

Gesture During Mental Rotation

Autumn B. Hostetter (Autumn.Hostetter@kzoo.edu)

Department of Psychology, 1200 Academy Street
Kalamazoo, MI 49006 USA

Martha W. Alibali (mwalibali@wisc.edu)

Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Anne E. Bartholomew (aebartholome@wisc.edu)

Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

Speakers gesture at high rates when explaining their solutions to spatial problems. The present work investigates one possible explanation for why speakers gesture so frequently during spatial problem solving, namely, that when speakers imagine the problem components in motion, they are particularly likely to gesture. We compared speakers' gesture rates as they actively solved a mental rotation problem and as they described the end state of the problem. Speakers gestured at a higher rate in the rotation condition. Thus, it appears that thoughts about the problem pieces in motion in the rotation condition do lead to increased gesture rates.

Keywords: gesture; spatial problem solving; mental imagery; simulation

Introduction

Speakers often gesture when they describe their solutions to spatial problems (e.g., Chu & Kita, 2008, 2011; Goldin-Meadow & Beilock, 2010). As one example, Hegarty, Mayer, Kriz, and Keehner (2005) found that approximately 90% of participants spontaneously gestured when asked to describe their solutions to mental animation problems. What is it about describing problem solutions that elicits gesture so reliably? The present work addresses this question.

One possible explanation is that gestures are simply a natural and effective way of expressing spatial information (e.g., Alibali, 2005; Emmorey & Casey, 2001; Kendon, 2004). The "expressive possibilities" of gesture (Kendon, 2004) are well suited for expressing information about actions and spatial relationships, so such information is very likely to be expressed in gestures.

Another possible explanation is that participants reactivate the actions involved in solving the problems during description, and that some aspects of these reactivations are expressed as gestures. Indeed, there is some evidence that the particular form of gestures that occur during problem descriptions appears to be related to the form of the action involved in solving the problem. For instance, Cook and Tanenhaus (2009) found that the specific form of the gestures produced by speakers as they described

their solutions to the Tower of Hanoi problem was related to whether they had solved the problem on a computer or by physically manipulating discs. When speakers had solved the problem with physical discs, they were much more likely to produce gestures that arced up and over (as though lifting the pieces) than when they had solved the problem on the computer.

The co-occurrence between the form of an action used in a particular problem and the form of the gesture used in descriptions of the problem is predicted by a recent account of gesture called the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). According to this view, gestures arise as outward manifestations of the simulations that occur during language production. Simulations are neural reenactments of perceptual and motor states that occur during interaction and experience with the world (see Barsalou, 2008; 2009). When speakers think about and describe experiences they have had, simulations of the experience cause cortical activation in the same motor and perceptual areas of the brain that were involved during the initial experience. The GSA framework claims that this cortical activation supports the production of gestures that partially or fully reenact the actions produced in the initial experience.

One implication of the GSA framework is that simulations that involve motor activation should be especially likely to result in gestures. The GSA framework distinguishes between simulations of motor imagery, which involve imagination of the body or percepts in motion, and simulations of visual imagery, which involve imagination of static percepts. Because motor imagery is particularly likely to rely on activation in motor areas of the brain (Richter et al., 2000), motor imagery should be more likely to result in gesture than visual imagery, other things being equal. Indeed, in the first experimental test of the GSA framework, Hostetter and Alibali (2010) demonstrated that speakers gesture at a higher rate when they have specific, relevant motor experience with the information they are describing than when they do not.

From the perspective of the GSA framework, then, the prevalence of gestures with descriptions of spatial problems,

such as mental animation problems (e.g., Hegarty et al., 2005) or gear problems (Schwartz & Black, 1999) may be because speakers are particularly likely to imagine the *actions* of the problem elements as they are describing them. Consistent with this view, past research has shown that speakers produce gestures more often with speech about “spatial transformations” (such as actions) which involve change over time, than with speech about static spatial relationships (Trafton, Trickett, Stitzlein, Saner, Schunn, & Kirschenbaum, 2005).

This view contrasts with the alternative view that speakers may gesture with descriptions of spatial problems because gestures are a natural way of expressing the spatial elements of the problem, regardless of whether those elements are imagined in motion (e.g., Kendon, 2004).

The present study will test these possibilities by comparing the gestures speakers produced as they solve mental rotation problems to the gestures they produce when they describe the end states of the rotations. If gestures occur during problem solving primarily because gestures are adept at expressing spatial information, then speakers should gesture at similar, high rates, whether they are describing the end state of a problem or actually solving the problem. On the other hand, if gestures occur during problem solving as the result of simulations of action, such as imagining the problem elements in motion, then gestures should occur more frequently when speakers are actively solving the problems than when they are describing the end states of the rotations.

Method

Participants

Thirty-eight undergraduates (13 male) at the University of Wisconsin-Madison volunteered to participate in exchange for extra credit in their Psychology course. All participants were native English speakers, and 85% claimed Caucasian descent. The remaining participants were African American (5%), Hispanic (5%), or mixed ethnicity (5%). The average age of the participants was 18.9 years.

Stimuli

Ten patterns, each containing two arrows, were constructed for the experiment. Each pattern depicted a blue arrow and a purple arrow positioned in two of the four quadrants of the screen. Each arrow pointed either straight up, down, left, right, or diagonally toward one of the four corners of the screen. Each pattern was paired with a rotation direction (clockwise or counterclockwise) and rotation angle (45, 90, or 135 degrees). These patterns and the accompanying rotation angles were shown to participants in the Rotation condition. A second version of each pattern was designed that showed each arrow correctly rotated to its end position. See Figure 1. These versions of the patterns were shown to participants in the Non-Rotation condition. Four similar patterns were created and used as examples and practice patterns. The patterns were presented using Psyscope

software (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Macintosh notebook computer.

Procedure

Participants were told that the experiment was about their ability to transform and describe mental images. Participants described 5 patterns from each of two conditions: Rotation and Non-Rotation. In the Rotation condition, participants were instructed to imagine the arrows in the pattern as they would appear rotated either an eighth of a turn, a quarter turn, or three eighths of a turn clockwise or counterclockwise. During the instructions, the experimenter explained that these increments correspond to 45 degrees, 90 degrees, and 135 degrees of rotation, although throughout the rest of the experiment, degree terminology was avoided in order to discourage participants from describing the orientation of the patterns in terms of degrees. In the Non-Rotation condition, participants were instructed to describe the pattern exactly as it was presented.

During each trial, the pattern appeared on the screen for 4 sec. The pattern then disappeared and was replaced by instructions, which depended on the condition. In the Rotation condition, the instructions included a fraction indicating the degree of rotation (either 1/8, 1/4, or 3/8) and a picture of a semi-circle (like a piece of pie) that corresponded to the degree of rotation. Additionally, an arrow above the piece of pie indicated direction of rotation: right for clockwise and left for counterclockwise. In the Non-Rotation condition, the words “Do not rotate” appeared on the screen. In both conditions, the instructions remained on the screen for 1 sec. As soon as the instructions disappeared, a beep sounded and the word “Describe” appeared on the screen. Participants were instructed to begin describing as soon as the beep sounded. Participants were asked to describe the location, orientation, and color of each arrow in every pattern and to take as much time as they needed for each description. See Figure 1.

Each participant saw and described each pattern only once, in either the Rotation or the Non-Rotation condition. Participants were randomly assigned to one of two presentation orders. The 10 patterns were in identical serial positions across the orders but varied in whether they were the Rotation or Non-Rotation version. Each pattern in the first order that was presented in the Rotation condition was presented in the Non-Rotation condition in the second order and vice versa. This resulted in similar descriptions for each pattern across the two orders. For example, participants in order 1 described the first pattern in its final position after it had been rotated; participants in order 2 described this same pattern in its final position while imagining it rotating.

Before the start of the first trial, the experimenter described an example pattern in each condition. The experimenter produced comparable scripted gestures during the example in both conditions. The participant then viewed and described two practice patterns before beginning the first experimental trial.

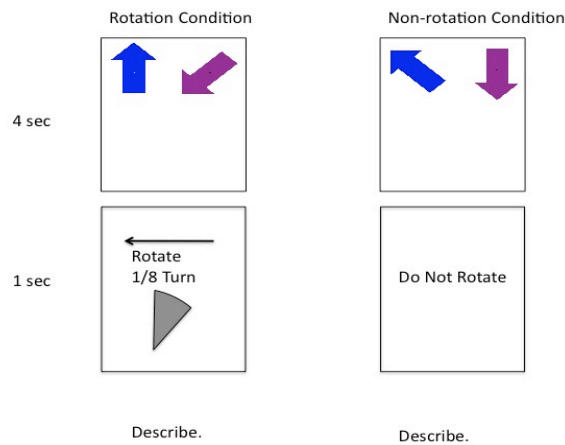


Figure 1. A sample pattern and instructions in the Rotation Condition and the corresponding pattern in the Non-rotation condition. The arrow on the left was blue and the arrow on the right was purple.

All descriptions were audio taped with the participants' knowledge. In addition, a hidden video camera recorded all speech and gesture produced during the task. Following the last trial, all participants were told about the hidden camera and the interest in gesture production and were given the opportunity to withdraw their video data from the study. All participants consented for their video data to be included.

Coding. All speech was transcribed. Gestures were coded by a research assistant who was unaware of the experimental condition of each description. All representational gestures were identified and coded. Representational gestures were defined as hand movements that represent semantic meaning by virtue of handshape or motion trajectory. For example, when a speaker said "it's pointing straight up" while holding the index finger of her right hand up and then moving the entire hand upwards vertically, this was coded as one representational gesture. Representational gestures were counted if they occurred during the description or if they occurred before the speaker began talking, either in the one-second time period when the instructions were on the screen, or after the beep but before the speaker began his or her description. Gestures that occurred before speech were rare, and they account for less than 5% of the data.

Each speech description was also coded for accuracy. We considered whether the description of each arrow accurately conveyed the location and orientation of the arrow. Responses that began by incorrectly describing either the location or orientation of the arrow were not considered correct, even if the speaker later self-corrected. Responses that were incomplete because they did not describe both the orientation and the location of the arrow were also considered incorrect. Finally, descriptions that were ambiguous were also considered incorrect. For example, if a speaker said, "The purple arrow is in the left corner," without specifying whether it was in the top left or bottom

left, the description was not considered correct. Thus, only descriptions that were both accurate and precise were considered correct.

Results

Description Accuracy

We first analyzed the accuracy of participants' descriptions across the two conditions. Participants may make more mistakes when describing patterns they had to rotate than when describing the patterns they did not have to rotate.

To analyze accuracy of spoken descriptions across the conditions, whether each pattern was described accurately or not was used as the dependent variable in a mixed logit regression analysis with condition (Rotation vs. Non-rotation) as a fixed effect and participant and pattern as random effects. We utilized logistic regression rather than analysis of variance because of the conceptual issues involved in treating proportional data as continuous (see Jaeger, 2008). The logistic regression model explains more variance than a model that includes only participant and pattern as random effects (and does not include condition), $\chi^2(1) = 50.48, p < .001$. The odds of an accurate description were 5.75 times higher in the Non-rotation condition than in the Rotation condition ($\beta = 1.75, SE = 0.25, z = 7.10, p < .001$). Thus, speakers were clearly less accurate in describing the end state of the rotation when they had to generate that end state on their own than when it was shown to them.

The large number of errors could affect gesture production in the rotation condition for two reasons. First, when speakers realize they have made an error, they may be more likely to produce a gesture than if they have not made an error. Second, recall that the patterns shown to participants in the Non-rotation condition were the end states of the corresponding patterns shown in the Rotation condition, so that when the patterns are described correctly, the speech in the two conditions should be comparable. However, if speakers make many more errors in one condition than in the other, the speech is no longer controlled across the two conditions, and any differences in gesture could be due to differences in the speech produced.

In order to remove this potential confound, the analysis was limited to only those descriptions that were accurate. However, in order to avoid losing too much of the data, particularly in the Rotation condition, descriptions were excluded by arrow rather than by entire pattern. For example, if a participant described the purple arrow correctly in pattern 1 but not the blue arrow, the words and gestures that occurred during the description of the blue arrow were excluded, but the words and gestures regarding the purple arrow were included. Approximately 10 percent of the descriptions did not contain an accurate description of either arrow and were therefore completely excluded from the analysis.

Gesture Rate by Condition

The words and gestures produced during accurate portions of each description were used to calculate a gesture rate per 100 words for each pattern. These rates were then analyzed in a mixed linear regression with condition (Rotation vs. Non-rotation) as a fixed effect and participant and pattern as random effects. Mixed linear regression is a more powerful analysis technique than analysis of variance, particularly when there is large individual variation in the dependent variable (see Baayen, Davidson, & Bates, 2008). This model was a better predictor of gesture rates than a model that included only the random effects, $\chi^2(1) = 4.92$, $p = .026$. Speakers produced more gestures per 100 words when they described rotating a pattern than when they described the end state of a rotation, $\beta = 1.09$ ($SE = 0.49$). See Figure 2.

Discussion

The present study suggests that speakers gesture at higher rates when they imagine problem elements in motion than when they imagine them in their static end state. This finding is compatible with previous reports that speakers gesture when solving spatial problems (e.g., Chu & Kita, 2008; Hegarty et al., 2005); however, this is the first experiment to have included a control condition in which participants described the same images involved in the problem solution without imagining them in motion. The inclusion of such a control condition allows a test of the claim put forward under the GSA framework (Hostetter & Alibali, 2008) that gestures are particularly likely to occur when speakers are simulating motion.

These findings suggest that gestures can be thought of as stemming from simulated actions in the mind of the speaker. The term “simulation”, as used by Hostetter and Alibali (2008) and as we use it here, refers to the off-line reenactment of neural experiences that occurred during an

initial event. This use of the term aligns with the usage of many other investigators, including Barsalou (2008, 2009), Niedenthal et al. (2010), and Hurley (2008). However, recently, a more specific meaning of the term “simulation” has been proposed by Willems, Toni, Hagoort and Casasanto (2010), on the basis of neural evidence that simulation and mental imagery are doubly dissociable constructs in the brain. According to Willems et al. (2010), *simulation* is a fast, effortless process that occurs during language comprehension. Relying primarily on areas in the premotor cortex, simulation is hypothesized to be predictive, in the sense that its computational goal is to predict the bodily action necessary to interact with the environment. In contrast, Willems et al. describe *imagery* as a slow, effortful process that engages primary motor cortex. The computational goal of imagery is hypothesized as reflective, as a means of recreating physical experiences.

As we use it here, the term “simulation” encompasses both of the more specific processes distinguished (and given different names) by Willems et al. (2010). Simulation (in Willems et al.’s sense) and imagery serve a similar function in the GSA framework, in that both engender neural activation in areas that could hypothetically support gesture production (i.e., premotor or motor cortex). Importantly, we believe that either recreative or predictive activation has the potential to support gesture production. In the present paradigm, the motor imagery in the rotation condition was presumably largely predictive, in that speakers used it to predict the end result of rotating the arrows. In contrast, the motor imagery involved in Hostetter and Alibali (2010) was probably largely recreative, in that participants were using imagery to recreate the actions they had produced during their encoding of the patterns.

In future work, it would be interesting to directly compare the gestures that occur with predictive motor imagery with those that occur with recreative motor imagery in a controlled setting. For example, participants could manually rotate the arrows in one condition before describing the end state, while in another condition, they could mentally rotate the arrows as they are describing them. In such a design, the motor imagery involved in the mental rotation condition would be largely predictive while the motor imagery involved in the manual rotation condition would be largely recreative. The GSA framework does not make a prediction about which of these situations would result in more gesture, and instead predicts that both would result in more gesture than a baseline condition in which the end state of the rotation is being described.

These findings do not speak to the issue of whether gestures actually facilitate the process of solving spatial problems. This is a separate question from the question being addressed here; however, it is worth noting that previous research is somewhat mixed on the extent to which gestures not only *reflect* the spatial and motor processes involved in problem solving, but also *facilitate* those processes. For example, Hegarty et al. (2005) found that prohibiting speakers from gesturing did not interfere with

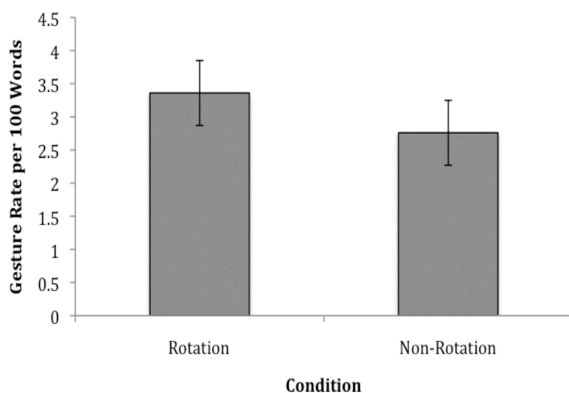


Figure 2. Average gesture rates per 100 words in the Rotation and Non-rotation conditions. Error bars represent the standard error of the condition effect in the regression.

individuals' ability to correctly solve mental animation problems. On the other hand, Chu and Kita (2011) found that speakers who are encouraged to gesture during a mental rotation task solve more problems correctly than speakers who are prohibited from gesturing. Regardless of whether gestures actually help speakers predict the outcome of a mental simulation, the present data suggest that gestures occur, at least in part, as the result of this mental simulation occurring in the mind of the speaker.

In addition to their possible role in problem solving, gestures have also been shown to facilitate communication about motor information (Hostetter, 2011). It is possible that some of the gestures observed in both conditions here were intended by the speakers to help an audience understand their descriptions. However, in both conditions, the participant did not know that he or she was being video taped and there was no other physical audience present who stood to benefit from the gestures. Thus, it is unlikely that very many of the gestures produced here were primarily for communicative reasons. Further, because the communicative demands were identical in the two conditions, it seems quite unlikely that differences in communicative demands are driving the difference observed here.

In conclusion, the present study has provided evidence that the gestures that occur with descriptions of mental animation problems are not produced simply in an effort to describe spatial aspects of those problems. Instead, the mental process of imagining problem elements in motion is particularly likely to give rise to gestures. In this way, simulated action supports gesture production.

Acknowledgments

We thank Elizabeth Farley for her assistance with data coding. We also thank Chelsea Baumgarten, Cierra Gillard, Keith Moreno, Andrea Potthoff, and Erin Sullivan for their comments on the manuscript.

References

- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating and thinking about spatial information. *Spatial Cognition & Computation*, 5, 307-331.
- Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production. *Journal of Memory and Language*, 44, 169-188.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Research Society*, 364, 1281-1289.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.
- Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, 137, 706-723.
- Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140, 102-116.
- 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.
- Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, 113, 98-104.
- Emmorey, K., & Casey, S. (2001). Gesture, thought, and spatial language. *Gesture*, 1, 35-50.
- Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5, 664-674.
- 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. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137, 297-315.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15, 495-514.
- Hostetter, A. B., & Alibali, M. W. (2010). Language, Gesture, Action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language*, 63, 245-257.
- Hostetter, A. B., & 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 representational gesture rates. *The Journal of Nonverbal Behavior*.
- Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences*, 31, 1-58.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7, 54-60.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) Model: Embodied Simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, 33, 417-480.

- Richter, W., Somorjai, R., Summers, R., Jarmasz, M., Memon, R. S., Gati, J. S., et al. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience*, 12, 310-320.
- Schwartz, D. L., & Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 116-136.
- Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience*, 22, 2387-2400.