussion and comments on the design of these experiments. We thank Andrea Potthoff for her help in running Experiment 3. We also thank Anne Bartholomew, Chelsea Baumgarten, Elizabeth Farley, and Erin Sullivan for their assistance with transcription and coding reliability.

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.
Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.
Bruckner, C. T., & Yoder, P. (2006). Interpreting kappa in observational research: Baserate matters. *American Journal of Mental Retardation, 111*, 433–441.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1*, 149–219.

Christenfeld, N., Schachter, S., & Bilous, F. (1991). Filled pauses and gestures: It's not coincidence. *Journal of Psycholinguistic Research, 20*, 1–10.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*, 43–71.

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.

Cornoldi, C., & Vecchi, T. (2003). *Visuo-spatial working memory and individual differences*. New York: Psychology Press.

de Ruiter, J. P. (1998). *Gesture and speech production.* Doctoral dissertation, Nijmegen, The Netherlands: Katholieke Universiteit Nijmegen.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review, 3*, 357–370.

Dijkstra, K., Kaschak, M. P., & Zwaan, R. A. (2006). Body posture facilitates retrieval of autobiographical memories. *Cognition, 102*, 139–149.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology, 91*, 451–471.

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.

Gerwing, J., & Bavelas, J. B. (2004). Linguistic influences on gesture's form. *Gesture, 4*, 157–195.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1–55.

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

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

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in the motor and premotor cortex. *Neuron, 41*, 301–307.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (in press). Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychological Science*.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology, 26*, 4–27.

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and representational gesture production. *Gesture, 7*, 73–95.

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.

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*, 228.

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage, 14*, S103–S109.

Kelly, S. D., Barr, D. J., Church, R., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language, 40*, 577–592.

Kelly, S. D., & Church, R. (1999). A comparison between children's and adult's ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*, 85–93.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction, 27*, 175–200.

Kendon, A. (2004). *Gesture: Visible action as utterance Cambridge*. Cambridge University Press.

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

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

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge, UK: Cambridge University Press.

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

McNeill, D. (2005). *Gesture and thought*. Chicago: The University of Chicago Press.

Melinger, A., & Kita, S. (2007). Conceptualization load triggers gesture production. *Language and Cognitive Processes, 22*, 473–500.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology, 117*, 411–424.

Morsella, E., & Krauss, R. M. (2005). Muscular activity in the arm during lexical retrieval: Implications for gesture–speech theories. *Journal of Psycholinguistic Research, 34*, 415–427.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Brain and Behavioral Sciences, 24*, 939–1031.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46*, 688–704.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews, 6*, 1–6.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21*, 793–797.

Sperry, R. (1952). Neurology and the mind–brain problem. *American Scientist, 40*, 291–312.

Spivey, M. J., Richardson, D. C., & Gonzales-Marquez, M. (2005). On the perceptual-motor and image-schematic infrastructure of language. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 246–281). Cambridge, UK: Cambridge University Press.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*, 273–281.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 830–846.

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.

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

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

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain & Language, 101*, 278–289.

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 397–412.