WCER Working Paper No. 2012-3
March 2012

Gesture as Model Enactment: The Depictive and Causal Roles of Gesture in Mental Model Construction When Learning from Text

Mitchell J. Nathan
Department of Educational Psychology/
Wisconsin Center for Education Research
University of Wisconsin–Madison
mnathan@wisc.edu

Chelsea V. Johnson
Wisconsin Center for Education Research

Wisconsin Center for Education Research
School of Education • University of Wisconsin–Madison • http://www.wcer.wisc.edu/

WCER working papers are available on the Internet at http://www.wcer.wisc.edu/publications/workingPapers/index.php. Recommended citation:

> Nathan, M. J., & Johnson, C. V. (2012). *Gesture as model enactment: The depictive and causal roles of gesture in mental model construction when learning from text* (WCER Working Paper No. 2012-3). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Gesture as Model Enactment:
The Depictive and Causal Roles of Gesture
in Mental Model Construction When Learning from Text

Mitchell J. Nathan and **Chelsea V. Johnson**

The mind has a rich capacity for representing ideas and the state of the world, both real and imagined. As an example, consider the information conveyed by these sentences, from the work by Bransford, Barclay, and Franks (1972):

(1a). Three turtles rested on a floating log, and a fish swam beneath them.

(1b). Three turtles rested on a floating log, and a fish swam beneath it.

(2a). Three turtles rested beside a floating log, and a fish swam beneath them.

(2b). Three turtles rested beside a floating log, and a fish swam beneath it.

A purely propositional account of reading comprehension, like one that might be performed by a simple computer program, would not confuse (1a) with (1b), because the words *them* and *it* do not perfectly match. However, people who read (1a) are very likely to confuse it with (1b), because, when we read, we do not just represent the words; we elaborate, make inferences, and construct for ourselves mental imagery of the objects, actions and relations of the situation that is referred to. This rich internal representation does not distinguish (1a) from (1b) since the fish swims beneath them both. In contrast, people who read (2a) (the *beside* relation) almost never confuse it with (2b), since here the two situations clearly differ both spatially and propositionally.

Mental models (hereafter, MM; Gentner & Stevens, 1983; Johnson-Laird, 1980, 1983) are so named because they capture and internally represent the spatial, temporal, and causal relations implied by modeling the meaning of the events and relations of a text (van Dijk & Kintsch, 1983; Kintsch, 1998), and often automatically include unstated but probable information about the situation in order to make this internal representation more complete. MMs are often described as imagistic, holistic, and dynamic, and can include runnable, simulation-like accounts of a phenomenon (e.g., Clement, 1983).

In contrast, the *textbase* is described as a propositional, sequential, and static representation of meaning, which is developed directly from the information explicitly presented in a passage (Kintsch, Welsh, Schmalhofer, & Zimny, 1990; Fletcher & Chrysler, 1990; Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995). With our vast facilities for producing and understanding language, it is unclear why MMs are formed by the mind at all. Why isn't a propositional structure sufficient? And why, once MMs are formed, do they exhibit holistic and spatial qualities? It is as though when comprehending language we are "of two minds”: one that addresses propositional aspects of the emic world, and another that holistically captures its spatial and causal relations.

This propositional-spatial duality appears to be fundamental to language, and to thinking more generally. For example, while uttering propositional information speakers often *gesture* with their hands when they communicate to one another (Kendon, 2004; McNeill, 1992). Gestures are intriguing because, like MMs, the nature of the information conveyed is holistic, nonlinear, and spatial (Kendon, 2004); gestures can reveal or enact the spatial, temporal, and causal relations of a problem (Goldin-Meadow, 2003); they often serve to complement the information presented verbally (Goldin-Meadow, 2003; Nathan, 2008); and they, too, have been likened to simulations of mental actions (Hostetter & Alibali, 2008).

Progress in answering questions about the nature and potential interrelationship of gestures and MMs could advance scientific understanding of how we represent our knowledge of the world, and how we think, learn, and communicate what we know to others. Specifically, if gestures play a role in MM formation, this would provide further evidence of the body’s role in cognitive processes. Revealing this connection may contribute to theories of MM formation, model-based reasoning, and prevailing accounts of the role action plays in knowledge, communication, and learning. Practically speaking, such connections can have direct implications for methods of assessing knowledge, as when we test students on what they have learned when reading from text.

General Characteristics of Mental Models

Broadly construed, MMs are psychological constructs, created on the spot (Vosniadou & Brewer, 1992), which depict the relations among elements of some physical or conceptual system. Typically, these relations are spatial or causal (Zwaan & Radvansky, 1998), though MMs can also capture temporal, intentional, and perspectival relations (e.g., Goldvarg & Johnson-Laird, 2001; Schaeken, Johnson-Laird, & d’Ydewalle, 1996; Taylor & Tversky, 1992, 1996). Investigations of reasoning and problem solving, eye fixations, reaction times (Hegarty & Just, 1993; Rinck, Hahnel, Bower, & Glowalla, 1997; Taylor & Tversky, 1996), and neuroimaging (e.g., Knauff, 2009) show that MMs carry analogical information (e.g., relative proximity or cause and effect relations), while also depicting abstract relations among classes of entities, regardless of whether the information is acquired through reading, listening, or by observing a physical system in action. MMs also consolidate information that an individual has learned about a domain, enabling the individual to rapidly access that information when making judgments about new situations (Johnson-Laird, 1994; Vosniadou & Brewer, 1992). Some scholars (e.g., Barsalou, 1999) suggest that MMs should be considered for their role in providing relational and predictive information that may prepare people for pending situated action. Glenberg (1999) argues that readers’ sensibility judgments depend on their abilities to combine the perceived affordances invoked by the situation under consideration with their goals. Empirical studies also show that people are able to adjust their MMs when they encounter new information by accessing stored memory related to the information and then interpreting it based on this stored knowledge (Clement, 2008; de Vega, 1995; Glenberg, Meyer, & Lindem, 1987; Johnson-Laird, 1983; Singer, Radinsky, & Goldman, 2008).

Situation Models and Learning from Text

The role of MMs in meaning making is particularly relevant to studies of reading and learning from text. Reading is thought to produce a unitary structure that addresses the complex network of relations among ideas (episodic text memory; Kintsch, 1998). However, for purposes of modeling, predicting, and assessing reading comprehension and learning, it has proved valuable to consider analytically separable levels of text representation such as those identified by van Dijk and Kintsch (1983)—namely, the surface level, textbase, and *situation model* (SM). The surface level and textbase relate to memory for perceptual and propositional information, respectively, derived directly from the given text. The SM is an extratextual construct formed when integrating propositions within the text with the reader's prior knowledge. The SM may be peripheral or play no role at all in *memory* of a text (Kintsch, 1998). However, when the focus is on deep learning (Cutica & Bucciarelli, 2008) or reading with understanding, "learning from text . . . requires the formation of a situation model" (Kintsch, 1998, p. 295).

SMs incorporate ideas that must be generated because they are not directly stated by or retrievable from the text (Kintsch, 1993; Schmalhofer & Glavanov, 1986). There are several reasons for including SMs in theories of reading (Zwaan & Radvansky, 1998): SMs help account for the ways in which readers integrate information across propositions (e.g., Hess, Foss, & Carroll, 1995), across modalities, such as verbal and visual information (e.g., Baggett, 1979; Glenberg & Langston, 1992), and across multiple text sources (e.g., Wiley & Voss, 1999).

It must be acknowledged that there seldom are clear lines between learning and memory, so evaluations of a reader's textbase and SM are not cleanly separated. Investigators have developed methods, however, that focus to a greater degree on one or the other. Measures that go beyond the given text, such as questions directed at inferences or drawings of entities from a text and their spatial and functional relations, serve as valuable methods (Butcher, 2003; Butcher & Kintsch, 2004). For example, Butcher (2006) demonstrated that including appropriate illustrations enhanced readers' formation of SMs regarding the circulatory system, as measured by inference making and drawing quality. Readers of the nonillustrated text formed a weaker SM, even though the text itself was sufficient to support the reasoning needed for all of the posttest measures.

Another influence on memory and SM formation is the presence of gestures. Cutica and Bucciarelli (2008) presented participants with video-recorded stories both with and without accompanying gestures. Stories presented with gestures led to better recollection and more discourse-based inferences, though surface structure recognition was superior in the no-gesture condition. On this basis, Cutica and Bucciarelli concluded that individuals who listen to a discourse accompanied by gesture develop a richer SM, but a weaker verbatim representation of the narrative. While this advances our understanding of the role of gesture in SM formation, Cutica and Bucciarelli's design does not address whether learners' gestures influence the representations those learners create. Furthermore, in their study, gestures served primarily a discourse function, much like the presence or absence of illustrations in Butcher's (2006) study (and with similar results). Consequently, Cutica and Bucciarelli's work speaks to the design of multimedia materials but not necessarily to learners' cognitive strategies for enhancing comprehension and learning. Finally, their work examined learners' comprehension of spoken discourse (with or without gestures), which leaves open the question of whether the findings apply to learning from written text, which cannot incorporate gestures per se.

General Characteristics of Gestures

One of the earliest insights about gesture, other than its ubiquity, is that there are different types. Even in the face of broad cultural variation (e.g., Kita, 2009), gestures exhibit a universality that signals its natural place in communication (Gulberg, 1998; Kendon, 2004). Though several categorical schemes have been proposed, McNeill (1992) has offered a widely adopted one that considers both a gesture's form and its interrelationship with concurrent speech. *Iconic gestures* generally depict shape or motion and make reference by virtue of the similarity of the gesture form to the object shape or motion path. *Metaphoric gestures* can appear much like iconics yet convey abstract ideas when the form they depict is metaphorically related to the focal object. For example, invoking a large object with one's hands when speaking about a theory expresses intellectual importance though size and heft. These two gesture types have often been combined for practical purposes under the more general category of *representational gestures* (e.g., McNeill, 2000), since both refer to gestures that pictorially "bear a close formal relationship to the semantic content of speech" (McNeill, 1992, p. 12). *Pointing* or *deictic gestures* are used to indicate objects, locations, inscriptions, or events, and operate in an indexical manner. Pointing gestures can be invoked when what is being indexed is absent (or does not exist), permitting deictics to convey abstract concepts. Non-semantic beat gestures, which stress prosody or the structure of speech, and interactive gestures, such as managing turn taking, are commonly excluded in analyses of learning and cognition (e.g., Hostetter & Alibali, 2008) and will not be addressed here.

The Role of Gestures in Communication

During speech, gestures often serve interpersonal functions among interlocutors (Kendon, 1994; McNeill, 1985, 1992). Speakers will gesture more when listeners can see them (Alibali, Heath, & Myers, 2001). Gestures can depict spatial information directly (Kita & Özyürek, 2003; Krauss, 1998; Krauss, Chen, & Gottesman, 2000). Teachers use gestures to provide referential correspondences between familiar and new representations, and their gesture use increases when new ideas are introduced (Alibali & Nathan, 2007).

The Role of Gestures in Reasoning and Learning

One of the great insights from the scientific study of gestures is that they convey information about the speaker's mental state. Goldin-Meadow and her colleagues have shown in a variety of settings and activities that gesture reveals how children think about the tasks they perform, what they notice, and what they are ready to learn (e.g., Alibali & Goldin-Meadow, 1993; Church & Goldin-Meadow, 1986; Perry, Church, & Goldin-Meadow, 1988). Gesture may convey knowledge that is based more on motor simulation and not readily verbalizable, as when performing mechanical reasoning (Hegarty, 2004). Gesture can also exhibit the refinement of representation that problem-solvers go through. For example, when problem-solvers transition from simpler to more sophisticated solution representations, their gestures may provide a visual trace of the evolving nature of the solvers' MMs (e.g., Schwartz & Black, 1996).

In addition, gesture production is implicated as a factor in reasoning, learning, and development. Children in one study produced greater numbers of iconic (but not pointing) gestures when asked to *reason* about conservation of objects in a Piagetian task than when asked to *describe* how the objects appeared, presumably because of the greater difficulty and more substantive reasoning involved in the conservation task (Alibali, Kita, & Young, 2000). Gestures also appear to reduce the cognitive demands of some tasks (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner, Nusbaum, & Goldin-Meadow, 2004), facilitating visual coding of stimuli (Butterworth & Hadar, 1989), and providing alternative ways of representing information (Goldin-Meadow, 2003). Since gestures also provide unique sources of perceptuo-motor information, it has been proposed that the act of gesturing can itself be a source of feedback that alters thinking and thereby directly influences the learning process (Goldin-Meadow & Beilock, 2010).

One avenue of study for understanding the role of gestures in reasoning is to inhibit or otherwise occupy the ability to gesture. When speakers' gestures are inhibited, they produce less imagery when they talk (Rimé, Schiaratura, Hupert, & Ghysselinckx, 1984) and fewer perceptually rich explanations (Alibali & Kita, 2010). Children prohibited from gesturing during a Piagetian conservation task were less likely to refer to perceptually present information and more likely to refer to non-present information than when gesture was permitted (Alibali & Kita, 2010). Inhibition itself seems to generally load the cognitive system, impairing performance on a secondary task even when that task is not overtly motoric or spatial (Goldin-Meadow et al., 2001). Yet gesture inhibition also seems to selectively impair some forms of reasoning, such as causal relations within a mechanical system (Hegarty, Mayer, Kriz, and Keehner, 2005).

An Embodied Cognition Account of Gesture Production and Learning

The Theory of Gesture as Simulated Action (GSA; Hostetter & Alibali, 2008) provides an embodied cognition account of the causes of speech-accompanied gestures. Neural areas employed by action and perception that are essential for functioning in the world (Gibson, 1979) are also activated during *simulated* action and mental (i.e., offline) processing of language, imagery, and planning. As suggested by Figure 1 (next page), gestures arise when pre-motor activation, formed in response to motor or perceptual imagery, is activated beyond a speaker's current gesture-inhibition threshold. The threshold inhibits motor activation of the body during speech but can be overcome by several factors, such as the current task demands (e.g., strength of motor activation when processing spatial imagery), individual differences (e.g., level of spatial skills), and situational considerations (e.g., communicative contexts).

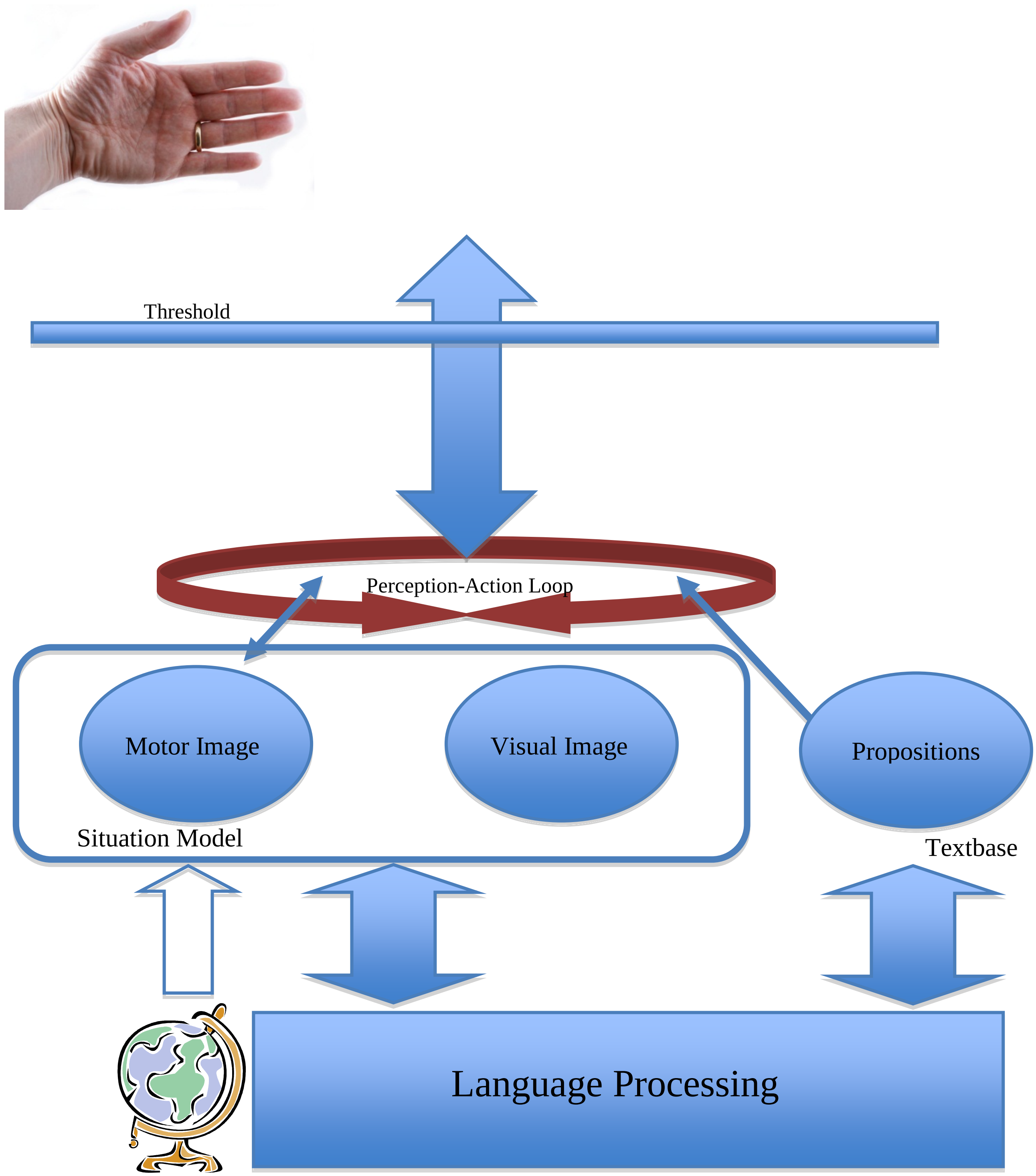


***Figure 1*. Modified version of GSA model (Hostetter & Alibali, 2008) that distinguishes between the textbase and the situation model constructs when learning from text.**

Hypotheses

In this investigation, we propose three hypotheses that extend our current understanding of the relationship between gesture production and SM formation when learning from text.

Hypothesis 1: gesture production is associated with model-based reasoning. Operationally, we predict greater gesture production when readers respond to inference-based test items than other test items, because inference items draw more directly on readers' SMs.

Hypothesis 2: gesture production is responsive to model-based information. Operationally, we expect to see that texts that contribute to a weaker SM (as when illustrations accompanying the text are removed; Butcher, 2006) will yield significantly lower rates of gesture production.

Hypothesis 3: manipulation of gesture production will influence the quality of readers' SMs. Operationally, we predict that readers whose gesture production is inhibited will form weaker SMs than those free to gesture, as indicated by a significantly lower performance on inference items. Since gesture inhibition can produce a general drop in cognitive processing, even for tasks that are not spatial (Hostetter & Alibali, 2008; Goldin-Meadow et al., 2001; Wagner et al., 2004), we can expect to see this drop in performance across all of the posttest items with gesture inhibition. However, due to the combined effect of general attentional load and a targeted loss of model-based resources, we expect the greatest performance drop on inference items for participants whose gestures are inhibited.

General Method

Several common aspects of the empirical method are used across the three experiments.

Materials

Using a set of tutorial and test activities about the human heart and circulatory system developed by Wolfe et al. (1998) and Butcher (2006), participants in each experiment completed a two-part pretest, followed by a computer-based tutorial and two-part posttest. The pre- and posttests each included drawing and verbal tasks.

Drawing tasks. The interviewer handed participants a piece of paper and a marker and instructed them to "use this sheet of paper and marker to draw a picture of the heart and its relationship to the circulatory system as best you can. After you're done drawing, I will ask you to hold your drawing in front of the camera and explain it to me." Participants were not given a time limit to create the drawing or to explain it. After participants finished explaining the drawing, the interviewer placed it facedown and out of sight so participants could not refer to it later. Participants completed identical drawing tasks before and after the computer-based tutorial.

Verbal tasks. While the pre- and posttest questions were adopted from Butcher (2006; also see Wolfe et al., 1998), the GSA framework (Hostetter & Alibali, 2008) notes the importance of engaging the speech production system to facilitate gesture production. Consequently, in the current study the task demands are designed so that participants *verbally* respond to the pre- and posttest items. (In contrast, Butcher [2006] had participants type their answers.)

The pretest included “general knowledge” questions designed to assess participants’ current understanding of the human heart and circulatory system (i.e., “What is a capillary?”). The posttest included these general knowledge questions and textbase and inference questions. Textbase questions asked participants to recall information presented in the computer-based tutorial (i.e., “Where does blood entering the right ventricle come from?”). Inference questions required that participants extrapolate information learned from the text to explain the effects of hypothetical situations related to the heart and circulatory system (i.e., “What would be the consequences of a large hole in the septum that separates the left and right sides of the heart?”). A total of 31 questions were asked: 18 general knowledge questions, 8 textbase questions, and 5 inference questions.

Tutorial. After the pretest, participants completed a tutorial focusing on the circulatory system developed and used in prior studies of SM development (Butcher, 2006). The tutorial was presented as a computer-based text comprising 43 web pages. Each page had a small amount of text about the heart and circulatory system, and 32 pages featured text accompanied by a relevant color illustration (see Appendix A). Participants advanced through the tutorial by using a mouse to click to the next page, but could not move backwards. Participants were not given a time limit.

Procedure

In each experiment, participants completed one-on-one interviews conducted in a private office furnished with two desks, two chairs, and a video camera visible to participants. Participants were videotaped as they completed an initial drawing activity, a pretest of general knowledge of the circulatory system, the computer-based tutorial, and several posttest activities: a drawing activity, a test of general knowledge, a test of textbase questions that could be answered directly from the tutorial, and a test of inference-based questions that went beyond the text of the tutorial and required model-based processing. Participants sat in front of the video camera at a desk with a computer and were able to turn to face the interviewer throughout the session.

The tutorial was presented on an iMac computer using the Safari web browser operating offline. Participants viewed the text and images on the screen and were directed to select a button on the screen that would advance to the next page of text or text plus illustration. The location of the button shifted throughout the task. The tutorial typically took approximately 30 minutes to complete. Participants were not permitted to take notes.

During the pre- and posttest activities, participants were asked to explain their drawing and answer questions to the best of their abilities. They were encouraged to guess if they were unsure of an answer, and they were permitted to respond, “I don’t know.” The interviewer did not indicate if participants answered correctly or incorrectly. Participants were told that the posttest included familiar as well as new questions. Items in both the pre- and posttest were always asked in the same order. After completing the final posttest items, participants were thanked for their participation and debriefed on the purposes of the study.

Coding and Scoring of Participants' Data

Drawings. Participants' drawings at both pretest and posttest were scored according to a 6-point rubric established by Chi, de Leeuw, Chiu, and LaVancher (1994). Both the drawing and the participant's explanation of it, including the verbal report and any accompanying gestures, were considered when scoring. See Appendix B for an example of the rubric applied to participants' drawings.

Verbal responses. The interview tasks were transcribed and coded for speech and gesture using Transana video analysis software (Woods & Fassnacht, 2009). Verbal utterances were unitized at sentence and clausal boundaries (Kintsch, 1998), and codes were assigned for both speech and gesture.

***Speech codes*.** Speech codes identified segments of speech that fell into a certain category related to SM formation. Every unitized segment of participant speech was assigned at least one code from the following set.

- *Original speech.* Participants' language not derived from the tutorial, e.g., "Uh, it's like the really thin blood vessels that are like the smallest ones so that there can be exchange of like gasses and stuff."
- *Paraphrase.* Restatements of material presented in the tutorial, e.g., "A ventricle is one of the bottom chambers where blood is pumped out of the heart."
- *Textbase.* Statements taken directly from the tutorial text, e.g., "A capillary is a microscopic vein or little veins."
- *Metacognitive.* Statements that demonstrated participants' reflection on their knowledge, e.g., "No. Yes. I'm gonna go with that, although I might have 'em confused."
- *I don't know.* Statements in which the participant did not attempt answer the question posed, e.g., "I have no idea."
- *Unscripted question.* Applied when participants asked a question.

Gesture codes. Each segment of participant speech was assigned codes from the following set.

- *Gesture utterance.* Segments of speech that included gesture.
- *Non-gesture utterance*. Segments of speech without gesture.
- *Point*. Segments of speech that included deictic gestures.
- *Beat*. Segments of speech that included gestures that did not display semantic content but occurred in rhythm with the participants' speech.

- *Representational*. Segments of speech that included gestures that displayed semantic content, including iconic and metaphoric gestures.

Gesture rate, an accepted measure of gesture production (Alibali & Kita, 2010; Hostetter & Alibali, 2010), was calculated for each task in the interview (predraw, pretest of general knowledge, postdraw, posttest of general knowledge, posttest of textbase material, and posttest of inference material). Gesture rate was calculated as the number of gestures produced per 100 words spoken. Only pointing and representational gestures were used. Speech that was coded as "I don't know" was excluded from the word counts because this speech does not rely on readers' knowledge from the tutorials.

Participants' responses in the pre- and posttests were scored according to a rubric established by Butcher (2006). The pre- and posttests of general knowledge included 18 questions worth a total of 38 points; the posttest textbase portion included 8 questions worth a total of 17 points; the posttest inference portion included 5 questions worth a total of 11 points.

Experiment 1

The goal of Experiment 1 was to examine the relationship between gesture use and participants' model-based reasoning when learning from a text. Participants read an illustrated tutorial about the human heart and circulatory system (see Appendix A) and were asked a series of pre- and posttest questions about the topic. *Textbase questions* were taken directly from information stated in the tutorial text. *Inference-based questions* required readers to go beyond information explicitly stated and draw upon knowledge derived from their SMs. Here we test Hypothesis 1, which states that rate of gesture production will be highest when readers' responses draw most heavily from their SMs. We therefore expect to see higher gesture rates when responding to inference-making items as compared to textbase or general-knowledge items. Furthermore, because spatial demands and task difficulty could also account for elevated gesture production, we examine whether either of these factors are predictive of the data.

Method

Participants. Twenty-two participants completed the one-on-one interviews and heart text. Participants were students over the age of 18 recruited from undergraduate Educational Psychology courses at the University of Wisconsin–Madison. They were compensated with extra credit in that course upon completion of the experiment.

Materials and procedure. Sessions were run individually and videotaped in their entirety. Each person completed the same series of pre- and posttest activities described in the General Method section. The materials were presented in the following order: predraw activity, pretest of general knowledge, computer-based tutorial, postdraw activity, posttest of general knowledge, posttest of textbase materials, and posttest of inference-based materials. Each session lasted approximately 40 minutes.

Results and Discussion

Task performance. Participants took an average of 11 minutes and 47 seconds to complete the heart text, with times ranging from 5 minutes and 10 seconds to 16 minutes and 12 seconds. General knowledge questions and the drawing tasks were the only test items that support direct comparison before and after reading. As noted, drawing scores were determined by a 6-point rubric developed by Chi and colleagues (1994). Table 1 shows the mean scores and standard deviations for the drawing tasks. Scores on the drawing tasks were used as measures of participants' initial and final models of the circulatory system as a whole, capturing aspects of its interrelations and dynamic qualities. The predraw activity showed a mean score of 2.68 (SD = 1.32; range 1–5 out of 6 maximum). The mean score on the postdraw activity was 4.14 (SD = 1.28, range 1–6), and was a significant gain over predraw performance, $t(19) = 5.97$, $p < .001$, Cohen's $d = 1.12$.

***Table 1.* Mean scores on drawing tasks in Experiment 1 (*n* = 22).**

Task	Mean score	SD
Predraw	2.68	1.32
Postdraw	4.14	1.28

Results from the general knowledge pretest (Table 2) showed a mean of 35.3% correct responses (SD = 25.90; range 3.85%–79.50%). Participants' baseline general knowledge was, on average, significantly greater than zero, $t(19) = 5.93$, $p < .01$. The general knowledge posttest showed a mean of 54.50% (SD = 15.15). Comparing general knowledge pre- and posttest performance yielded evidence of significant gains, $t(19) = 6.17$, $p < .001$, Cohen's $d = 1.08$.

Posttest scores for textbase items ranged from 38.50% to 100% correct, with a mean of 71.90%. In contrast, inference scores ranged from 13.60% to 81.80% correct, with a mean of 45.90%. Performance was significantly lower on inference than textbase questions, $t(19) = 6.33$, $p < .001$, Cohen's d = 1.42, as was expected from prior research (e.g., Butcher, 2006).

***Table 2.* Mean scores on verbal tasks in Experiment 1 (*n* = 22).**

	Task	% correct	SD
Pretest			
	General knowledge	35.30	25.90
Posttest			
	General knowledge	54.50	15.15
	Textbase	71.90	20.78
	Inference	45.90	16.82

As the data in Tables 1 and 2 suggest, participants generally increased their performance and exhibited significant gains in both verbal and drawing measures. Inference making showed lower scores than answering the textbase items that drew directly from the reading passage.

Gesture rate. Table 3 shows the mean gesture rates and standard deviations for all of the verbal tasks included in Experiment 1. Gesture rates during the pre- and posttest of general knowledge were comparable ($F < 1$). The main contrast of interest for testing Hypothesis 1 is between the gesture rate of inference-based items and the other test items. An analysis of variance showed that gesture rates differed significantly among the verbal interview tasks, F (5, 120) = 9.46, $p < .001$. Post-hoc analyses using Fisher's Least Square Difference method indicated that the average gesture rate was significantly higher during the inference questions than the pretest of general knowledge questions ($p < .001$), the general knowledge posttest ($p < .001$), and the posttest textbase questions ($p < .05$). As a further test, we performed a Wilcoxon Signed Rank test to examine the pattern at the level of each participant and found that gesture rate was higher during inferences than textbase responses on a person-by-person basis (Z = -2.57, $p < .01$).

Table 3. **Mean gesture rates per hundred spoken words for verbal tasks in Experiment 1.**

Task	Gesture rate	SD
Pretest		
General knowledge	7.97	4.43
Posttest		
General knowledge	7.73	5.45
Textbase	9.91	5.45
Inference	13.46	6.95

As predicted, gesture rates were significantly greater when responding to the inference-based test questions than textbase or general knowledge questions when learning from text. It is unlikely that gesture production is epiphenomenal in this context since it increases systematically during certain task demands, and this held in the aggregate and on a person-by-person basis. These results are in line with the predictions made in Hypothesis 1.

One interpretation is that the gesture production system is more actively engaged during model-based reasoning because of SM information is highly action based. However, two alternative interpretations must also be addressed. One is that spatial demands from engaging spatial reasoning systems (Hostetter & Alibali, 2008) or lexical access of spatial terms (Krauss, Chen, & Gottesman, 2000) drive the differences between test items. All of the textbase and inference items were coded as predominantly spatial (cf. Butcher, 2006), but among general knowledge items more than one-third (9/25) were spatial and nearly two-thirds (16/25) non-spatial. We then examined whether gesture rates differed significantly between spatial and non-spatial general knowledge items. An analysis comparing gesture rates of spatial and non-spatial general knowledge items showed a marginal but non-significant effect on gesture rates, $t(19) = 2.05$, $p = .054$, Cohen's $d = 0.003$. This indicates that the spatial quality of test items alone is not likely to explain the observed difference in gesture rates.

A second alternative interpretation, that the relative difficulty of the test items determines gesture rate, is also unsupported (cf. Sassenberg & van der Meer, 2010). Comparing patterns of performance (Table 2) and gesture production (Table 3) show that textbase items have the highest levels of performance but are intermediate in their accompanying gesture rate, while general knowledge items exhibit the lowest gesture rate but are intermediate in difficulty. Thus, difficulty by itself is also an unlikely explanation of the patterns of gesture production that were observed.

The results of Experiment 1 indicate a reliable relationship between gesture production and model-based reasoning that is not simply dependent on their spatial content or level of difficulty. Yet the specific nature of the relationship remains unclear. If gesture rate is reliably affected by manipulating the quality of the SMs produced by readers, this would provide further evidence for the relationship between gesture production and model-based reasoning.

Experiment 2

Experiment 1 showed that responses to inference-based questions—those items that depend most strongly on readers' SMs—elicit significantly higher gesture rates than textbase or general knowledge questions. According to Hostetter & Alibali's (2008) GSA framework, this arises because of demands placed on the motor system through simulated action. Further analyses revealed that this was not simply due to the difficulty or spatial nature of the information in the test items, thus providing some support for Hypothesis 1 and the broader investigation of the role of gesture in SM formation.

Experiment 2 aims to sharpen these claims in three ways. First, Experiment 1 was correlational, therefore causal claims are not warranted. If the hypothesized relationship holds, then manipulation of the quality of readers' SMs in Experiment 2 will produce expected changes in gesture rates. Previous research by Butcher (2006) demonstrated that including appropriate illustrations contributes to the development of SMs for the circulatory system. Specifically, Butcher found that readers of the non-illustrated text formed a more impoverished SM and exhibited lower performance on inference-making items. This held even though the text by itself was sufficient to support the reasoning needed for all of the posttest questions. Because in-text illustrations have been shown to provide readers with a richer base for SM formation, participants who read a non-illustrated version of the heart tutorial are expected to develop a weaker SM, and consequently exhibit measurably lower gesture rates and lower test scores on inference making. Second, it is still likely that variability of gesture rate across test items is driven by the different spatial demands involved in making inferences in the domain of the circulatory system. By excluding the illustrations from the original text, we can compare between-subject gesture rates using high (illustrated) versus low (non-illustrated) spatial stimuli.

Finally, Experiment 2 provides an occasion to replicate the finding that readers gesture more during inference making (Hypothesis 1).

Method

Participants. Fifty-four undergraduate students, all over 18 years of age, were recruited from Educational Psychology courses at the University of Wisconsin–Madison. They were compensated with extra credit in that course upon completion of the experiment. Participants were randomly assigned to read either an illustrated or a non-illustrated version of the circulatory text based on order of arrival to the laboratory. Each session was conducted individually and lasted approximately 40 minutes. Because six participants engaged in activities during testing that interfered with their gesture production they were excluded, resulting in data from 48 participants for the final analyses.

Materials and procedure. Participants completed the same series of pre- and posttest activities in the same order as in Experiment 1. During the tutorial, however, 24 students saw a non-illustrated version of the heart text while the remaining 24 saw the illustrated version used in Experiment 1. The non-illustrated tutorial used the same text and was presented in the same manner as the illustrated tutorial; the only difference being the presence or lack of illustrations. Appendix A shows a page from the tutorial in both its illustrated and non-illustrated forms.

Participants' drawings, speech, and gestures were coded using the system introduced earlier. As before, speakers' gesture rate per 100 words uttered was calculated for each task.

Results

Test performance. Scores on the drawing task (Table 4) improved significantly from pre- to posttest overall, $F(3,52) = 78.73$, $p < .000$, MSE = 1.04, $\eta_p^2 = 0.60$, and in both conditions, $p < .01$. Group differences at pretest were not significant, and there was no significant text x time interaction. Postdraw scores also showed no difference, $F < 1$.

Performance on general knowledge questions showed no pretest differences (Table 5), and significant improvement from pre- to posttest, $F(3,52) = 145.44$, MSE = 96.19, $p =.000$, $\eta_p^2 = 0.74$, with no effect of text type ($F < 1$) and no text x time interaction ($F < 1$). Together, these results suggest that, overall, students improved their understanding of the circulatory system after reading either tutorial, with no apparent advantages on these measures for including illustrations.

***Table 4*. Mean drawing scores in Experiment 2 for illustrated and non-illustrated texts.**

	Illustrated tutorial		Non-illustrated tutorial	
Task	**Mean score**	**SD**	**Mean score**	**SD**
Predraw	2.89	1.42	3.07	1.03
Postdraw	4.70	1.51	4.74	1.29

As the main focus of this experiment, we expected to see higher inference scores for readers of the illustrated text, following results reported by Butcher (2006). Contrary to expectations, however, scores on the inference questions for the illustrated (*M* = 47.14%, *SD* = 24.96, ranging from 0% to 90%) and non-illustrated texts (*M* = 41.69%, *SD* = 18.67, ranging from 18% to 81%) were comparable between conditions, $F < 1$ (Table 5). As a further check, participants' scores on textbase questions in the illustrated (*M* = 69.28%, *SD* = 24.77, ranging from 17% to 100% correct) and non-illustrated (*M* = 68.08%, *SD* = 24.91, ranging from 6% to 100%) conditions did not reliably differ, $F < 1$. As in Experiment 1, we found the expected main effect of test items, with performance on textbase items significantly higher than inference items, $F(1,52) = 78.91$, $p =. 000$, MSE = 195.54, $\eta_p^2 = 0.6$.

***Table 5*. Mean scores on verbal tasks in Experiment 2 for illustrated and non-illustrated texts.**

	Illustrated tutorial		Non-illustrated tutorial	
Task	**% correct**	**SD**	**% correct**	**SD**
Pretest				
General Knowledge	29.63	24.75	31.29	23.35
Posttest				
General Knowledge	54.14	21.33	52.73	17.66
Textbase	69.28	24.77	68.08	24.91
Inference	47.14	24.96	41.69	18.67

It appears that those who read the non-illustrated text form a SM of the human circulatory system at a level comparable to that formed by readers of the illustrated text, in contrast with earlier findings (Butcher, 2006). We explore this further in the Discussion section.

Gesture rates. Table 6 shows the gesture rates for each verbal task. Participants in the non-illustrated condition gestured more frequently during the pretest of general knowledge than did participants in the illustrated condition, but this difference did not reach significance (F < 1). Gesture rates for all participants, regardless of text condition, were significantly higher during inference than textbase questions, $F(1,52) = 10.74$, $p = .002$, MSE = 63.48, $\eta_p^2 = 0.17$, replicating the effect found in Experiment 1, and providing additional support for Hypothesis 1.

***Table 6*. Mean gesture rates for verbal tasks in Experiment 2 for illustrated and non-illustrated tutorials.**

	Illustrated tutorial		Non-illustrated tutorial	
Task	**Gesture rate***	**SD**	**Gesture rate***	**SD**
Pretest				
General knowledge	2.98	2.68	4.98	4.62
Posttest				
General knowledge	4.70	3.09	5.35	5.26
Textbase	6.21	3.92	6.24	4.27
Inference	7.80	5.82	9.83	8.94

* Gesture rate for each participant is calculated as the number of gestures observed per 100 words uttered.

What of gesture production during inference making for readers of the non-illustrated text? Recall that the non-illustrated text was expected to support a less well-formulated SM, and therefore exhibit lower gesture rates. While the drawing and inference test performance data reveal comparable levels of SM development for members of both treatment groups, the gesture data tell a different story. Planned post-hoc Bonferroni comparisons ($\alpha < .05$) show that participants who read the non-illustrated text exhibited *significantly higher* gesture rates during inference questions than those in the illustrated group, $p < .05$ (Figure 2). This surprising finding is used to inform our discussion and the design of Experiment 3.

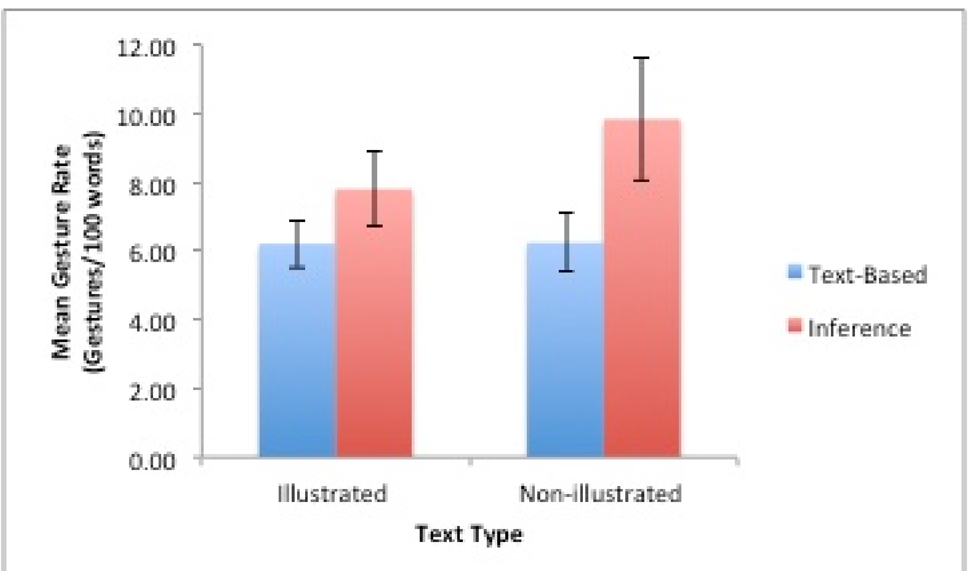


***Figure 2*. Experiment 2 mean gesture rates during posttest on textbase and inference-based questions as a function of treatment. (Error bars indicate standard errors.)**

Discussion

Findings form Experiment 2 support Hypothesis 1 but are contrary to Hypothesis 2 and previous empirical findings on learning from text (Butcher, 2006) showing that manipulation of the presence of helpful illustrations with the text would lead to corresponding differences in posttest inference scores. Specifically, those in the non-illustrated condition showed comparable scores on model-based inference questions and drawing tasks despite reading from a text devoid of images that had previously been shown to improve the quality of readers' SMs. Yet, along with comparable levels of performance we found—also unexpectedly—that the gesture rates of readers of the non-illustrated text were significantly higher when responding to the inference-based items. Although both the performance and gesture production results are surprising, they point to a common explanation: With less direct support for SM formation, a typical reader in the non-illustrated condition might use gesture production as a compensatory mechanism for constructing a more complete SM, one that reaches the level attained with the inclusion of appropriate figures. Gestures, then, are implicated as a body-based resource in SM construction. When other resources that aid SM construction (e.g., illustrations) are not available, participants use gestures adaptively in SM construction.

These findings also speak to the issue of differential spatial processing. Theories of gesture production that argue exclusively for the spatial role of gestures (e.g., Krauss, 1998) would presumably predict higher gesture rates for the illustrated treatment since presence of more spatial information invites greater spatial processing. In contrast, we found that the non-illustrated condition led to more gesturing. Prior research found increased gesture production in the absence of pictures (Wesp, Hesse, Keutmann, & Wheaton, 2001); yet, the pictures had been seen previously, and gestures served to reinstate that spatial information from memory.

This raises the question of how spatial reasoning plays into gesture production when other process demands are being made. Readers in the non-illustrated condition were exposed to less spatial information: they had fewer spatial objects to use for pointing and tracing, and they never saw the illustrations previously. While spatial demands certainly influence gesture production, the current findings make it clear that there are other influential factors involved. These data suggest that model-based reasoning in the form of inference making is itself a factor influencing gesture production.

One puzzling finding is that these results do not substantiate Butcher's earlier finding that the absence of illustrations leads to less inference making using the same reading and assessment materials. Yet there are other differences in the research methods worth noting. In Butcher's original (2006) study, participants responded to the verbal tests by writing or typing on a computer, and did so in non-communicative settings. As previously noted, the communicative setting and overt speech production are important preconditions to rich gesture production, both for its social engagement and the activation of speech motor programs. Furthermore, because of the written response methods used in Butcher's study, her participants also had their hands engaged while producing their answers. Participants' in Butcher's non-illustrated condition could not draw on gestures to contribute to SM formation in the absence of illustrations. By preoccupying one's hands, the formation of participant's SMs may be inadvertently affected. In contrast, participants in the current study were free to gesture during testing. In other words, the manipulation of gesture production might lead to variation in measures of model formation. It is this topic that we turn to in the final experiment.

Experiment 3

Experiments 1 and 2 demonstrated a relationship between gesture use and readers' reporting of model-based information when learning from text. Experiment 1 showed that gestures are invoked to a greater degree when readers' respond to inference-based questions than general knowledge or textbase questions during the posttest. Results from Experiment 2 suggest that gestures might serve as a resource during inference making, contributing to SM formation and compensating for those situations where model information (in the form of useful illustrations) is scarce. In the current experiment we investigate the hypothesis that gestures are causally involved in model creation by examining how manipulation of gesture production affects the quality of readers' SMs. When gestures are restricted, speakers produce less imagistic speech (Rimé, Schiaratura, Hupert, & Ghysselinckx, 1984) and fewer perceptually rich explanations (Alibali & Kita, 2010). If, as we predict, gesture production influences SM construction, then participants who are allowed to gesture freely should outperform those who experience gesture restrictions. Furthermore, we introduce two forms of gesture restriction. Simple tapping merely interferes with gesture production. But a more demanding spatial tapping condition requires that participants continually plan, execute, and monitor new patterns of motor movement. SM differences between the simple and spatial tapping condition would directly implicate the action control system above and beyond mere gesture restriction.

Method

Participants and materials. Seventy-nine undergraduate students were recruited for individual sessions that lasted approximately 50 minutes. Participants completed the same series of pre- and posttest activities (though the order differed slightly; see below), and read a modified version of the illustrated tutorial on the human circulatory system used in Experiment 1.

Procedure. Each participant was assigned to one of three experimental conditions based on order of arrival: Spatial tapping, simple tapping, and the gesture (control) condition. Regardless of condition, study tasks were administered in the following order: a measure of spatial reasoning, predraw task, pretest of general knowledge, computer-based tutorial (illustrated), posttest of general knowledge, posttest of textbase questions, posttest of inference-based questions, and finally the postdraw task.

Measure of spatial ability. Because some participants would be engaging in a secondary spatial task during the posttest activities, a measure of spatial reasoning was included to better predict performance on the verbal tasks and gesture production (Hostetter & Alibali, 2007). The paper folding test (Ekstrom, French, Harman, & Dermen, 1976), a 2-page worksheet designed to evaluate spatial ability, was included as a covariate measure and administered at the beginning of the interview session. Each page contained 10 multiple-choice problems. Participants were asked to imagine what a given piece of paper would look like if it were folded, punctured with a single hole through all of the folded layers, and unfolded. Participants were presented with each page one at a time and were given 3 minutes to complete each worksheet. If participants finished the first page before time elapsed, they immediately moved to the second page. If they had not completed a worksheet within the allotted 3 minutes, they were not allowed extra time to complete it.

Spatial tapping condition. Participants assigned to the spatial tapping condition (n = 26) were free to gesture during the pretest tasks. While reading the tutorial and responding to the posttest tasks, however, participants were instructed to use one hand, of their own choosing, to tap out a particular spatial pattern on a sheet of paper that displayed 4 numerals arranged in a 2 x 2 array, derived from a protocol established by Hegarty, Mayer, Kriz, and Keehner (2005; see Appendix C for an example of the tapping stimulus). Participants of the spatial tapping condition were instructed to match their tapping to the rhythm of 100 beats per minute established by the blinking light of a silent metronome situated in full view of each participant.

Though an initial tapping pattern was presented at the beginning of the tutorial (e.g., 2 3 4 1), the pattern changed throughout the tutorial and posttest to prevent participants from automating to any one spatial arrangement. In order to present new patterns throughout the study, the original illustrated tutorial was modified to include six extra screens, spaced ten screens apart, which cued participants to shift to the new tapping pattern. Each modified screen showed only the four digits of the new pattern presented horizontally on the screen along with a button at the bottom that took them to the next page of the tutorial.

During the oral posttest, the interviewer presented participants with a 5” x 8” card that showed a new tapping pattern. A different card was introduced after every five questions. The patterns of numbers introduced during the tutorial and posttest were not the same. No specific direction regarding gesture use was given to the participants in the spatial tapping condition, but it was expected that the constraints of the tapping task would greatly reduce their gesture use.

Simple tapping condition. A second tapping condition (n = 26) was used to control for the specific engagement of the hand, but without the more demanding requirement to follow a changing tapping pattern. As with those in the spatial tapping condition, participants in the simple tapping condition were free to gesture during the pretest portion of the interview. During the tutorial and posttest, however, participants in the simple tapping condition were instructed to use either hand to tap the number “2” on the 2 x 2 array of numbers (see Appendix C). Here, manual movement is restricted, but without greater demands on the action control system that mediate action planning and monitoring. As in the spatial tapping condition, participants were instructed to tap along with the 100 beats per minute rhythm set by a metronome with a blinking light. Although the tutorial contained the additional screens used to direct those in the spatial tapping condition to change their tapping pattern, participants in the simple tapping condition were told to continue tapping the assigned number, and click through to the next tutorial screen.

Participants also saw the numbered cards during the posttest interview, but were instructed to maintain their original tapping behavior. Those in the simple tapping group were not instructed to change their gesture use in any way.

Gesture condition. Participants in the gesture condition (n = 27) served as a control group. They were exposed to the same metronome and the same numeric patterns during the tutorial and posttest but were instructed to continue with the assigned task. They were free to gesture during all of the testing and reading activities.

Coding and Scoring

The paper folding test was scored according to its own rubric (Ekstrom et al., 1976). Responses to the pre- and posttest questions were scored according to the rubrics used in Experiment 1. Speech and gesture were coded according to the criteria used in the previous experiments. An additional gesture code ("tapping interference") was introduced to account for interruptions during the spatial tapping condition. The code was applied to a segment of speech if a participant stopped tapping to gesture. Interference was quite low in the spatial tapping condition (only 3 of 26 spatial tapping participants), so it was excluded in the final analyses.

Results

Task performance. Despite the potential interference of the tapping tasks, performance improved from pre- to posttest during the drawing activity for all three conditions, $F(2, 77) = 757.69$, MSE = 3.15, $\eta_p^2 = .91$, $p < .001$. Planned post-hoc Bonferroni comparisons ($\alpha = .05$) showed that no group's drawings were significantly better than the others (Table 7).

***Table 7*. Drawing scores for each condition used in Experiment 3.**

	Gesture (control)	Simple tapping	Spatial tapping
Predraw	3.15	2.41	3.35
Postdraw	4.89	4.44	4.96

Figure 3 shows the mean scores for all verbal tests included in Experiment 3. Using participants' score on the paper folding test as a covariate did not yield significant results on test scores ($F < 1$), which suggests that participants with high spatial ability were not able to differentially overcome the demands of the tapping conditions. The covariate was removed and a three-way multivariate analysis of variance (MANOVA) was conducted with gesture condition as an independent variable and score on each of the verbal tasks as dependent variables. The results show a main effect of gesture condition on scores, $F(4,75) = 6.0$, $MSE = 253.1$, $p < .01$. Planned post-hoc Bonferroni comparisons ($\alpha = .05$) showed that the control group scored significantly higher than the spatial tapping group on textbase ($p < .05$) and inference questions, $p < .01$. Additionally, participants in the simple tapping group scored higher on inference questions than did participants in the spatial tapping group, $p < .05$ (see Table 8). These results are consistent with the hypothesis that restricting gesture production while learning from text impairs model-based reasoning. Furthermore, the advantage of the simple tapping group over the spatial tapping group shows that it is not merely engaging one's hands that matter; rather it is engagement of those mental resources that mediate action systems that impairs SM formation.

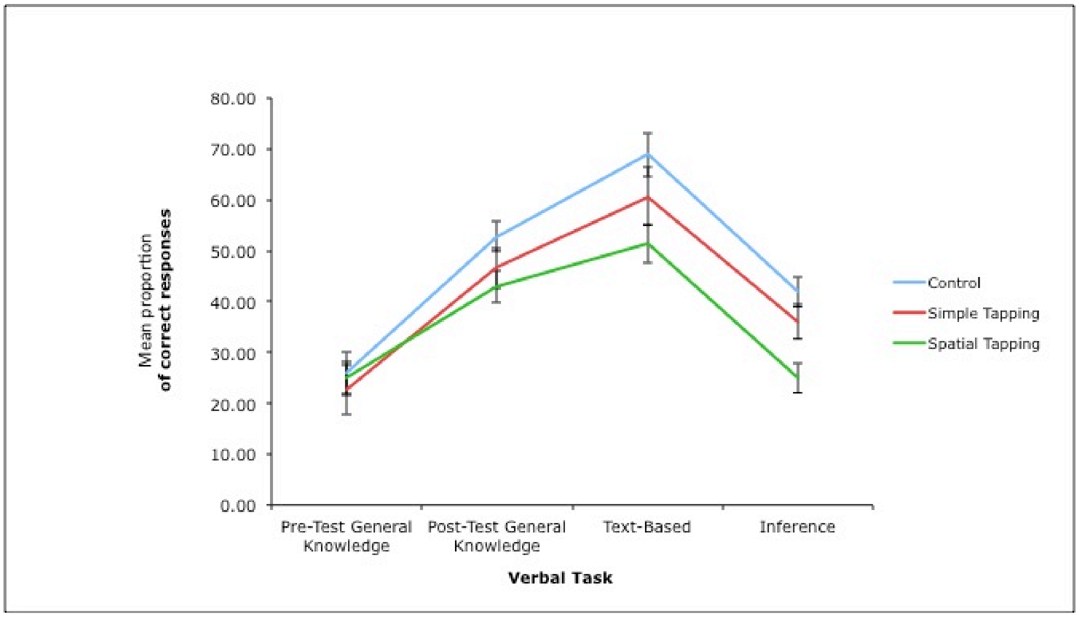


***Figure 3*. Experiment 3 scores on verbal tasks during pretest and posttests as a function of gesture condition.**

***Table 8*. Mean scores for performance on verbal tests in Experiment 3**

	Gesture (control)[a]	Simple tapping[b]	Spatial tapping[b]
Pretest			
General knowledge	25.93	22.66	24.90
Posttest			
General knowledge	52.78	46.59	42.91
Textbase	68.85	60.68	51.47*
Inference	41.92	36.03***	25.00**

[a]*n* = 27. [b]*n* = 26. * *p < .05*. ** *p < .05*. *** *p < .001*.

Gesture rate. It was expected that the simple tapping condition would reduce gesture production but would not engage the deeper motor control processes that might be shared by model-based reasoning. Participants' gesture rates in the simple tapping group were significantly higher than zero for all three posttest tasks (see Figure 4). The difference between mean gesture rates on all tasks between the simple tapping and the control condition was not significant, $F(1,6) = 3.1$, $p > .10$ (see Figure 4). Participants' gesture rates overall in the spatial tapping group were significantly greater than zero for all three posttest tasks (see Figure 4). Though the spatial tapping activity did not completely eliminate gesture production, it dramatically impaired it, consistent with findings reported elsewhere (Alibali & Kita, 2010; Hegarty et al., 2005).

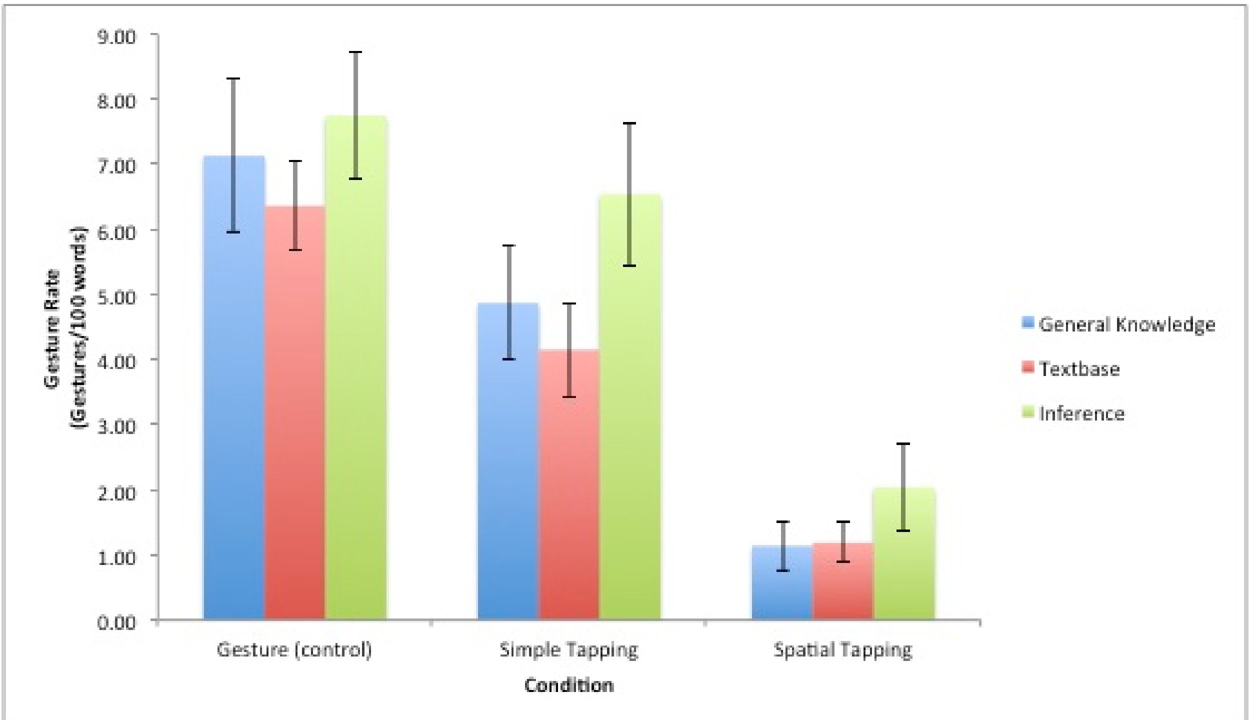


***Figure 4*. Gesture rate by condition for verbal posttest tasks in Experiment 3.**

Scores on the paper folding test did not reveal a significant effect on gesture rate in any of the tasks, $F(4, 73) = 2.1$, $p > .08$, so it was removed as a covariate measure. A three-way MANOVA was conducted with gesture condition as an independent variable and gesture rate on the three posttest tasks as dependent variables. The results revealed a main effect of condition on gesture rates during the verbal tasks, $F(4, 75) = 10.6$, $MSE = 23.4$, $p < .001$. Planned post-hoc Bonferroni comparisons ($\alpha = .05$) revealed that participants in the gesture (control) condition gestured more frequently during the posttest of general knowledge than did participants in either the simple tapping ($p < .01$) or spatial tapping condition, $p < .001$. During textbase questions, participants in the gesture condition gestured more than participants in either gesture-restricted condition ($p < .05$ for simple tapping; $p < .001$ for spatial tapping), while participants in the simple tapping condition gestured significantly more than those in the spatial tapping group, $p < .01$. During the inference-based questions, gesture was used more frequently by participants in the control group and in the simple tapping group than by those in the spatial tapping group, $p < .001$, while participants in the simple tapping group gestured more often than those in the spatial tapping group, $p < .01$.

Discussion

The results of the analyses for both test performance and gesture rates indicate that participants in the simple tapping and control conditions gestured more frequently than participants in the spatial tapping condition and performed better on the inference-based questions. The results of this experiment support Hypothesis 3, with its expectation that inference making is influenced by gesture production. Both tapping conditions reduced gesture production. However, the greater demands of planning and monitoring of motor control from the spatial tapping condition resulted in significantly greater reduction in inference making. Thus, the effect appears to be due to motor control influences on inference making. Internal processes used for SM enactment are implicated for their role in model-based reasoning. Next, in the General Discussion section, we examine the implications of these findings more broadly.

General Discussion

We consider our findings in terms of action-based theories of language and cognition, as well as implications they offer for reading comprehension, theories of mental model formation, and educational assessment. First, however, we summarize the major findings and address some current limitations of this work.

Summary and Limitations

Over three experiments, we found that gesture production was related to model-based reasoning in the form of inference making.

- Gestures were produced at a significantly higher rate during inference making (Hypothesis 1; Experiments 1 & 2).
- Gestures appeared to serve a compensatory role in SM formation in the absence of helpful illustrations (contrary to Hypothesis 2; Experiment 2).
- Gesture inhibition that loaded motor control systems had a substantial, negative impact on inference-making performance (Hypothesis 3; Experiment 3).

Though gesture production is often associated with spatial task demands in language and problem solving (Hostetter & Alibali, 2008; Krauss, 1998), spatial demands do not seem to account for all of the findings reported here. Gesture production did not differ when comparing responses to spatial versus non-spatial general knowledge test questions (Experiment 1). Furthermore, gesture production increased in the *absence* of more spatial information (Experiment 2). Finally, test item difficulty does not account for the pattern of observed findings. Taken together, these results contribute to a view that puts action in a central, causal role in the production of SMs when participants learn from reading a scientific text.

While these findings are promising, there are also important limitations to the current study. First, these findings all came from experiments using a single text and a common set of assessment items. In particular, the text addresses many spatial concepts associated with the circulatory system. Second, all of the experiments reported here used college students for whom reading and test taking are common practices. Extending these findings to other texts, including texts that are not organized around spatial topics, and to other populations of readers, is necessary to ensure their generalizability. Third, the current study explored only variants of the tapping method of gesture inhibition. The spatial tapping method used in Experiment 3 restricts gesture while also potentially making considerable spatial demands on the reader. The generalizability of these findings would also be enhanced if they arose using non-spatial inhibition methods such as holding an object (Alibali & Kita, 2010).

An Embodied Cognition Account of Mental Models

The GSA framework (Hostetter & Alibali, 2008) reviewed above provides an empirically rooted account for how cognitive processes that invoke resources involved in motor control can lead to gesture production. Yet, as some scholars point out (Alibali & Kita, 2010; Goldin-Meadow & Beilock, 2010), gesture can affect the course along which those cognitive processes proceed. For example, teaching specific movements to children can lead them to learn a grouping procedure used to solve mathematical equivalence problems (Goldin-Meadow, Cook, & Mitchell, 2009). Building on this work, we propose the Gesture as Model Enactment (GAME) framework, which posits that the content of SMs builds upon plans for hierarchical, goal-directed action. Within the GAME framework, SMs formed from reading are, in effect, cognitive simulations of coordinated actions that express the spatial and motoric imagery evoked during language comprehension (cf. Glenberg, 1999). Following GSA, we expect gesture production expresses this imagery during reading because the spatial demands of the text activate the associated motor systems. GAME further proposes that the motor movement from gestures can influence a currently activated SM when common motor control programs are engaged during the construction and running of these models. To achieve this, two sets of modifications to the GSA architecture are proposed (Figure 5).

For the first set of modifications, we draw on contributions from the modular selection and identification for control (MOSAIC) system (Haruno, Wolpert, & Kawato, 2001; Wolpert & Kawato, 1998), an architecture for regulating motor control in service of goal directed behavior. MOSAIC and HMOSAIC, its hierarchical variant (Haruno, Wolpert, & Kawato, 2003), have been used to provide computational accounts in a number of areas, including action production in an uncertain environment, social interaction and Theory of Mind (Wolpert, Doya, & Kawato, 2003), the emergence of giftedness (Vandervert, 2007), and action based accounts of language comprehension (Glenberg & Gallese, 2011).

Central to the HMOSAIC model is the simultaneous production of multiple, paired predictor-controller modules, each of which anticipates some plausible next state of the motor system and provide feedforward and feedback signals to regulate muscle behavior toward the intended goal. There is competition among the modules, and the system favors those modules that most closely track the real-time behavior of the body, even as adjustments are implemented.

As in HMOSAIC, motor-directed actions in GAME are modeled by parallel predictor-controller modules, C_i-P_i (Figure 5). In particular, each predictor, P_i, receives a copy of an intended action, and the efference copy provides rapid access to the projected future state of the motor system. That state is compared to the goal state, from which P_i generates future predictions. One key change we propose to the HMOSAIC architecture is that each specific prediction generated by P_i also produces a provisional **sm_i**, a potential modification to the currently activated SM in long-term memory based on the predictions from unit *i*. Selection of the most likely sm_i is used to update the reader's current SM.

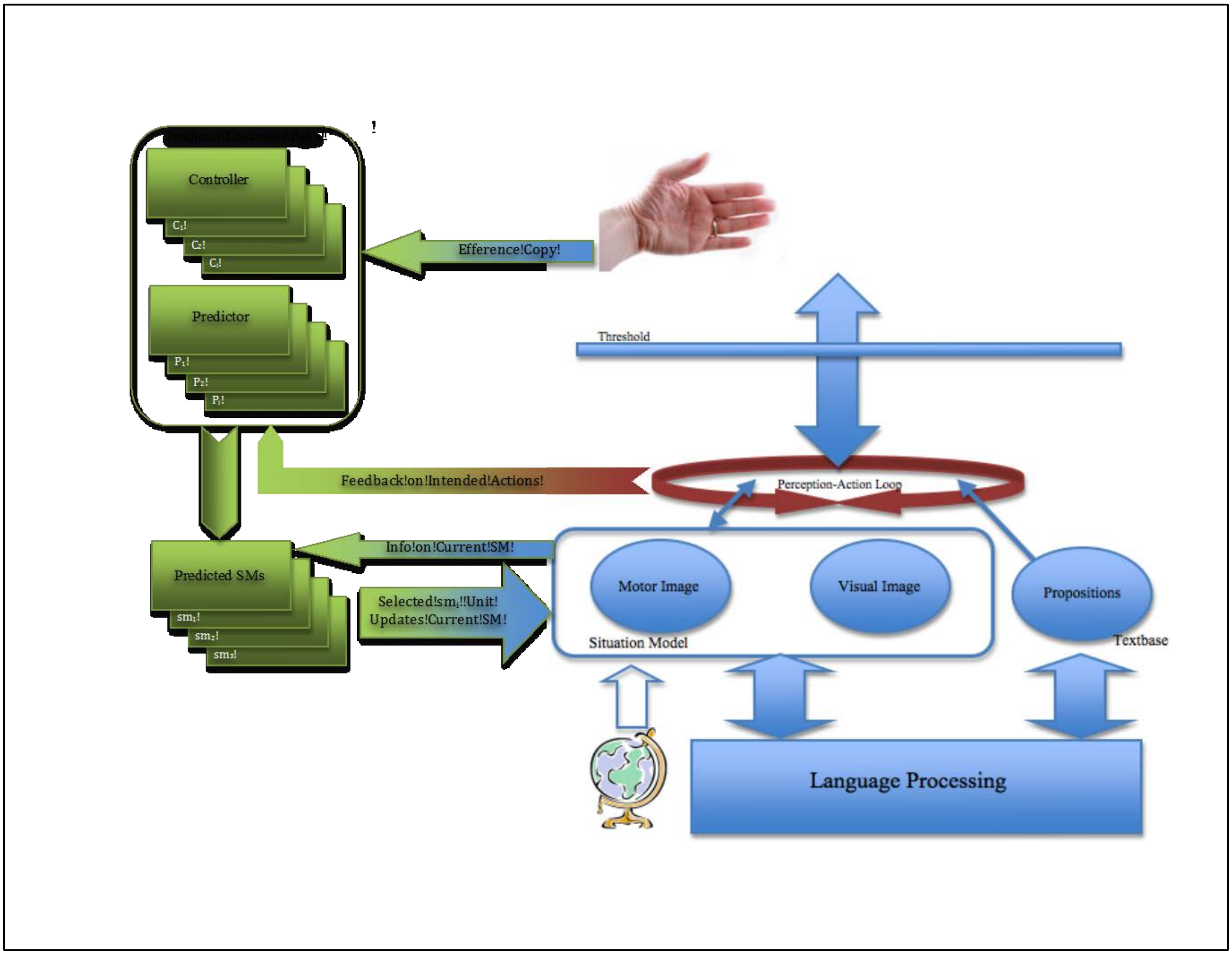


***Figure 5*. GAME model of gesture-mediated situation model formation**. Each predictor-controller pair, C_i-P_i, receives an efference copy of actual movement during gesture production that provides rapid access to the projected future state of the motor system during SM simulation, which is compared to the goal state. Provisional situation model modifications, sm_i, are produced by the parallel predictor-controller modules. Selection of the most likely sm_i is used to update the reader's current SM.

The second modification recognizes that the original GSA framework was designed to account for behavior during a range of problem solving activities, rather than reading comprehension and learning from text. In Figure 1, we modified the inputs from the original GSA model in order to distinguish input that is propositional from the imagery input that arises when activating a SM when learning from text (van Dijk & Kintsch, 1983; Kintsch, 1988). This distinction is especially important when modeling the effects of motor activity on cognition. There is now a body of evidence suggesting that some of the same neural resources used for reading (Pulvermüller, 2005), imagining (Jeannerod, 1994; Kosslyn & Thompson, 1999), reasoning (Kosslyn, Ball, & Reiser, 1978), or viewing actions (Martin et al., 1999) are recruited when actions are actually carried out. The same brain areas involved in action and perception are implicated when making linguistic inferences (Feldman & Narayanan, 2004) of the kind associated with SM-based reasoning. Based on this evidence, we posit that language understanding and model-based reasoning literally activate the same neural structures that are invoked during action and perception (Gallese & Lakoff, 2005).

We can thus articulate two novel predictions that follow from the GAME framework. First, we can predict that processing of SM based information is more likely to engage action production systems, such as gestures, because of their model-building role, rather than solely because of the activating role of spatial content. This prediction was supported by data from Experiments 1 and 2. Second, we can predict that the execution of motor control programs during gesture production will influence the ways that SMs are constructed. This prediction received support from Experiment 3.

Action-based accounts of higher order cognition such as language comprehension have recently received some support. Glenberg & Gallese (2011) propose that meaning in language derives from simulated actions that draw from the same systems that mediate hierarchical, goal-directed (i.e., planned) body movement. In like fashion, the theoretical account of situation model formation offered by the GAME framework emerges from such an action-based account of language. From this perspective, SMs are, in effect, cognitive simulations of coordinated plans of actions invoked during learning and testing (cf. Glenberg, 1999).

Motor-directed action plans in the proposed GAME framework rely on predictor-controller modules to generate multiple, parallel SM-like units to track the state of the imagined “world” referenced in the text, favoring those predictions that resemble the state of things as they unfold, and dispensing with those $\mathbf{sm_i}$s that appear to be least accurate. Selection of the most appropriate change to one current SM is likely to be based on several cognitive and affective factors, but for the current discussion we highlight two in particular: the degree to which the reader can formulate an action-based expression of a portion of the SM that is currently attended to (following our proposed modification of GSA); and simulated sensory feedback that compares the current state of the motor system to the predicted (simulated) state (following HMOSAIC).

Taken together, the contributions of imagery-enabled gesture production as modeled in GSA (Figure 1) and the influences of motor control on comprehension processes, particularly SM formation, provide the components of a complete GAME framework. In this way, gesture production in GAME has two complementary characteristics: It is expected that gestures will express the imagery readers’ activate during comprehension and SM reporting; and gesture production is implicated as a causal factor in model formation, as common motor control resources are engaged during the construction and execution of SMs. While we propose a motor control account, where the SM formation process is influenced by hand movements that are driven by simulated actions, we do not claim that *all* SM formation stems from motor behavior.

In such a coarse description of a model there are many details left unspecified. One aspect of particular importance is an explication of where the provisional SM model alterations—the $\mathbf{sm_i}$—come from. In fact, Glenberg & Gallese (2011) argue that the core functionality of the HMOSAIC "predictor corresponds to a mental model."

Reciprocity of Action Control and Cognitive Processes

The GAME framework, and its explicit regard for the *reciprocity* between action and cognitive processes, raises a question as to whether the intentionality of one's actions matters in learning. A number of findings suggest that the actions themselves play a consequential role and can be effective even when participants unintentionally engage in them. In one line of research, facial expressions that are commensurate with affective states (e.g., smiling to indicate happiness) but induced without the agent's awareness (e.g., directed to hold a pen without touching one's lips) can influence behavior in accordance with the simulated emotion (Havas, Glenberg, & Rinck, 2008; Niedenthal, 2007; Tom, Pettersen, Lau, Burton, & Cook, 1991). In language comprehension several studies have induced motor movements that are compatible or incompatible with the physical or metaphoric actions represented in words—the Action Compatibility Effect—to show that the action system affects language processing (e.g., de Vega & Urrutia, 2011; Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006). In insight problem solving, Thomas & Lleras (2007) showed that those participants who were directed to shift their gaze in a pattern compatible with the solution to Duncker's Radiation Problem were the most successful, even though participants were not consciously aware of the relation of the eye gazing patterns and the problem solution. Communicative or semantic intention may ultimately provide some additional resources that modulate the effect actions have on cognitive processes, but there is substantial evidence suggesting that intention is not necessary to do so.

Implications for Computer-Based Testing

This study also failed to replicate a prior finding in the literature (Butcher, 2006) showing that simple, relevant illustrations help readers form better-developed SMs. We surmise that methodological difference in typing versus speaking one's test responses played a central role in this difference in performance patterns across the two investigations. Here, we consider some of the practical implications of these differences; specifically, what these findings suggest in terms of knowledge assessment practices. Participants in our study whose gesture production was severely impaired by the demands to produce an ever-changing tapping pattern during testing showed significantly lower performance on items testing inference making, even though they showed learning gains for general knowledge items. Because of the close similarity of spatial tapping to typing, this finding raises questions about the effect that computer-based assessments may have on certain cognitive processes of test takers. If typing responses interferes with the motor control systems in ways that impair SM development, then we could see degraded test performance on tasks such as inference making and other forms of higher order cognition (for a number of hierarchies of cognitive complexity, see Bloom, 1956, 1987; Kawanaka & Stigler, 1999; Mehan, 1979; Nathan & Kim, 2009; Nystrand, Gamoran, Kachur, & Prendergast, 1997; Wells & Arauz, 2006). From a methodological perspective this may not be an issue in experimental psychology. However, from the standpoint of educational testing, it would be a serious limitation to the enterprise of computer-based testing if typed assessments inadvertently target some of the most sought-after performance objectives in reading, while showing relatively little effect on subordinate processes such as recognition, repetition, and factual recall. The growing practice of online testing and typed responses needs to be carefully monitored in light of these findings. The result could be a distorted picture of the capabilities of readers and the effectiveness of literacy and language arts programs. Alternatively, if gesture production effectively enhances areas of higher order cognition such as inference making, one may use these findings as justification to explore alternative assessment methods that encourage richer forms of responses, like those that occur in communicative interactions. By engaging both verbal production and body-based resources such as gestures during assessment we may encounter new insights about the nature of reading comprehension and learning from text that can fundamentally alter the ways we measure and describe learning.

Conclusions

The spatial, dynamic nature of MMs that people construct when learning through reading would seem to have its basis, not in the proposal structures and symbolic networks of the text itself, but in the bodily realm of actions that the text invokes. Participants in the studies reported here gestured more frequently when responding to model-based inferences, used gestures apparently as a body-based resource to support inference making when other resources were absent, and exhibited variation in inference making when gesture was inhibited. The causal role of gesture in model formation aligns with an emerging view that considers the reciprocal influences between action and cognitive processes that may lead to new insights about the nature of human behavior and contribute to new ways of fostering learning.

References

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture*, *10*(1), 3–28.

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*(2), 169–88.

Alibali, M. W., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, *15*, 593–613.

Alibali, M. W., & Nathan, M. J. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 349–365). Mahwah, NJ: Erlbaum.

Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 333–356.

Barsalou, L. W. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, *28*(1), 61–80.

Bloom, B. S. (1956/1987). *Taxonomy of educational objectives, Book 1: Cognitive domain*. New York, NY: Longman.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*(2), 193–209.

Butcher, K. R. (2003). *Effects of diagram complexity on comprehension processes and learning outcomes* (Unpublished doctoral dissertation). University of Colorado, Boulder, CO.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*(1), 182–197.

Butcher, K. R., & Kintsch, W. (2004). Learning with diagrams: Effects on inferences and the integration of information. In A. Blackwell, K. Marriott, & A. Shimojima (Eds.), *Diagrammatic representation and inference: Third international conference, diagrams 2004, Cambridge, UK, March 22–24, 2004: Proceedings* (pp. 337–340). New York, NY: Springer.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review, 96*(1), 168–174.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71.

Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299–324). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Dordrecht, The Netherlands: Springer.

Cutica, I., & Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science*, *32*(5), 921–935.

de Vega, M., & Urrutia, M. (2011). Counterfactual sentences activate embodied meaning. An action-sentence compatibility effect study. *Journal of Cognitive Psychology, 23*(8), 962–973. doi: 10.1080/20445911.2011.590471

de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 373–385.

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Services.

Farah, M. J. (1996). The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.

Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, *89*(2):385–392.

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes*, *13*(2), 175–190. doi: 10.1080/01638539009544752

Gallese, V., & Lakoff, G. (2005). The brain’s concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, *22*(3), 455–479.

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

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

Glenberg, A. (1999). Why mental models must be embodied. *Advances in Psychology, 128,* 77–90. doi: 10.1016/S0166-4115(99)80048-X

Glenberg, A. M., & Gallese, V. (2011). Action-based language: A theory of language acquisition, comprehension, and production [Special issue: Research report]. *Cortex*. doi:10.1016/j.cortex.2011.04.010

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language*, *31*(2), 129–151.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*(1), 69–83.

Goldin-Meadow, S. (2003). *Hearing gestures: How our hands help us think*. Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674. doi:10.1177/1745691610388764

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. 2009. Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–272. doi: 10.1111/j.1467-9280.2009.02297.x

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*(4), 565–610.

Gulberg, M. (1998). *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish*. Lund, Sweden: Lund University Press.

Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation 13*(10), 2201–2220.

Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, *14*(3), 436–441.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*(6), 280–285.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*(6), 717–742.

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation*, *5*(4), 333–356.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*(1), 62–82.

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skill and gesture production. *Gesture*, *7*(1), 73–95. doi: : 10.1075/gest.7.1.05hos

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 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*(2), 245–257.

Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation,13*, 2201–2220.

Haruno, M, Wolpert, D. M., & Kawato M. (2003). Hierarchical MOSAIC for movement generation. *International Congress Series, 1250*: 575-590, 2003.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–245.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, 71–115.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, 50, 189–209.

Kawanaka, T., & Stigler, J. W. (1999). Teachers' use of questions in eight-grade mathematics classrooms in Germany, Japan, and the Unites states. *Mathematical Thinking and Learning, 1*(4), 255–278.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction*, *27*(3), 175–200. doi: 10.1207/s15327973rlsi2703_2

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

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes*, 16, 193–202.

Kintsch, W. (1998) *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W., Britton, B. K., Fletcher, C. R., Kintsch, E., Mannes, S. M., & Nathan, M. J. (1993). A comprehension-based approach to learning and understanding. *The Psychology of Learning and Motivation*, 30, 165–214. doi: 10.1016/S0079-7421(08)60297-2

Kintsch, W., Welsh, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*(2), 133–159. doi: 10.1016/0749-596X(90)90069-C

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, *24*(2), 145–167.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32.

Knauff, M. (2009). A neuro-cognitive theory of deductive relational reasoning with mental models and visual images. *Spatial Cognition and Computation*, *9*(2), 109–137.

Kosslyn, S. M., & Thompson, W. L. (1999). Shared mechanisms in visual imagery and visual perception: Insights from cognitive science. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 47–60.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*(2), 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). New York, NY: Cambridge University Press.

Martin, A., Ungerleider, L. G., & Haxby, J. V. (1999). Category specificity and the brain: The sensory/ motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1023–1036, 2nd ed.). A Bradford book. Cambridge, MA: MIT Press.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92(3), 350–371. doi: 10.1037//0033-295X.92.3.350

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

McNeill (2000). *Language and gesture*. Cambridge, UK: Cambridge University Press.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

Nathan, M. J. (2008). An embodied cognition perspective on symbols, grounding, and instructional gesture. In M. de Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 375–396). Oxford, UK: Oxford University Press.

Nathan, M. J. & Kim, S. (2009). Regulation of teacher elicitations in the mathematics classroom. *Cognition and Instruction*, *27*(2), 91–120.

Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York, NY: Teachers College Press.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*(4), 359–400.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582. doi: 10.1038/nrn1706

Rimé, B., Schiaratura, L., Hupert, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*(4), 311–325. doi: 10.1007/BF00991870

Rinck, M., Hahnel, A., Bower, G. H., & Glowalla, U. (1997). The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 622–637.

Sassenberg, U., & van der Meer, E. (2010). Do we really gesture more when it is more difficult? *Cognitive Science*, *34*(4), 643–664.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition*, *60*(3), 205–234.

Schmalhofer, E., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, *25*(3), 279–294. doi: 10.1016/0749-596X(86)90002-1

Schwartz, D., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, *20*(4), 457–497. doi: 10.1016/S0364-0213(99)80012-3

Singer, M., Radinsky, J., & Goldman, S. (2008). The role of gesture in meaning construction. *Discourse Processes*, *45*(4), 365–386.

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, *31*(2), 261–292. doi: 10.1016/0749-596X(92)90014-O

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, *35*(3), 371–391. doi: 10.1006/jmla.1996.0021

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, 14, 663-668.

Tom, G., Pettersen, P., Lau, T., Burton, T., & Cook, J. (1991). The role of overt head movement in the formation of affect. *Basic and Applied Social Psychology*, *12*, 281.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension*. New York, NY: Academic Press.

Vandervert, L. R. (2007). Cognitive functions of the cerebellum explain how Ericsson's deliberate practice produces giftedness. *High Ability Studies*, *18*(1), 89–92.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*(4), 395–407.

Wells, G., & Arauz, R. M. (2006). Dialogue in the classroom. *Journal of the Learning Sciences*, *15*(3), 379–428.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, *114*(4), 591–600.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 1–11.

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by latent semantic analysis. *Discourse Processes*, *25*(2), 309–336.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 593–602. doi: 10.1098/rstb.2002.1238

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks, 11*(7-8), 1317–1329. doi: 10.1016/S0893-6080(98)00066-5

Woods, D., & Fassnacht, C. (2009). Transana (version 2.41) [Qualitative analysis software]. http://www.transana.org. Madison, WI: The Board of Regents of the University of Wisconsin System.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297. doi: 10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., Magliano, R. A., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386–397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185.

Zwaan, R., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11.

Appendix A: Example of Tutorial Text

When you breathe out, you get rid of this carbon dioxide. Oxygen from the air sacs passes into the blood capillaries, and the circle begins again. Even though it is the same blood that carried the carbon dioxide wastes, when it has unloaded them and taken on a new cargo of oxygen, we can think of it as fresh blood -- it is as good as new.

Continue

Page 31 of the illustrated tutorial (used in Experiments 1, 2, and 3)

When you breathe out, you get rid of this carbon dioxide. Oxygen from the air sacs passes into the blood capillaries, and the circle begins again. Even though it is the same blood that carried the carbon dioxide wastes, when it has unloaded them and taken on a new cargo of oxygen, we can think of it as fresh blood -- it is as good as new.

Continue

Page 31 of the non-illustrated tutorial (used in Experiment 2 only)

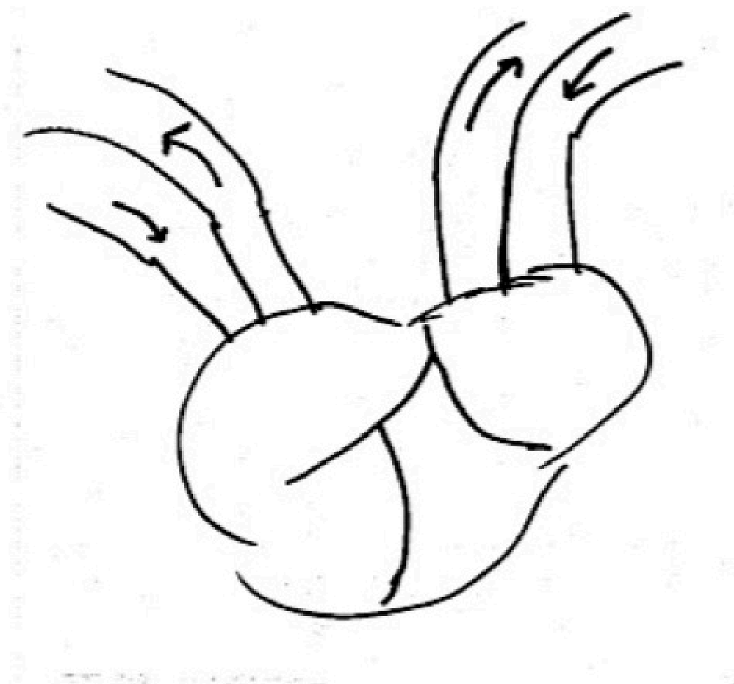
Appendix B: Drawing Scoring Rubric and Examples

B-1. No Loop (1 point)

1. Blood is pumped from the heart to the body.
2. Blood does not return to the heart.

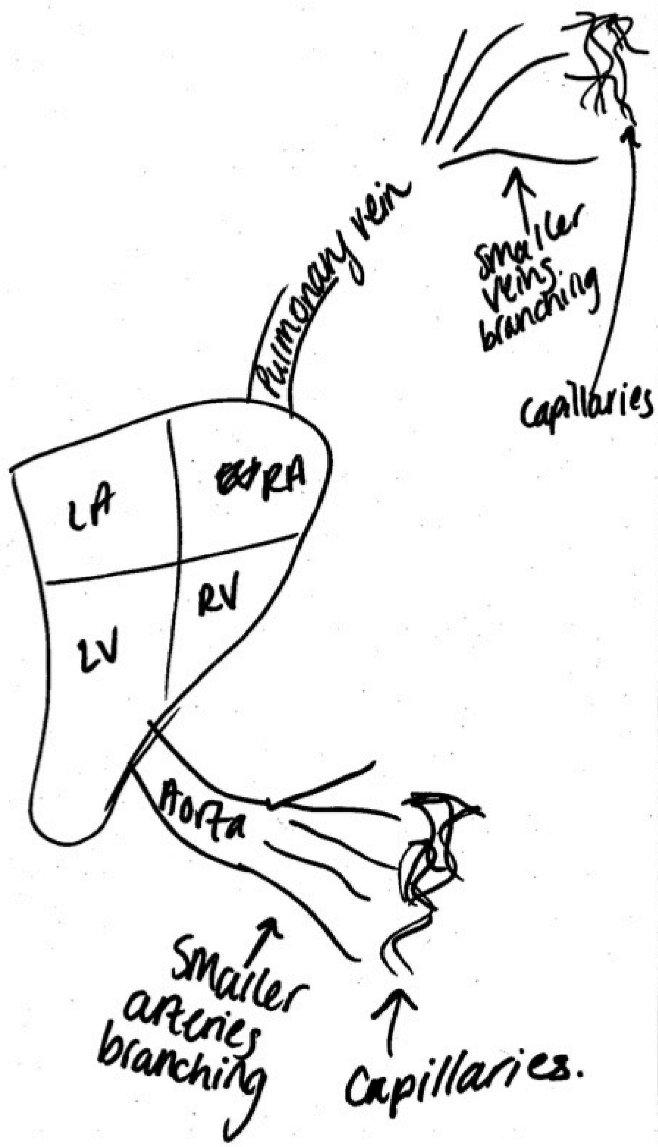
B-2. Ebb and Flow (2 points)

1. Blood is primarily contained in blood vessels.
2. Blood is pumped from the heart to the body.
3. Blood returns to the heart by way of some blood vessel.



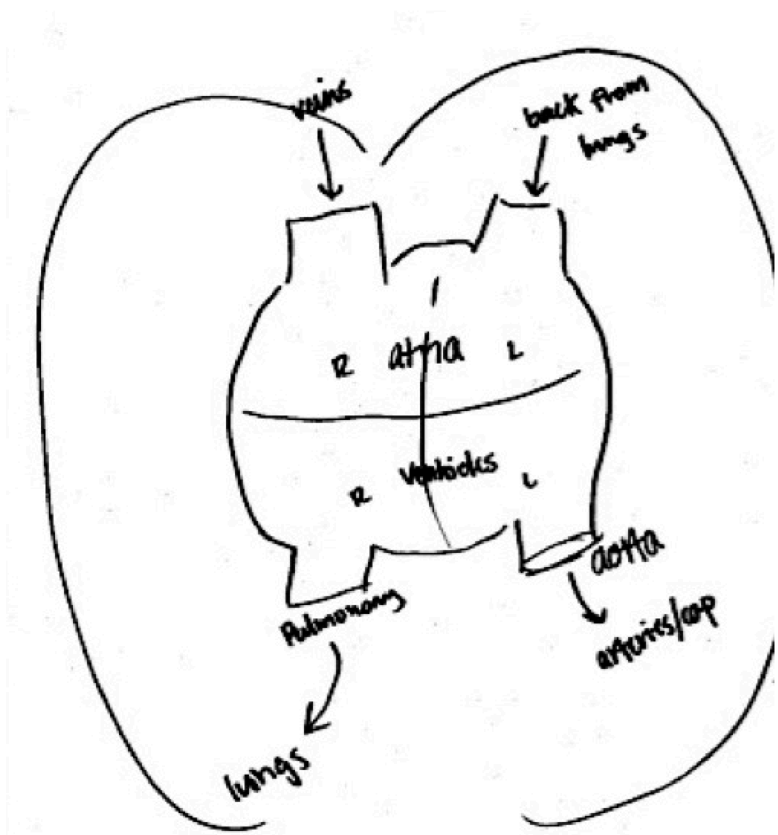
B-3. Single Loop (3 points)

1. Blood is primarily contained in blood vessels
2. Blood is pumped from the heart to the body
3. Blood returns to the heart from the body.



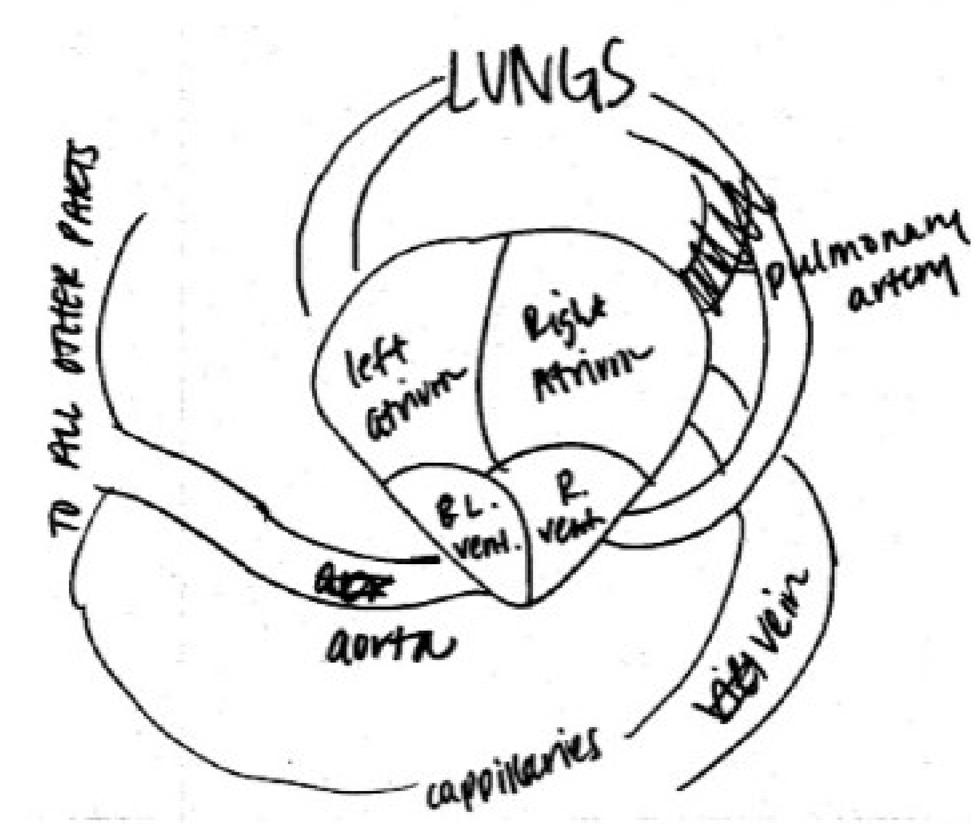
B-4. Single Loop with Lungs (4 points)

1. Blood is primarily contained in blood vessels
2. Heart pumps blood to body or to lungs.
3. Blood returns to heart from body or from lungs.
4. Blood flows from lungs to body or from body to lungs without return to heart in between.
5. Lungs play a role in the oxygenation of blood.



B-5. Double Loop-1 (5 points)

1. Blood is primarily contained in blood vessels.
2. Heart pumps blood to body.
3. Blood returns to heart from body.
4. Heart pumps blood to lungs.
5. Blood returns to heart form lungs.
6. Lungs play a role in the oxygenation of blood.



B-6. Double Loop-2 (6 points)

1. All features from Double Loop-1
2. Heart has four chambers
3. Septum divides heart lengthwise - sense of preventing mixing of blood.
4. Blood flow through heart is top to bottom.
5. At least three of the following:
   a. Blood flows from right ventricle to the lungs
   b. Blood flows from lungs to left atrium.
   c. Blood flows from left ventricle to body.
   d. Blood flows from body to right atrium

Appendix C: Tapping Stimulus Used in Experiment 3 for Simple and Spatial Tapping Conditions.

1	2
3	4