Gesture as model enactment: the role of gesture in mental model construction and inference making when learning from text

Mitchell J. Nathan\textsuperscript{a} & Chelsea V.J. Martinez\textsuperscript{b}
\textsuperscript{a} Department of Educational Psychology, University of Wisconsin-Madison, Madison, WI, USA
\textsuperscript{b} Education and Psychology Faculty, Highland Community College, Freeport, IL, USA
Published online: 01 Apr 2015.

To cite this article: Mitchell J. Nathan & Chelsea V.J. Martinez (2015) Gesture as model enactment: the role of gesture in mental model construction and inference making when learning from text, Learning: Research and Practice, 1:1, 4-37, DOI: 10.1080/23735082.2015.1006758

To link to this article: http://dx.doi.org/10.1080/23735082.2015.1006758

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &
Gesture as model enactment: the role of gesture in mental model construction and inference making when learning from text

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

aDepartment of Educational Psychology, University of Wisconsin-Madison, Madison, WI, USA; bEducation and Psychology Faculty, Highland Community College, Freeport, IL, USA

(Received 14 June 2014; accepted 2 January 2015)

The question of the relationship between gesture production and mental models was explored in three experiments focusing on inference making when learning from reading a scientific text. Participants engaged in one-on-one interviews after reading an illustrated tutorial on the human circulatory system. Participants gestured more frequently when responding to mental model-based inference test questions (Experiment 1; N = 22) than textbase or general knowledge questions. Learning from a text lacking illustrations resulted in comparable inference making (Experiment 2; N = 48) but elevated gesture production. Restricting gesture production with hand tapping led to inferior inference making (Experiment 3; N = 79). Results support an embodied cognition view of mental model formation where inference making stimulates gesture production, and restricting gestures selectively impairs inference making. The results are interpreted within the Gesture as Model Enactment (GAME) framework. Activation of motor control systems via gesture support simulated actions that contribute to mental model formation. Implications for learning and assessing knowledge are explored.

Keywords: embodied cognition; gesture production and inhibition; inference making; knowledge assessment; learning from text; mental models; situation models; spatial relations

The mind has a rich capacity for representing ideas and the state of the world, both real and imagined. As an example, consider how readers represent the information conveyed by these sentences, from the work by Bransford, Barclay, and Franks (1972, p. 195):

(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 strictly 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) when asked, because when we read, we do not just represent the words; we elaborate, make contextually relevant inferences, and construct for ourselves mental models (hereafter, MM) of the objects, actions and relations of the situation that is referred to (Gentner & Stevens, 1983; Johnson-Laird, 1980). MMs are often described as

*Corresponding author. Email: MNathan@Wisc.edu

© 2015 Taylor & Francis
imagistic, holistic, and dynamic, and can include runnable, simulation-like accounts of a phenomenon (e.g., Clement, 1983). In contrast to the MM, the textbase (TB) is described as a propositional, sequential, and static representation of meaning, which is developed directly from the information explicitly presented in a passage during reading (Fletcher & Chrysler, 1990; Kintsch, Welsh, Schmalhofer, & Zimny, 1990; Zwaan, Langston, & Graesser, 1995, Zwaan, Magliano, & Graesser, 1995). A reader’s MM may not distinguish (1a) from (1b) since the fish swims beneath them both. People who read (2a) (the beside relation) almost never confuse it with (2b), since the MMs of the two situations clearly differ both spatially and propositionally.

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, 1994; McNeill, 1992). Gestures are intriguing because, like MMs, the nature of the information conveyed is holistic, dynamic, and spatial (Kendon, 2004; McNeill & Duncan, 2000); which enables gestures to reveal or enact spatial, temporal and causal relations (Goldin-Meadow, 2003), and complement the information presented verbally (Goldin-Meadow, 2003; Nathan, 2008).

We start with a presumption that MMs are inherently dynamic (e.g., runnable) mental processes that model or simulate properties of the dynamic, situated world in which they are formed. It is via their dynamic, situated qualities that readers make spatial and causal inferences when learning from text. The following studies extend beyond this view to explore the conjecture that MMs are embodied, meaning they are grounded in body-based behaviours and systems such as those involved in perception, action planning, motor control, and physical interactions in the world, be they real or imagined (Barsalou, 2008; Wilson, 2002).

The embodied nature of MMs and model-based inference making is explored from two directions: First, that body-based resources in the form of gesture production increases with increasing demands to making inferences (Experiments 1 and 2); and second, that restricting gesture production selectively impairs inference making (Experiment 3). Spatial demands and inference-making difficulty are explored as alternative interpretations of the findings. Progress on questions of the nature and potential inter-relationship of gestures and complex cognition advances our scientific understanding of how we think and engage our bodies as we learn about the world.

**General characteristics of mental models**

Broadly construed, MMs are psychological constructs, often 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 also capture temporal, intentional, and perspectival relations (e.g., Goldvarg & Johnson-Laird, 2001; Schaecken, 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 analogue information (e.g., relative proximity), while also depicting abstract relations among classes of entities. MMs also consolidate information that an individual has learned about a domain, enabling the individual to rapidly access that information when making judgements about new situations (Johnson-Laird, 1994; Vosniadou & Brewer, 1992).
Some scholars (e.g., Barsalou, 1999; Glenberg, 1999) suggest that MMs provide relational and predictive information that may prepare people for pending situated action.

The situation model (SM) is a type of MM formed when integrating propositions within the text with the reader’s prior knowledge (van Dijk & Kintsch, 1983). SMs are the mental models readers form when modelling the situation referred to (explicitly or implicitly) by words of a text. 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 document sources (e.g., Wiley & Voss, 1999). 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). For example, Perrig and Kintsch (1985) showed that inference making from a text (describing the layout of a fictitious town) depended upon the formation of an SM, and that the form the SM took (manipulated as the route while driving through the town versus the geographical survey of the layout of the town) affected the types of inferences readers made. Butcher (2006) investigated readers’ SMs of the human circulatory system. Readers of a non-illustrated version of the text formed a weaker SM than readers of the text with simple illustrations, as measured by inference making and drawing quality, 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 hand gestures. Cutica and Bucciarelli (2008) presented participants with video-recorded stories with or without accompanying gestures. Stories presented with gestures led to better recollection and more inferences, though surface structure recognition was superior in the no-gesture condition. Cutica and Bucciarelli concluded that individuals who listen to a discourse that is accompanied by gesture develop a richer SM, but a weaker verbatim representation of the narrative.

Glenberg (1999), taking an embodied cognition perspective, argues that systems that only represent meaning of the text, but do not simulate the affordances of the referenced objects and events – their shapes, weights, flexibility, and sizes, for example – provide an impoverished account of how people actually process texts. One conjecture flowing from this that guides our work is that readers may engage action systems in the body as a way to apprehend and simulate these affordances, which may result in an increased level of body-based behaviours, such as increased gesture production. It also follows, in reciprocal fashion, that restrictions on relevant body-based behaviours may impair one’s ability to simulate these affordances, result in an impoverished MM that does a poor job supporting inference-making processes. Since gestures offer one way to capture these affordances, we turn next to briefly review that literature.

**General characteristics of gestures**

Much of gesture research can be considered along two inclusive lines of research: gestures as a multimodal resource that is used for communication in social interactions; and gesture as a cognitive resource that exhibits and influences one’s thinking.

Gestures are part of a multimodal system of interpersonal communication that are ubiquitous in social interactions (e.g., Goodwin, 2000; Kendon, 1980). They play an important role in managing conversation such as turn taking (Duncan, 1972). Gestures are
clearly meant to be seen and are used more frequently when interlocutors can see the speaker than when the speaker is hidden (Alibali, Heath, & Myers, 2001).

Along with conversation, gesture is integral to teaching and learning settings (Alibali & Nathan, 2012; Arzarello & Edwards, 2005; Edwards, 2009; Nathan, Eilam, & Kim, 2007; Radford, 2000; Roth, 2001). Studies of classroom instruction reveal that teachers use gestures as part of multimodal communication to link ideas and representations (Alibali, Nathan, Wolkgram, et al., 2014), and to foster and maintain common ground (Alibali, Nathan, Church, Wolkgram, Kim, & Knuth, 2013). From an interaction analysis perspective, Scopelitis, Mehus, and Stevens (2010) view gestures as a medium for representation and for coordinating external representations, so gestures serve as both technical and communication resources. Furthermore, gestures are influential, and can enhance interlocutors’ understanding and learning (Alibali, Young, Crooks, Yeo, Wolkgram, Ledesma, Nathan, Church, & Knuth, 2013; Hostetter, 2011).

In addition to its presence in social interaction, gestures convey information about the speaker’s mental state, reasoning, and cognitive development. 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). 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 more substantive reasoning involved in conservation (Alibali, Kita, & Young, 2000).

Gesture may convey knowledge that is based on motor simulation and not readily verbalizable, as when performing mechanical reasoning (Hegarty, 2004). Gesture analyses conducted by Abrahamson (2004) characterize student understanding of mathematics concepts as negotiations between how they imagine a concept and the formal, culturally constituted representations of that concept. 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 reasoning and learning (Alibali & Kita, 2010; Goldin-Meadow & Beilock, 2010; Nathan, 2014). Restricting gestures disrupts that feedback channel, leading to less frequent imagery-based language when they talk (Rime, Schiaratura, Hupert, & Grysselinckx, 1984). 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 may selectively impair causal reasoning (Hegarty, Mayer, Kriz, & Keehner, 2005).

The theory of Gesture as Simulated Action (GSA; Hostetter & Alibali, 2008) provides an embodied cognition account of the multimodal production of speech-accompanied gestures. Under GSA, the activation of neural areas employed by action and perception are also invoked via simulated action and perception for mental (i.e., off-line) processing of language, imagery and planning. As suggested by Figure 1, gestures in the GSA framework arise during speaking when pre-motor activation, formed in response to motor or perceptual imagery is activated beyond a speaker’s current gesture threshold. The threshold is the level of motor activation needed for a simulation to be expressed in overt action, such as gesture; this threshold can vary depending on 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). In the modified view of GSA (shown in Figure 1), the model explicitly includes inputs from the textbase and SM during reading.
Hypotheses

In this investigation, we propose three specific hypotheses that extend our current understanding of the relationship of gesture production and SM formation when learning from text. Hypothesis 1 states that 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 states that gesture production is responsive to model-based information. Operationally, we expect to see that stimuli 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 states that the manipulation of gesture production will influence 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 overtly spatial (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Hostetter & Alibali, 2008; Wagner, Nusbaum, &
Goldin-Meadow, 2004, we expect to see a broad drop in performance across all of the post-test items. However, due to the combined effect of general attentional load and a targeted loss of model-based resources, we expect performance on inference items for participants whose gestures are inhibited will exhibit the greatest performance drop.

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

**Materials**
Participants in each experiment completed a set of tutorial and test activities about the human heart and circulatory system developed by Wolfe et al. (1998) and Butcher (2006). Participants completed a two-part pre-test followed by a computer-based tutorial and a two-part post-test. The pre- and post-tests each included a drawing activity and verbal test items comprised of a series of questions about the heart and circulatory system.

**Drawing tasks**
During the drawing activity, the interviewer handed participants a blank 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”. Thus, the drawings were created in the presence of the experimenter, and once completed were thoroughly described to the experimenter before moving on to the next activity. Participants were not given a time limit for the creation of the drawing or their explanation of it. After participants finished explaining the drawing, the interviewer placed the drawing face down and out of sight of participants so they could not refer to the drawings in later interview tasks. Participants completed identical drawing tasks before and after the computer-based tutorial.

**Verbal tasks**
The pre- and post-test questions were adopted from Butcher (2006; also see Wolfe et al., 1998). Participants verbally responded to the pre- and post-test items, in contrast to Butcher (2006), who had participants type their answers. Pre-test and questions included “general knowledge” (GK) questions designed to assess the participants’ current understanding of the human heart and circulatory system (i.e., “What is a capillary?”). Post-test questions also included TB and inference (INF) questions. TB questions asked participants to recall information that had been presented in the computer-based tutorial (i.e., “Where does blood entering the right ventricle come from?”). INF 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 GK questions, 8 TB questions, and 5 INF questions.
Tutorial

After completing the pre-test, participants read a tutorial about the human heart and circulatory system used in prior studies (Butcher, 2006). The tutorial was presented as a computer-based text. It was comprised of 43 web pages; each page had a small amount of text about the heart and circulatory system, and 32 pages of the tutorial featured text accompanied by a colour illustration relevant to the text. See Appendix 1 for an example of the tutorial text and illustrations.

Participants

All participants were undergraduate students recruited from Educational Psychology courses at the University of Wisconsin-Madison, and over 18 years of age. They were compensated with extra credit points in that course upon completion of the experiment.

Procedure

In each experiment, participants completed one-on-one interviews with a member of the investigative team. The interviews were conducted in a private office furnished with two desks, two chairs, a computer monitor, and a video camera visible to participants. Participants were videotaped as they completed an initial drawing activity, a pre-test of general knowledge of the circulatory system, the computer-based tutorial, and several post-test activities: a post-drawing activity, a post-test of general knowledge, textbase assessment, and inference making.

Participants sat in front of the video camera at a desk with a computer monitor and were able to turn and face the interviewer throughout the session. Participants advanced through the tutorial by using a mouse to click to the next page, but could not move backwards. The location of the button for advancing to the next page of text shifted screen position throughout the task to avoid accidental advancement through the tutorial. The tutorial typically took about 30 minutes for participants to complete. Participants were not permitted to take notes. Participants were not timed as they completed the tutorial and were not given an upper time limit.

Coding and scoring of participants’ data

Drawings

Participants’ drawings at both pre-test and post-test were scored according to a six-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 their responses. Rubric scores were based on the accurate function and interrelations of anatomy and physiological processes, rather than shape or physical resemblances. See Appendix 2 for examples of how the rubric was 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). Codes were assigned according to the following criteria.
Speech codes. Speech codes were used to identify 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: This applied to participants’ language that was not derived from the tutorial, for example, “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: This applied to restatements of material presented in the tutorial, that is, “A ventricle is one of the bottom chambers where blood is pumped out of the heart”.
Verbatim: Statements taken directly from the tutorial text, that is, “A capillary is a microscopic vein or little veins” were assigned this code.
Metacognitive: Metacognitive statements demonstrated participants’ reflection on their knowledge, that is, “No. Yes. I’m gonna go with that, although I might have ’em confused”.
I don’t know: This applied to statements made where the participant did not attempt answer the question posed, that is, “I have no idea”.
Unscripted question: This was applied when participants asked a question.

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

Gesture utterance: This was applied to every segment of speech that included gesture.
Non-gesture utterance: This was applied to every segment of speech without gesture.
Point: This was applied to speech segments of that included deictic gestures. Figure 2a provides an example of a pointing gesture.
Beat: This was applied to segments of speech that included gestures that did not display semantic content but occurred in rhythm with the participants’ speech.

Figure 2. Illustrations of gesture types observed in the study. (a) A pointing gesture. (b) An iconic gesture. (c) Producing an iconic gesture while making an inference about how blood would flow when the valves leading out of the ventricles did not properly close.
Representational: This was applied to segments of speech that included gestures that displayed semantic content, including iconic and metaphoric gestures. **Figures 2b and 2c show example representational gestures when reporting on general knowledge and when making an inference, respectively.**

Gesture rate, an accepted measure of gesture production (Alibali & Kita, 2010; Hostetter & Alibali, 2010), was calculated for each task in the interview (pre-draw, pre-test of general knowledge, post-draw, post-test of general knowledge, post-test of textbase material, and post-test of inference material). Gesture rate was calculated as the number of gestures produced per 100 words spoken. Only pointing and representational gestures were used when calculating gesture rates. 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 to questions in the pre- and post-tests were scored according to a rubric established by Butcher (2006). The pre- and post-tests of general knowledge included 18 questions worth a total of 38 points; the post-test textbase portion included eight questions worth a total of 17 points; the post-test inference portion included five questions worth a total of 11 points.

**Experiment 1**

The goal of Experiment 1 was to examine the relationship between participants’ gesture production and model-based reasoning when learning from a text. Here we test Hypothesis 1, which states that rate of gesture production will be highest when readers draw heavily from their SMs during inference making.

**Method**

The primary manipulation is question type, with the central prediction that gesture rate will be higher for inference test items than for items testing general knowledge or textbase recall.

**Participants**

Twenty-two participants (as described in the General Methods) completed the one-on-one interviews and heart text.

**Materials and procedure**

Participants were run individually and sessions were videotaped in their entirety. Each person completed the same series of pre- and post-test activities described in the General Method section in a one-on-one interview-type setting. The materials were presented in the following order: pre-draw activity, pre-test of general knowledge, computer-based tutorial, post-draw activity, post-test of general knowledge, post-test of textbase materials, and post-test of inference-based materials. Each session lasted about 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. Drawing scores were determined by a six-point rubric developed by Chi and colleagues (1994; see Appendix 2 for examples of how the rubric was applied to participants’ drawings). 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 pre-draw activity showed a mean score of 2.68 (SD = 1.32; range 1–5 out of 6 maximum). The mean score on the post-draw activity was 4.14 (SD = 1.28, range 1–6), and was a significant gain over pre-draw performance, t(19) = 5.97, p < 0.001, Cohen’s d = 1.12.

Results from the general knowledge pre-test (Table 2) showed a mean of 35.3% correct responses (SD = 25.90; ranging from 3.85% to 79.50%). Participants’ baseline general knowledge was, on average, significantly greater than zero, t(19) = 5.93, p < 0.01. The general knowledge post-test showed a mean of 54.50% (SD = 15.15). Comparing general knowledge pre- and post-test performance yielded evidence of significant gains, t(19) = 6.17, p < 0.001, Cohen’s d = 1.08.

Post-test 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 < 0.001, Cohen’s d = 1.42, as was expected from prior research (e.g., Butcher, 2006).

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 post-test 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, $F(5, 120) = 9.46, p < 0.001$. Post hoc analyses using Fisher’s Least Square Difference (LSD) method indicated that the average gesture rate was significantly higher during the inference questions than the pre-test of general knowledge questions ($p < 0.001$), the general knowledge post-test ($p < 0.001$), and the post-test textbase questions ($p < 0.05$). 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 ($Z = -2.57, p < 0.01$).

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. 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 inference making is an embodied, situated activity that stimulates action planning and motor systems, resulting in more gestures. However, two alternative interpretations must also be considered. One alternative interpretation 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. To test this possibility, GK items were coded for their spatial content based on the presence of location or circulation pathway information. About one-third (9/25) of GK were coded as spatial and two-thirds (16/25) as non-spatial. An analysis comparing gesture rates of spatial and non-spatial GK items showed a marginal but non-significant effect on gesture rates, $t(19) = 2.05, p = 0.054, Cohen’s d = 0.003$. This indicates that the spatial quality of test items is not likely to explain the observed difference in gesture rates.

A second alternative interpretation is that the relative difficulty of the test items determines gesture rate (cf. Sassenberg & van der Meer, 2010). Comparing patterns of performance (Table 2) and gesture production (Table 3) shows that TB items have the highest levels of performance but are intermediate in their accompanying gesture rate; while GK items exhibit the lowest gesture rate but are intermediate in difficulty. Thus, test item difficulty 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 primarily dependent on their spatial content or level of difficulty. Yet the specific nature of the relationship remains unclear. If gesture production were reliably affected by manipulating the quality of the SMs produced by readers, this would provide further evidence for the relationship between gestures and MMs.
Experiment 2

Experiment 2 sharpens our investigation in two ways. First, if the hypothesized relationship holds, then manipulation of qualities of the text that influence readers’ SMs will produce changes in gesture rates. Previous research by Butcher (2006) demonstrated that the inclusion of 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 as compared to those with illustrations. This held even though the text was, by itself, sufficient to support the reasoning needed for the post-test questions. Second, it is still possible 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 illustrations, 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) than other assessment activities.

Method

The primary manipulation is the presence or absence of illustrations previously determined to help in SM formation. The central predictions are that the presence of illustrations will be associated with superior inference making and increased gesture rates.

Participants

Fifty-four 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 about 40 minutes. Because six participants engaged in activities during testing that directly 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 post-test 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 between the two tutorials was the presence or lack of illustrations. Appendix 1 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 post-test overall, $F(3,52) = 78.73$, $p < 0.000$, $MSE = 1.04$, $\eta^2_p = 0.60$, and in both conditions, $p < 0.01$. Group differences at pre-test were not significant, and there was no significant text x time interaction. Post-draw scores also showed no difference, $F < 1$.

Performance on GK questions showed no pre-test differences (Table 5), and significant improvement from pre- to post-test, $F(3,52) = 145.44$, $MSE = 96.19$, $p = 0.000$, $\eta^2_p = 0.74$, 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.

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 item, with performance on textbase items significantly higher than inference items, $F(1,52) = 78.91$, $p = 0.000$, $MSE = 195.54$, $\eta^2_p = 0.6$.

It appears that those who read the non-illustrated text form an 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).

<table>
<thead>
<tr>
<th>Task</th>
<th>Illustrated tutorial</th>
<th>Non-illustrated tutorial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Score</td>
<td>SD</td>
</tr>
<tr>
<td>Pre-draw</td>
<td>2.89</td>
<td>1.42</td>
</tr>
<tr>
<td>Post-draw</td>
<td>4.70</td>
<td>1.51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
<th>% Correct</th>
<th>SD</th>
<th>% Correct</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Test General Knowledge</td>
<td>29.63</td>
<td>24.75</td>
<td>31.29</td>
<td>23.35</td>
</tr>
<tr>
<td>Post-Test General Knowledge</td>
<td>54.14</td>
<td>21.33</td>
<td>52.73</td>
<td>17.66</td>
</tr>
<tr>
<td>Textbase</td>
<td>69.28</td>
<td>24.77</td>
<td>68.08</td>
<td>24.91</td>
</tr>
<tr>
<td>Inference</td>
<td>47.14</td>
<td>24.96</td>
<td>41.69</td>
<td>18.67</td>
</tr>
</tbody>
</table>
Table 6. Mean gesture rates for verbal tasks in Experiment 2 for illustrated and non-illustrated tutorials.

<table>
<thead>
<tr>
<th>Task</th>
<th>Illustrated tutorial</th>
<th>Non-illustrated tutorial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gesture Rate*</td>
<td>SD</td>
</tr>
<tr>
<td>Pre-Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Knowledge</td>
<td>2.98</td>
<td>2.68</td>
</tr>
<tr>
<td>Post-Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Knowledge</td>
<td>4.70</td>
<td>3.09</td>
</tr>
<tr>
<td>Textbase</td>
<td>6.21</td>
<td>3.92</td>
</tr>
<tr>
<td>Inference</td>
<td>7.80</td>
<td>5.82</td>
</tr>
</tbody>
</table>

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

**Gesture rates**

Table 6 shows the gesture rates for each verbal task. Participants in the non-illustrated condition gestured more frequently during the pre-test 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 = 0.002, MSE = 63.48, \( \eta^2_p = 0.17 \), replicating the effect found in Experiment 1, and providing additional support for Hypothesis 1.

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-formed 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 (\( \alpha < 0.05 \)) show that participants who read the non-illustrated text exhibited significantly higher gesture rates during inference questions than those in the illustrated group, p < 0.05 (Figure 3). This surprising finding is used to inform our discussion and the design of Experiment 3.

![Figure 3](image-url)
Discussion
Findings from Experiment 2 support Hypothesis 1 but are contrary to Hypothesis 2 and previous empirical findings on learning from text (Butcher, 2006) showing that the presence of helpful illustrations with the text would lead to higher post-test inference scores. Yet, along with comparable levels of performance we found – 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 test 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.

Experiment 3
One puzzling finding from Experiment 2 is that the 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 methodological differences worth noting. In Butcher’s (2006) original study, participants responded to the verbal tests by writing or typing on a computer, and did so in non-communicative settings. 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 programmes. Furthermore, because of the written response methods used in Butcher’s study, her participants had their hands engaged while producing their answers. By pre-occupying one’s hands, the formation of participant’s SMs may be inadvertently affected. 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.

In the current experiment we investigate the hypothesis that gestures are causally involved in MM creation by examining how manipulation of gesture production affects the quality of readers’ SMs. If, as we predict, gesture production influences SM construction, then participants who are allowed to gesture freely should outperform those who experience gesture restrictions on model-based inference making.

Method
The primary manipulation is the gesture condition, with participants in the gesture control group predicted to perform better on inference test items than participants engaged in spatial tapping.

Participants and materials
Seventy-nine participants were recruited for individual sessions that lasted about 50 minutes. Participants completed the same series of pre- and post-test 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: Simple tapping, spatial tapping, and the gesture (control) condition. Regardless of condition, the study tasks were then administered in the following order: a measure of spatial reasoning, pre-draw task, pre-test of general knowledge, computer-based tutorial (illustrated), post-test of general knowledge, post-test of textbase questions, post-test of inference-based questions, and finally the post-draw task.

Spatial tapping condition. Participants assigned to the spatial tapping condition (n = 26) were free to gesture during the pre-test tasks. While reading the tutorial and responding to the post-test 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 four numerals arranged in a 2 x 2 array, derived from a protocol established by Hegarty et al. (2005; see Appendix 3 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 post-test 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 10 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 post-test, 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 post-test were never the same.

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 specific and changing tapping pattern. During the tutorial and post-test, 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 3). As in the spatial tapping condition, participants were instructed to tap along with the 100 bpm rhythm set by a metronome with a blinking light. Although the tutorial and post-test contained the additional cues 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.

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 post-test but were instructed to continue with the assigned task. They were free to gesture during all of the testing and reading activities.

Measure of spatial ability. Because some participants would be engaging in a secondary spatial task during the post-test 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, 1979), a two-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 as illustrated. Participants were presented with each page one at a time and were given three minutes to complete each worksheet. If participants finished the first page before the three minutes elapsed, they immediately moved to the second page. If they had not completed a worksheet within the allotted three minutes, they were not allowed extra time to complete it.

Coding and scoring

The Paper Folding Task was scored according to its own rubric (Ekstrom, French, Harman, & Derman, 1976). Responses to the pre- and post-test 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. Data were collected on the number of times a participant stopped tapping in order to gesture, but the number was so small that it was dropped from subsequent analyses.

Results

Task performance

Despite the potential interference of the tapping tasks, performance improved from pre- to post-test during the drawing activity for all three conditions, $F(2, 77) = 757.69$, MSE = 3.15, $\eta_p^2 = 0.91$, $p < 0.001$. Planned post-hoc Bonferroni comparisons ($\alpha = 0.05$) showed that no group’s drawings were significantly better than the others.

Figure 4 shows the mean scores for all verbal tests included in Experiment 3. Using participants’ scores 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

![Figure 4](image-url)  
**Figure 4.** Experiment 3 scores on verbal tasks during pre-test and post-tests as a function of gesture condition.
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, $F(4,75) = 6.0$, $MSE = 253.1$, $p < 0.01$. Planned post-hoc Bonferroni comparisons ($\alpha = 0.05$) showed that the control group scored significantly higher than the spatial tapping group on textbase ($p < 0.05$) and inference questions, $p < 0.01$. Additionally, participants in the simple tapping group scored higher on inferences questions than did participants in the spatial tapping group, $p < 0.05$ (see Table 7). 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 engaging those mental resources that mediate action systems that impairs SM formation.

**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 post-test tasks (see Figure 5). The difference between mean

![Figure 5. Gesture rate by condition for verbal post-test tasks in Experiment 3.](image-url)
gesture rates on all tasks between the simple tapping and the control condition was not significant, $F(1,6) = 3.1$, $p > 0.10$ (see Figure 5). Participants’ gesture rates overall in the spatial tapping group were significantly greater than zero for all three post-test tasks (see Figure 5). 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).

Scores on the paper folding test did not reveal a significant effect on gesture rate in any of the tasks, $F(4, 73) = 2.1$, $p > 0.08$, 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 post-test tasks as dependent variables. The results revealed a main effect of condition on gesture rates during the verbal tasks, $F(4, 75) = 10.6$, $MSE = 23.4$, $p < 0.001$. Planned post-hoc Bonferroni comparisons ($\alpha = 0.05$) revealed that participants in the gesture (control) condition gestured more frequently during the post-test of general knowledge than did participants in either the simple tapping ($p < 0.01$) or spatial tapping condition, $p < 0.001$. During textbase questions, participants in the gesture condition gestured more than participants in either gesture-restricted condition ($p < 0.05$ for simple tapping; $p < 0.001$ for spatial tapping), while participants in the simple tapping condition gestured significantly more than those in the spatial tapping group, $p < 0.01$. 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, $p < 0.001$, while participants in the simple tapping group gestured more often than those in the spatial tapping group, $p < 0.01$.

**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.

**General discussion**

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 and 2); gestures appeared to serve a compensatory role in SM formation in the absence of helpful illustrations (Experiment 2); and gesture inhibition had a substantial, negative impact on inference-making performance (Hypothesis 3; Experiment 3).

To qualify these findings, there are several limitations to the current investigation. 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). Finally, the spatial demands inherent in the gesture production system (e.g., Krauss, 1998) were not conclusively disentangled from model based reasoning in these studies, and it will require careful analytic work in the future to explore this dimension.

While it is likely that processes involved in spatial reasoning are part of both model-based reasoning and gesture-based behaviours, they do not seem to provide the dominant account for 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 was lower in the presence of more spatial information (Experiment 2). Test item difficulty presents another complex set of issues when interpreting these findings, as inference making is often more difficult than other types of assessment tasks. Post hoc examination of the results suggests, however, that test item difficulty does not provide the primary account for the pattern of observed findings. Taken together, these results contribute to a view that suggests gesture production serves a central, causal role in the formation of MMs when participants learn from reading a scientific text. Furthermore, comparison between less and more demanding hand tapping activities suggest that it is not merely engaging one’s hands that matters; rather it is engaging those mental resources that mediate action systems that affects SM formation and inference-making processes.

An embodied cognition account of learning with mental models

The GSA framework (Hostetter & Alibali, 2008) reviewed above provides an empirically based account for how cognitive processes that invoke resources involved in motor control can lead to gesture production. Yet, gesture can affect the course along which those cognitive processes proceed (Alibali & Kita, 2010; Goldin-Meadow & Beilock, 2010). 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 build 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 programmes are engaged during the construction and running of these models. To specify the GAME model, two sets of modifications to the GSA architecture are proposed (Figure 6).

For the first set of modifications, we draw on contributions from the MOSAIC system (Haruno, Wolpert, & Kawato, 2001; Wolpert & Kawato, 1998), an architecture for regulating motor control in service of goal directed behaviour. MOSAIC and HMOASAIC, 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, 2012).

Central to the HMOSAIC model is the simultaneous production of multiple, paired predictor–controller modules that each anticipate some plausible next state of the motor system and provide feedforward and feedback signals to regulate muscle behaviour toward the intended goal. There is continuous competition among the predictor–controller modules and the system favours those modules that most closely track the real-time behaviour of the body and the proprioceptive feedback from interactions with the environment, even as adjustments are implemented.

As in HMOSAIC, motor-directed actions in GAME are modelled by parallel predictor–controller modules (Figure 6). In particular, each predictor, \( P_i \), receives a copy of an intended action, 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 predictor \( P_i \) generates future predictions. One key change we propose to the HMOSAIC architecture is that each specific prediction generated by \( P_i \) also includes an \( sm_i \), a potential modification to the currently activated SM in long-term memory based on the predictions from unit \( i \).

The second modification to GSA recognizes that the original GSA framework was designed to account for behaviour during a range of problem solving activities, rather than reading comprehension and learning from text. In Figure 1, we have modified the inputs from the original GSA model in order to distinguish input that is propositional from the imagery input that arises when activating an SM when learning from text (Kintsch, 1998;
van Dijk & Kintsch, 1983). This distinction is especially important when modelling 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
(Farah, 1999; Jeannerod, 1994; Kosslyn & Thompson, 1999), reasoning (Kosslyn, Ball,
& Reiser, 1978), or actually viewing actions (Martin, Ungerleider, & Haxby, 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 predict that processing of SM based information is more likely to engage action
production systems, such as gestures, because of their model building role, in addition to
activation due to spatial content, as proposed in GSA. This prediction was supported by
data from Experiments 1 and 2. Second, we predict that the execution of motor control
programmes during gesture production will influence the ways that SMs are constructed.
Thus we expect to see more extensive model-based reasoning, such as higher rates of
inference making, when gestures are allowed, and lower rates of inferencing when gesture
production is restricted. This prediction received support from Experiment 3.

Action-based accounts of higher order cognition such as language comprehension
have recently received some support. Glenberg and Gallese (2012) 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, favouring those predictions that resemble the
state of things as they unfold, and dispensing with those sm’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 modelled
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 behaviour.

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 alterations – the sm – come from. In fact, Glenberg and Gallese (2012) argue that the core functionality of the HMOSAIC “the predictor corresponds to a mental model (Johnson-Laird, 1989),” as cited in Glenberg & Gallese, 2012, p. 4).

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. There are a number of findings suggesting that the actions themselves play a consequential role and can be effective even when participants unintentionally engage in them. 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 behaviour in accordance with the simulated emotion (Havas, Glenberg, & Rinck, 2007; Niedenthal, 2007). 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). Thomas and 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. Recently, Nathan, Walkington, et al. (2014) showed that directed actions could improve students’ mathematical reasoning during a proof production task, even when participants are unaware that the actions relate to the mathematical conjectures. 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 learning theories**

Having shown the benefits of gestures for inference making, it is reasonable to ask whether the influence of gestures is likely to be limited to settings utilizing learning from text. The literature suggests that it is a more general phenomenon. Body based behaviours, including gestures, serve as a resource for learning by supporting cognitive off-loading of current memory and processing demands (Wilson, 2002) and indexically linking information in the world with words and ideas (Glenberg & Robertson, 1999). Gestures serve a causal role in learning by providing an alternative, embodied way of representing new ideas (Cook, Mitchell, & Goldin-Meadow, 2008). Performing iconically compatible actions fosters learning in areas such as college level physics (Kontra, Goldin-Meadow, & Beilock, 2012) and high school geometry (Nathan, Walkington, et al., 2014). Thus, engaging action systems in certain ways can foster learning across a range of tasks and topics.

**Implications for assessing knowledge**

People naturally express themselves multi-modally, using speech as well as gestures. Assessment methods that focus exclusively on written, verbal information may give a biased account of what students actually know about a topic.

In addition to the added content of one’s responses to test questions, this investigation provides evidence that gestures influence the cognitive processes themselves, whether expressed verbally or through actions. Because of the close similarity of the spatial tapping condition to typing, this finding raises questions about unintended effect that typing may have during computer-based assessments. If typing responses interferes with
the motor control systems in ways that impair SM development, then we may see degraded test performance on tasks such as inference making and other forms of higher order cognition, while showing relatively little effect on subordinate processes such as recognition, repetition, and factual recall.

Conclusions
The spatial, dynamic nature of mental models that people construct when learning through reading seem to be influenced by simulated, situated actions that the text invokes. Participants in the studies reported here gestured more frequently when responding to model-based inferences, used gestures as a body-based resource to support inference making when helpful illustrations were absent, and exhibited inferior inference making when gesture was restricted. The potential causal role of gesture in mental model formation and inference making aligns with an emerging view of the embodied nature of knowledge and cognitive processes, and the reciprocal influences between simulated, situated action (i.e., gestures) and complex cognitive processes offer new insights about the nature of human behaviour and learning.

Acknowledgements
Portions of this research have been presented at the annual meeting of the American Education Research Association (2009, 2010, 2011) the International Conference of the Learning Sciences (2010), and in the unpublished doctoral dissertation of the second author.

References
Hearing gestures: How our hands help us think

M.J. Nathan and C.V.J. Martinez

Downloaded by [University of Wisconsin - Madison] at 19:14 23 April 2015


**Appendix 1**

**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.
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 2

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 from 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 3

Tapping Stimulus used in Experiment 3 for simple and spatial tapping conditions.