Visible embodiment: gestures as simulated action.

2 AUTHORS:

Autumn B. Hostetter
Kalamazoo College
23 PUBLICATIONS **727** CITATIONS

SEE PROFILE

Martha W. Alibali
University of Wisconsin–Madison
101 PUBLICATIONS **3,167** CITATIONS

SEE PROFILE

Visible embodiment: Gestures as simulated action

Autumn B. Hostetter and Martha W. Alibali
University of Wisconsin, Madison, Wisconsin

Spontaneous gestures that accompany speech are related to both verbal and spatial processes. We argue that gestures emerge from perceptual and motor simulations that underlie embodied language and mental imagery. We first review current thinking about embodied cognition, embodied language, and embodied mental imagery. We then provide evidence that gestures stem from spatial representations and mental images. We then propose the gestures-as-simulated-action framework to explain how gestures might arise from an embodied cognitive system. Finally, we compare this framework with other current models of gesture production, and we briefly outline predictions that derive from the framework.

Gestures are often considered to be valid evidence of the embodiment of language and cognition. Since people use their bodies (i.e., gestures) to express knowledge, it is argued, the knowledge itself must be deeply tied to the body (Alibali & Nathan, 2007; Gallagher, 2005; Gibbs, 2006; Hostetter & Alibali, 2004; McNeill, 2005; Nuñez, 2005). This claim may hold some truth, but to date, there have been no explicit proposals about *how* embodied knowledge comes to be reflected in speech-accompanying gestures. The goal of this article is to explore how gestures may arise from embodied thinking; we will approach this goal by integrating research about perception, action, cognition, language, mental imagery, and gesture production.

A variety of hand and body movements can be considered to be gestures, and much previous research has sought to describe these different types of gestures (e.g., Efron, 1972; Kendon, 2004; McNeill, 1992). In this article, we limit our discussion to *representational gestures*—that is, movements that represent the content of speech by pointing to a referent in the physical environment (deictic gestures), depicting a referent with the motion or shape of the hands (iconic gestures), or depicting a concrete referent or indicating a spatial location for an abstract idea (metaphoric gestures). We use the term *gestures* to refer specifically to representational gestures. Other types of gestures, including *beat gestures* (movements that emphasize the prosody or structure of speech without conveying semantic information) and *interactive gestures* (movements used to manage turn taking and other aspects of an interaction between multiple speakers), fall outside the scope of this article.

Our aim is to describe a mechanism that may possibly give rise to gestures. We do not seek to address the issue of the functions of gestures (e.g., in communication, in speech production). Although we review some ideas about the functions of gestures, we do not take a strong stance here regarding gesture function. The goal of our framework is to explain what gestures *are*, rather than what they *do*.

The article proceeds in seven sections. In the first section, we review ideas about links between perception and action and how these links are important in memory and cognition. In the second section, we review evidence that language processing is tied to the body. In the third section, we review evidence that mental imagery is also an embodied process. In the fourth section, we review evidence on the relationship between gesture production and mental imagery, and in the fifth, we propose a theoretical framework that explains how gestures may arise from the activation of perception and action. In the sixth section, we compare our framework with other models of gesture production. Finally, in the seventh section, we articulate predictions that derive from the framework.

Perception, Action, and Embodied Cognition

Many theorists have proposed that perception is for action (J. J. Gibson, 1979; Sperry, 1952); that is, the ability to perceive evolved from a need to interact with the world. Perception enables us to know the affordances of objects in our environment and thus enables us to guide our actions in a goal-directed way (see, e.g., Adolph, 1997). Without this knowledge, it would be virtually impossible to know which object is best for hiding behind, sitting on, climbing up, or any number of other potential actions that were presumably important for survival in the ancestral environment. It is thus difficult to imagine perception as having evolved as anything but an "ongoing process of finding appropriate responses to the environment" (Borghi, 2005, p. 9). Perception is adaptive because it enables quick and effective reactions to the stimuli we encounter.

If perception indeed evolved primarily to facilitate action, one might predict that perceptions should elicit appropriate actions quite automatically. Indeed, there is accumulating evidence that action plans are involuntary consequences of our perceptions. Tucker and Ellis (1998) found that responses that indicated whether a visually perceived object

A. B. Hostetter, abhostetter@wisc.edu

was upright or inverted were significantly faster when they were made with the hand that would be used to grasp the object. For example, participants were faster to respond to a picture of a teapot with the handle on the left side when they responded with the left hand than when they responded with the right. Ellis and Tucker (2000) showed a similar compatibility effect between the type of action response and the type of grip afforded by a previously viewed object. Participants were faster to respond by squeezing a cylinder after viewing objects that would be grasped with a squeezing hand shape (e.g., hammer, bottle, pan) than after viewing objects that would be grasped with a pinching hand shape (e.g., screw, coin, pen). Thus, viewing an object appears to prime the actions associated with grasping that object.

Neural evidence also suggests that perception automatically evokes involuntary action responses. Rizzolatti and colleagues have documented the presence of *mirror neurons* in the monkey (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) and, possibly, in the human brain (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Located in the premotor cortex, mirror neurons are activated both when perceiving another's actions and when producing actions oneself. These neurons are hypothesized to be integral in understanding and imitating the actions of others (Gallese & Goldman, 1998; M. Wilson & Knoblich, 2005). The fact that the very same cells are involved in both action and perception suggests that activating potential actions may be an automatic consequence of perception.

To date, the best evidence for the existence of mirror neurons comes from studies with the monkey brain and from situations in which the monkey is observing actions. However, there are also human data that suggest that premotor and motor areas of the brain are activated when one perceives some visual or auditory stimuli. For example, Grafton, Fadiga, Arbib, and Rizzolatti (1997) found activation of the left ventral premotor cortex when participants observed familiar tools. Such tools have strong associations with possible actions, and this may make activation of premotor or motor areas especially likely. Indeed, objects that can be easily manipulated, such as a shirt or an apple, activate areas of premotor cortex more strongly than do objects that cannot be easily manipulated, such as a traffic light or an animal (Gerlach, Law, & Paulson, 2002). Similarly, in the auditory domain, music perception activates the motor cortex of pianists, who have clear associations between auditory perception and the actions required to produce the music, but it does not activate the motor cortex of individuals who have no associations between the music and potential actions (Haueisen & Knösche, 2001). Thus, at least for stimuli with strong associations with possible actions, perception automatically activates the brain areas involved in producing those corresponding actions.

Both behavioral and neural evidence converge to support the notion that perception leads to automatic planning of actions. From the perspective that perception evolved to facilitate action, this is not surprising. Successful navigation and survival require a perceptual system that is capable of quickly turning veridical representations into possible actions. At the same time, however, a perceptual system that has this capability seems to require a body that can actively move through its environment. As J. J. Gibson (1979) argued, "We must perceive in order to move, but we must also move in order to perceive" (p. 223). Tactile perception requires movement of the arms, hands, and fingers in order to apprehend the perceived object's dimensions, contours, and identity. The same is true for visual perception: Movements of the eyes, head, and neck are crucial for perceiving and understanding the three-dimensional nature of the world. According to one account put forth by O'Regan and Noë (2001), visual perception relies on the governing laws of sensorimotor contingency. What is experienced as visual perception is knowledge of how actions lawfully relate to changes in visual sensations. For example, when our bodies move forward, the retinal pattern we perceive expands outward; in contrast, when our bodies move backward, the retinal pattern contracts inward. Understanding the lawful contingencies between actions and visual sensations is what it means to visually perceive.

The integral role of action in perception has been studied extensively in human infants. When infants learn to crawl, their perceptual skills change in many ways (Campos et al., 2000). For example, when placed at the edge of the *visual cliff* (E. J. Gibson & Walk, 1960), 7-month-olds who are experienced crawlers often refuse to cross, and their heart rates accelerate (a sign of fear) when they are urged to do so. However, babies of the same age who are not yet crawling do not show signs of fear (Campos, Bertenthal, & Kermoian, 1992). The crucial difference seems to be the experience of self-generated locomotion, rather than crawling per se. Precrawling infants who received experience with self-generated locomotion (using a "walker" they pushed with their feet) showed greater heart rate acceleration at the visual cliff than did control infants who had not received comparable experience (Bertenthal, Campos, & Kermoian, 1994). The walker provided infants with a new opportunity for action, and by performing this action, infants gained new information that led them to perceive the visual cliff differently and, therefore, to fear it.

Of course, the influence of action on perception is not limited to development. As one example, manual actions also inform visual perception in adults. Hecht, Vogt, and Prinz (2001) trained participants to make timed arm movements without visual feedback. After training, participants were better at judging the timing of a visually observed movement than were control participants who had not received the movement training. Thus, knowledge that was gained from action enhanced perceptual judgments.

Finally, actions can also inform perception by changing the state of the world being perceived. Kirsh and Maglio (1994) described *epistemic actions*, or actions that are produced not to bring the actor closer to an external goal, but to make perception and computation easier. They observed that many of the movements made by players of the video game Tetris seemed to be produced solely to make perception of the falling blocks and their possible locations at the bottom of the screen easier. By performing actions, players (and perceivers, more generally) are able to change the information in the environment and provide themselves with new perceptual information that makes computation easier.

Thus, it is not only that perception determines possible actions; actions also determine what and how we perceive. Perception and action are intricately linked, each influencing and determining the other in a broader system of coordination. No event can be thought of as being purely perceptual or purely motor. As John Dewey (1896) said in his discussion of reflexes, "The motor response determines the stimulus, just as truly as sensory stimulus determines movement" (p. 4). This constant cycle of perception determining action and action determining perception enables adaptive interaction with the world.

This tight coupling of motor and perceptual processes that is so important for physical interaction with the world may also be important for mental representation of the world. This is a central claim of a current framework in cognitive science known as *embodied cognition* (Gibbs, 2006; Shapiro, 2007). Proponents of the embodied cognition framework make a variety of claims (see M. Wilson, 2002, for a review), one of which is that offline cognition (i.e., cognition that occurs in the absence of relevant environmental input) is perceptually and motorically based. From this perspective, the ability to represent and manipulate information that is not currently perceptually present is accomplished through the activation of sensorimotor processes.

Thinking about a particular concept, for example, involves a perceptual and motor simulation of the properties associated with that concept, even when no exemplar of the concept is present in the current perceptual environment (Barsalou, 1999). Thinking about a chair involves activating (or partially activating) the same neural substrates that are involved in seeing, touching, and interacting with a real chair. Supporting this idea, Pecher, Zeelinberg, and Barsalou (2003) demonstrated that there is a cost involved in switching perceptual modalities in a conceptual judgment task, similar to the cost involved in switching modalities in a perceptual judgment task. Participants were slower to verify that *loud* is a property of a *blender* after having made a different-modality judgment (e.g., *cranberries* are *tart*) than after having made a same-modality judgment (e.g., *leaves* can *rustle*). Thus, perceptual mechanisms are involved in what is often thought to be a purely conceptual task. Solomon and Barsalou (2001) and Zwaan and Yaxley (2003) have provided additional evidence on this point.

According to this view, thought is shaped by the contingencies between perception and possible actions. It is extremely difficult to think about information in a way that contradicts the contingencies between perception and action that exist in the physical environment. Schwartz and Black (1999), for example, found that participants had difficulty running simulations of physical situations in reverse. It is relatively easy to determine when water will pour out of a glass if we imagine the glass tilting from its upright position, but it is much more difficult to make this same judgment when the glass is tilting upward from a horizontal position. This is not to say that we can never think about new concepts or situations; separate concepts can be simulated simultaneously and *meshed* together to create new ideas (Glenberg, 1997). However, the new concept will obey the sensorimotor possibilities of the meshed concepts, making it difficult to mesh concepts that have vastly different affordances (Glenberg & Robertson, 2000).

The idea that cognition is deeply rooted in what is possible, given our bodies, our physical environment, and the relationship between perception and action, is a sea change in cognitive science. Some critics of the embodied approach have argued that it cannot explain all aspects of cognition and memory, and that the field cannot dispense with arbitrary symbols or mental representations entirely (see Markman & Dietrich, 2000; M. Wilson, 2002). Others have suggested that findings about the role of the body in cognition could also be readily explained with a more traditional, symbolic approach (Shapiro, 2007). However, for our purposes, it is not at issue whether an embodied perspective can entirely replace or should rather be integrated with more traditional views of cognition; it is sufficient that accumulating evidence indicates that much of cognition is rooted in the body. Embodiment is at the center of the present discussion because of its relevance to speakers' use of the body in thinking and speaking—namely, in gestures. We believe that an embodied approach will provide promising new insights about gesture production. Before discussing how gestures might arise from an embodied mind, we consider in more detail how theories of embodied cognition apply to two cognitive processes that are related to gesture: language and mental imagery.

Embodiment of Language

The embodied approach to cognition suggests that the meanings of linguistic objects (words, phrases, sentences) are tied to perceptual experience, rather than derived from relationships among abstract, amodal symbols (Barsalou, 1999; Glenberg, 1997). According to such a view, language comes to have meaning because we can index words to the real world; language is grounded in our sensorimotor experience (Glenberg & Robertson, 2000; Zwaan & Madden, 2005). Theories that construe meaning in terms of perceptual experience offer a solution to one of the classic problems in cognitive psychology: the *symbol grounding problem* (Harnad, 1990). The symbol grounding problem arises when abstract symbols are considered only in terms of other abstract symbols; it is difficult for any of the symbols to become meaningful when defined only in terms of one another. Indeed, most current theories of semantic representation, even those that are not explicitly embodied, include connections between conceptual representations and perceptual experiences (e.g., Rogers & McClelland, 2004).

Data from eye-tracking studies suggest that language comprehenders constantly index the words they hear to the real-world objects around them. Participants typically focus their gaze on physically present objects they are hearing about (see Spivey, Richardson, & Gonzales-Marquez, 2005, for a review). It is even more compelling, however, that participants integrate the perceptual and motor affordances of objects into their linguistic representations by using them to predict what will be said and to resolve syntactic ambiguities (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers, Tanenhaus, & Magnuson, 2004). For example, Chambers et al. (2002) instructed participants to put one object from a visual display inside

another object from the display (e.g., "Put the cube inside the can"). As soon as the first object and preposition were mentioned (e.g., "Put the cube inside"), participants constrained their eye movements to objects that would afford the instructed action. In this example, they no longer looked at a can that was too small to put the cube in. It seems, then, that people comprehend language by indexing references in the environment and by considering what is possible given the affordances of the objects involved.

Eye-tracking studies are useful for determining how comprehenders process language that has physically present referents. Much of language, however, is not about objects that are physically present. Proponents of an embodied view of language claim that affordances of mentioned objects are evoked even when the objects themselves are not present. Zwaan and colleagues (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002) have demonstrated that readers create perceptual simulations of the sentences they read. Participants were faster to verify that a pictured object was in a sentence when the orientation of the pictured object matched the orientation implied by the sentence. For example, after reading *John hammered the nail into the floor*, participants were faster to verify that a nail was in the sentence when the pictured nail was pointing downward than when it was lying horizontally (Stanfield & Zwaan, 2001). Thus, people appear to understand sentences by evoking the sensorimotor affordances of objects and events, even when those objects are not physically present.

If comprehenders routinely evoke the motor and perceptual affordances of language, then concurrent motor or perceptual activity should affect their efficiency at doing so. Indeed, several studies have manipulated the relationship between the motor or perceptual state that is implied in a sentence and participants' actual motor or perceptual states while they are comprehending or responding to the sentence (Borghi, Glenberg, & Kaschak, 2004; Glenberg, Havas, Becher, & Rinck, 2005; Glenberg & Kaschak, 2002; Kaschak et al., 2005; Kaschak, Zwaan, Aveyard, & Yaxley, 2006; D. C. Richardson, Spivey, Barsalou, & McRae, 2003). For example, Glenberg and Kaschak asked participants to read sentences that implied either motion away from the body (e.g., *Close the drawer*) or motion toward the body (e.g., *Open the drawer*). Half the time, participants verified that sentences were sensical by making arm movements away from themselves, and half the time they responded by making arm movements toward themselves. The results demonstrated an *action-compatibility effect* between the motion implied in a sentence and the motion of the response: Participants were faster to respond when their responses matched the direction of motion in the sentences.

This behavioral evidence is complemented by neuroimaging studies that compare the brain regions involved in reading about actions with those involved in producing actions (Pulvermüller, 1999, 2005). For example, Hauk, Johnsrude, and Pulvermüller (2004) compared fMRI scans taken while participants moved their tongues, fingers, and feet with scans taken while they passively read words like *lick*, *pick*, and *kick*. The comparison revealed very similar patterns of activation in the motor cortex for movements and corresponding words. Thus, reading words, even with no instructions to actively simulate the meaning of those words, automatically activates the brain areas involved in turning those words into actions.

The evidence discussed so far has involved words and sentences about fairly concrete objects and events. What about language about less tangible concepts? The idea that even abstract concepts are routinely mapped onto our understanding of more concrete, physical concepts (or *image schemas*) is gaining support (Barsalou & Wiemer-Hastings, 2005; Gibbs, 2003, 2006; Johnson, 1987; Lakoff, 1987). Johnson, for example, describes how physical attraction between individuals is often conceptualized as an extension of physical force. Words like *radiant*, *attractive*, *knockout*, and *bombshell* that are used to describe the forces of physical objects are extended to describe feelings about someone's appearance. D. C. Richardson, Spivey, Edelman, and Naples (2001) demonstrated empirically that common spatial images underlie the perceptions of abstract (as well as concrete) verbs. Participants consistently drew sentences like *John respected Mary* as depicting a vertical spatial relationship between John and Mary. In fact, the underlying spatial representations for such abstract verbs appear to be so strong that processing them interferes with subsequent processing of vertically presented images (D. C. Richardson et al., 2003). Along similar lines, Glenberg and Kaschak (2002) also found an action-compatibility effect for abstract sentences in which the motion was metaphorical (e.g., *She told me the story* implies metaphorical movement of information toward the body). Finally, N. L. Wilson and Gibbs (2007) demonstrated that comprehension of metaphorical phrases (e.g., *Grasp a concept*) was facilitated by producing or imagining an appropriate body movement (e.g., grasping) prior to reading the sentence. These findings suggest that the meanings of sentences are simulated in physical terms, even when the sentences do not describe physical movement.

The claim that language is processed by indexing words to real-world referents and simulating their perceptual and motor properties has been supported for language about physically present referents, language about concrete absent referents, and language about abstract concepts. All of this evidence, however, deals specifically with the processes involved in language comprehension; there is virtually no direct evidence about the role of sensorimotor simulations in language production. Nevertheless, the claims of embodied cognition presumably also apply to language production because successful production, much like comprehension, relies heavily on conceptual representations.

In many ways, the processes involved in language production can be thought of as being a reversal of the processes involved in language comprehension. Rather than turning an acoustic or visual signal into a conceptual message (as comprehenders do), producers start with a conceptual message and turn it into an acoustic or written signal (Dell, Chang, & Griffin, 1999; Levelt, 1989). According to theories of embodied cognition (Barsalou, 1999; Glenberg, 1997), concepts are primarily sensorimotor; thus, when speakers activate concepts in order to express meaning, they are presumably activating perceptual and motor

information, just as comprehenders do when they activate meaning from language input. In theory, then, language producers must start with sensorimotor representations of meaning, just as language comprehenders end there. These sensorimotor representations that underlie speaking, we argue, are the bases for speech-accompanying gestures. The type of sensorimotor representation that has received the most attention in the literature on gesture production is mental imagery. We turn now to evidence for the embodiment of mental imagery.

Embodiment of Mental Imagery

A mental image is an analogue representation of a perceptual object or motor event. Mental images can be contrasted with *verbal representations*, which are natural language descriptions, and with *propositional representations*, which are lists of amodal, abstract symbols and predicates that describe how the symbols are related to each other. For example, the picture in Figure 1 could be represented verbally with a sentence like *The circle is on top of the triangle*. It could also be represented propositionally as ON(CIRCLE, TRIANGLE), where the words in capital letters do not represent the words themselves, but rather are symbols for the concepts expressed by those words. Alternatively, it could be represented with a mental image that strongly resembles Figure 1 itself. Rather than encode the picture in propositional or verbal symbols, the mental image retains the spatial and physical features of the picture itself.

Theories of memory and cognition that incorporate mental imagery provide a parsimonious account of many aspects of memory and cognition. For example, in the 1960s, Paivio (1963, 1965) observed that successful verbal recall is strongly associated with how easy a concept is to visualize. This effect appears to be due to how easy it is to form an image of a concept, rather than to other related factors, such as meaningfulness of verbal associations or concreteness (Paivio, 1968). The mnemonic benefits of imagery are twofold. First, images provide an additional code for retrieval and memory, such that highly imageable concepts can be recalled either by remembering their verbal labels or by remembering their visual images (Begg, 1972; Paivio, 1975). Second, images provide a means of integrating separate concepts into one connected entity, making imageability particularly important for remembering pairs of items (Paivio, 1963, 1965). These ideas are articulated in *dual coding theory* (Paivio, 1971, 1991), which explains memory as being the product of both symbolic (e.g., verbal or propositional) and nonsymbolic (e.g., imageable) representations.

The notion that mental imagery is integral to thinking and memory is highly compatible with theories of embodied cognition (Gibbs & Berg, 2002). Because mental images retain the spatial, physical, and kinesthetic properties of the events they represent, they are dependent on the same relationships between perceptual and motor processes that are involved in interacting with physical objects in the environment. Although mental images can correspond to any of the senses, *visual mental imagery* and *motor mental imagery* have the clearest relationships with perception and action. In both cases, imagery utilizes sen-
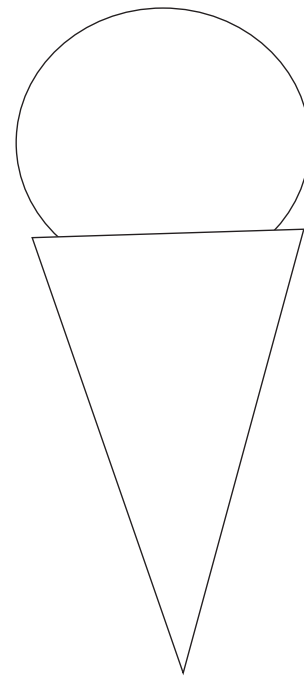

Figure 1. An image that could be mentally represented verbally as "A circle is on top of a triangle," propositionally as "ON(CIRCLE, TRIANGLE)," or imagistically as an image that resembles the figure.

sory and motor processes to enable offline cognition. We now discuss these two types of imagery in turn.

Visual mental imagery is "the experience of viewing a stimulus in the absence of appropriate sensory input" (Kosslyn, 2005, p. 334). Imagining what Figure 1 looks like without looking back at it involves visual mental imagery. The phenomenological experience of visual mental imagery is very similar to vision, and, indeed, the brain areas involved in visual mental imagery overlap with those activated during vision by as much as 90% (Ganis, Thompson, & Kosslyn, 2004). Visual mental images afford many of the same processes afforded by currently viewed images. They can be interpreted in multiple ways (Mast & Kosslyn, 2002; Peterson, Kihlstrom, Rose, & Glisky, 1992) and decomposed into smaller pieces (Helstrup & Andersen, 1991). Attention can be shifted to different places in a visual mental image, similar to the way attention is shifted in a visual scene by moving the eyes. The amount of time needed to shift attention to an object in an imagined scene is strongly correlated with the distance that must be scanned to find the object in a viewed scene (Kosslyn, 1973). This isomorphism is present regardless of whether the visual mental image is a re-creation of a viewed image or a construction of a layout that was verbally described in text (Denis, Goncalves, & Memmi, 1995).

As a close analogue to visual perception, visual mental imagery is shaped by action. Just as visual perceptions are linked to the actions they afford—as well as to the actions that inform them—visual mental images are similarly tied to simulated movement. The role of motor processes in visual mental imagery is especially apparent when we consider imagery processes that require a large

amount of cognitive control, such as *image transformation*. Image transformation is perhaps best exemplified by the classic Shepard and Metzler (1971) mental rotation task (see also Shepard & Cooper, 1982). In this task, participants are asked to make a judgment about whether two three-dimensional shapes presented at different rotational orientations are the same shape or mirror images. The typical strategy for solving the task is to imagine rotating one shape until it is in the same orientation as the second shape. Once the two shapes are imagined at the same orientation, a direct comparison can be made.

Evidence suggests that mental rotation relies on processes analogous to those involved in physical rotation (Cooper, 1976; Freyd & Finke, 1984; Shepard & Cooper, 1982; Shepard & Metzler, 1971). The most compelling evidence is that mental and physical rotation interfere with one another; judgments about mental rotations are influenced by concurrent physical rotation (Wexler, Kosslyn, & Berthoz, 1998; Wohlschläger & Wohlschläger, 1998). Wexler et al. asked participants to complete a mental rotation task while simultaneously rotating a joystick in a direction either congruent or incongruent to the direction of mental rotation. Mental rotation was slower when it conflicted with the direction of physical rotation and faster when it matched the direction of physical rotation. Furthermore, when asked to physically rotate the joystick at a slower rate, participants performed mental rotation more slowly as well. Such interference suggests that physical and mental rotation rely, at least to some extent, on overlapping resources.

It seems, then, that when thinkers use mental images to represent aspects of the environment, they are using sensory and perceptual systems. Furthermore, when thinkers make decisions about mental visual images, they often evoke simulations of motion that allow them to "see" how the image "looks" from different perspectives. This is analogous to the way perceivers sometimes use actions to change their current perceptual input (see, e.g., Kirsh & Maglio, 1994).

Sometimes the goal of mental imagery is not to understand how something would look at a different orientation, but rather to understand what actions need to be taken to meet a particular perceptual goal. This type of mental imagery, motor mental imagery, is thought to consist of representations of our bodies in motion (Jeannerod, 1994, 1995). That is, rather than imagining a mental visual image moving (as is the case in the mental transformations described above), motor imagery "requires a representation of the body as the generator of acting forces, and not only of the effects of these forces on the external world" (Jeannerod, 1995, p. 1420). Imagining how to tie a shoe or how to land a perfect triple axel (if one were a figure skater) requires motor imagery.

Motor imagery appears to rely on the same neural mechanisms that are involved in movement. Brain areas that are involved in movement, such as the cerebellum, basal ganglia, and supplementary motor areas, are also activated when participants are asked to imagine moving (see Jeannerod, 2001, for a review). Additionally, the amount of time it takes to imagine performing a move-

ment and the amount of time it takes to actually perform the movement are strongly correlated (Decety, Jeannerod, & Prablanc, 1989). Similarly, imagined motor actions are performed more slowly when they are more difficult, as is true of physical movements (Decety & Jeannerod, 1995). This isomorphism between physical and imagined movements has led to the claim that motor imagery involves premotor preparation and planning.

There are two similar but distinct views about the basis of motor imagery. According to *simulation theory*, motor images occur when a premotor plan is prepared but not executed (Jeannerod, 1994). Alternatively, *emulation theory* (Grush, 2004) posits that every motor plan is run both by the body and by a mental *emulator* that mimics the movements and their effects on the musculoskeletal system and environment. According to emulation theory, motor imagery occurs when the emulator has planned and executed a movement that is not also planned and executed by the body. In contrast, according to simulation theory, the motor plan has not been executed by any system, physical or mental. These two theories are difficult to distinguish empirically, and for our purposes, their similarities are more important than their differences. Both claim that motor images are directly analogous to physical movements.

The view that motor images are analogues of physical actions is parallel to the view, discussed above, that visual images are analogues of visual perceptions. As is the case with visual images, motor images are offline representations that utilize perceptual and motor systems. In summary, then, mental imagery (both visual and motor) is an embodied process that relies on simulation of perceptions and actions. Just as we saw in the discussion of language comprehension, mental imagery is possible because we can mentally re-create our experiences with the world. We argue that the use of sensorimotor systems during both language production and mental imagery gives rise to gestures.

Thus far, we have reviewed evidence that both language processing and mental imagery rely on sensorimotor processes. We next consider research linking gesture and sensorimotor processes—specifically, mental imagery.

Gestures and Mental Imagery

People often use their bodies to gesture when they describe their mental images. Gestures are particularly good at expressing spatial and motor information (Alibali, 2005). For example, a speaker who is giving directions may say, "Take a left after the light," while moving his hand to the left. This movement utilizes space and motion to express the spatial concept *left*, and reflects the property of leftness that is presumably present in the speaker's mental image. Rather than transferring spatial or motor properties into verbal or propositional codes, gestures express these properties directly.

McNeill has argued that, unlike language, gestures are both *global* and *synthetic* (McNeill, 1992; McNeill & Duncan, 2000). They are global in the sense that their meaning is not determined from the sum of their parts in the way that the meaning of a sentence is derived from the sum of its words. Instead, the meaning of a gesture is determined from its form as a whole. The individual

parts of the gesture (e.g., hand shape, trajectory, placement) can be interpreted only within this global meaning. A particular hand shape, for example, does not convey a specific meaning common to all gestures in which that hand shape is used.

Gestures are synthetic in the sense that they do not rely on analytic rules (such as word order and syntax) to combine meanings. Instead, gestures can synthesize several meanings into single symbols. For instance, consider a speaker who says "She climbed up the ladder" while wiggling his fingers as his hand moves vertically up in front of him. In this example, the speaker is combining three meanings (she + climbed + up) in a single gesture.

The global and synthetic properties of gestures are similar to the global and synthetic properties of images (both physical and mental). Images, like gestures, convey meaning globally, such that the entire image's meaning influences the interpretation of each part. Figure 1, for example, may be interpreted as being an ice cream cone. Once this global interpretation has been made, we interpret the triangle as being a cone and the circle as being a scoop of ice cream, even though in another context these shapes could represent something completely different. Just as a particular hand shape does not always carry the same meaning, a particular shape (e.g., a triangle or a circle) does not always represent the same object in an image. Similarly, images are synthetic because they can integrate several meanings. *She climbed up the ladder*, for example, could be drawn or visualized as a single image, just as it could be expressed in a single gesture.

The isomorphism between images and gestures is manifested in their frequent co-occurrence. Gestures often occur when speakers express information that evokes images. Beattie and Shovelton (2002) found that judges rated written clauses that had originally been accompanied by gestures to be more evocative of images than they rated clauses that had not been accompanied by gestures. Similarly, gestures occur more often with speech about spatial information than with speech about nonspatial information (Alibali, Heath, & Myers, 2001; Krauss, 1998; Rauscher, Krauss, & Chen, 1996). For example, Zhang (as cited in Krauss, 1998) asked speakers to define words that varied in the extent to which they were concrete, active, or spatial. Participants spent the greatest proportion of time gesturing while defining words that were highly spatial, and this effect was independent of the other two dimensions.

Gestures are a natural way to convey spatial information; but do gestures actually stem from spatial and imagistic representations? Or are they the result of verbal and propositional representations that are expressed in a spatial medium, perhaps for communicative purposes? Wagner, Nusbaum, and Goldin-Meadow (2004) investigated this issue by using a task they viewed not to be inherently spatial; they asked participants to explain their solutions to factoring problems, such as $x^2 + 5x + 6 = (x + 2)(x + 3)$. Participants were required to remember either spatial or verbal information during their explanations, and for some of the explanations, they were prohibited from gesturing. Wagner et al. reasoned that gestures produced during the explanation task must arise either from the propositional representations involved in the task or from spatial representations created specifically for the purpose of gesturing. If the latter, then gesturing during the explanation task should compete with resources in spatial working memory; therefore, participants should have poorer memory for spatial information when they gesture. Instead, gesturing during the explanation task facilitated memory for both verbal and spatial information. Wagner et al. interpreted this evidence to suggest that gestures do not compete for spatial resources and can thus be produced directly from propositional representations.

However, an alternative explanation for these data should be considered. It seems likely that at least some of the representations involved in the factoring task were, in fact, spatial. Speakers often think about adding, subtracting, and grouping numbers as being similar both to grouping and to separating physical quantities—processes that clearly have spatial components (Lakoff & Nuñez, 2001). Furthermore, participants' problem representations may have included spatial information about their solution processes, such as which numbers they had written and in which positions. If some of the representations involved in the explanation task were spatial to begin with, it is possible that expressing these spatial representations in gestures may have alleviated demands on spatial resources, rather than increasing them, as Wagner et al. (2004) hypothesized. That is, rather than being held in working memory, these spatial representations may have been off-loaded to gesture, thus allowing more resources to be devoted to the spatial memory task. According to this view, participants remembered more unrelated spatial information when they had gestured than when they had not gestured, because expressing spatial information in gesture is less resource-intensive than holding it in working memory. Viewed in this way, the Wagner et al. data are not definitive evidence that the representations underlying gestures are not spatial.

Wagner et al. (2004) notwithstanding, abundant evidence suggests that gestures do, in fact, stem from spatial representations. For example, Hostetter and Hopkins (2002) found that speakers gestured more while retelling a cartoon story if they had watched the cartoon than if they had read a verbal description of it. Speakers who had watched the cartoon presumably based their retellings on richer spatial representations than did speakers who had read the verbal description. Further evidence comes from studies of individual differences. Individuals who have weak spatial visualization skills gesture less than do individuals with stronger spatial skills (Hostetter & Alibali, 2007). Similarly, stroke patients who have visuospatial deficits gesture less than do age-matched controls (Hadar, Burstein, Krauss, & Soroker, 1998). This evidence converges to suggest that gestures derive, at least in part, from spatial or imagistic representations.

The evidence reviewed thus far suggests that gestures reflect and stem from speakers' mental images, and it may be tempting to conclude that gestures are simply an epiphenomenon of active mental images. However, increasing evidence suggests that gestures actually facilitate speech about mental images. There are several theories about how this facilitation may occur.

One theory, the *image maintenance theory*, states that gestures activate images in working memory (de Ruiter, 1998; Wesp, Hesse, Keutmann, & Wheaton, 2001). According to this view, the images activated during prelinguistic thinking will quickly decay if not continuously updated, and gestures are a means of reactivating visuospatial information. In support of this view, Wesp et al. found that speakers gestured more than twice as often when describing a painting from memory than they did when describing it while it was visually present.

Other theories suggest a more specific role for gestures in translating mental images into verbal output. The *lexical access hypothesis* (Krauss, Chen, & Chawla, 1996; Krauss, Chen, & Gottesman, 2000) posits that gestures serve as cross-modal primes that facilitate the retrieval of lexical items for spatial and motor ideas. Two types of evidence support this idea. First, gesture rates increase when lexical access is difficult, such as when stimuli are difficult to name (Morsella & Krauss, 2004) or when names have not been rehearsed (Chawla & Krauss, 1994). Second, prohibiting gestures makes speech less fluent (see, e.g., Rauscher et al., 1996). Along these lines, Frick-Horbury and Guttentag (1998) found that restricting participants from gesturing while they were in tip-of-the-tongue states increased retrieval failures (but see also Beattie & Coughlan, 1998, for conflicting findings). It seems, then, that speakers often gesture when they are having trouble accessing words, and that gesturing can reduce retrieval failures, at least in some situations.

In addition to its benefits for lexical access, the highlighting of features of images through gestures may also have benefits for grammatical organization. When translating an image into a verbal stream, speakers must choose which aspects of the image to convey and in what order. The *information packaging hypothesis* (Kita, 2000) suggests that gestures help to break images down into smaller chunks that are suitable for verbalization. Support for this hypothesis has come from studies that have increased the difficulty of conceptualization for speaking without also increasing the difficulty of lexical access (Alibali, Kita, & Young, 2000; Hostetter, Alibali, & Kita, 2007; Melinger & Kita, 2007). For example, Hostetter et al. (2007) found that participants gestured more when describing dot patterns that had to be parsed into separate geometric shapes than when describing the same dot patterns that had already been broken down into discrete shapes.

Finally, highlighting mental images through gestures may also benefit speaking because gestures can bypass the need to verbally encode imagistic ideas. Although most gestures occur with speech that conveys related semantic content, gestures sometimes stand alone or provide richer information than does the co-occurring speech (e.g., Alibali, Bassok, Solomon, Syc, & Goldin-Meadow, 1999; Church & Goldin-Meadow, 1986). For example, Melinger and Levelt (2004) instructed speakers to convey particular information about objects (e.g., size, shape) and found that speakers sometimes conveyed the target information in gestures and not speech.

Speakers may be more likely to convey imagistic information uniquely in gestures when images are difficult to encode verbally. In one illustration, McNeill and Duncan (2000) described how speakers of verb-framed languages, such as Spanish, express manner of motion. In verb-framed languages, manner is not expressed in the main verb as is typical in satellite-framed languages, such as English (e.g., *roll*, *spin*, *twirl*). Instead, Spanish speakers typically encode manner in a separate phrase following a verb that conveys path (e.g., *he descended by rolling*). Because of this difference in linguistic structure, expressing manner verbally is more difficult in Spanish than in English, and indeed, Spanish speakers often omit manner information from their verbal descriptions. When manner is a particularly important aspect of a story, however, Spanish speakers often convey manner through their gestures. For example, a Spanish speaker might say the equivalent of *He descended* while making a repetitive, circular gesture. In this case, the speaker uses gesture to convey an important aspect of the mental image (rolling) that is difficult to encode linguistically.

It seems, then, that gesture and speech sometimes work in collaboration to fully express or, perhaps, even to form ideas. This perspective has been described in detail in *growth point theory* (McNeill, 1992, 2005). According to this view, each new idea in a discourse is a *growth point* that can be fully understood only by considering speech, gesture, and the discourse context. Rather than focusing on how gestures facilitate the translation of mental imagery into verbal output (as the theories discussed above do), growth point theory focuses on how gestures reflect the imagistic components of speech and how they are shaped by the discourse context at the moment of speaking.

All of the theories discussed thus far are broadly compatible with embodied views of cognition. That is, speaking involves activation of mental images—images that rely on simulations of perception and action. Gestures are outward manifestations of these simulations. Expressing simulations through gestures may be a natural by-product of our thoughts, as growth point theory contends. Doing so may also reinforce our mental images and help maintain them in memory, as image maintenance theory contends. Finally, expressing images in gestures may help speakers identify which features of an image to mention, as the information packaging hypothesis contends, or which word is most appropriate for describing a given feature, as the lexical access hypothesis contends. In the next section, we outline a framework for how embodied simulations of perception and action may come to be expressed in gestures.

Gestures as Simulated Action

Figure 2 presents the gesture-as-simulated-action (GSA) framework, which asserts that gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery. In line with embodied theories of cognition, the link between perception and action is at the center of the system. The circular arrows represent their influence on each other; perception determines potential action, just as action determines what is perceived. This relationship is crucial for online interaction with the world. According to the embodied framework, however, the relationship between perception and action is also crucial for offline thinking about the

world. We understand text, create mental images, and plan movements by utilizing the same perception and action mechanisms that we use for interacting with the world. Thus, language processing and mental imagery (depicted at the bottom of Figure 2) are accomplished via simulations of perception and action—simulations that activate or reinstantiate perception and action states.

Simulating actions involves activating neural areas that are involved in planning physical actions (Jeannerod, 2001). Simulating perceptions involves activating neural areas that are involved in perceiving physical objects (Kosslyn, 2005), and, as a result, also involves activating neural areas that are involved in using or reacting to objects (Gerlach et al., 2002; Grafton et al., 1997). Oftentimes, this activation occurs covertly and is inhibited from realization as an overt motor plan. Other times, however, this activation is realized as motor output. Mirror neurons, discussed previously as being neural links between perception and action (e.g., Rizzolatti et al., 1996), work in a similar way. Located in the premotor cortex, mirror neurons fire both when an action is produced and when an action is observed. Activation of the neurons leads to overt movement, however, only when the activation spreads from the premotor cortex to the motor cortex during action production. The GSA framework proposes a similar mechanism. Simulation involves activating premotor[1] action states; this activation has the potential to spread to motor areas and to be realized as overt action. When this spreading activation occurs, a gesture is born.
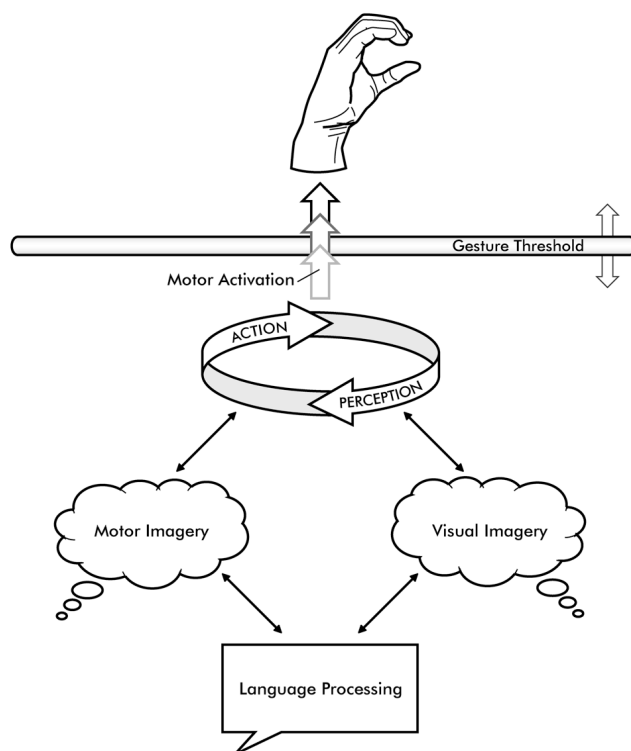

Figure 2. A schematic depiction of the gesture-as-simulated-action framework, which specifies how an embodied cognitive system gives rise to gesture production during mental imagery and language production.

The GSA framework proposes that three factors contribute to whether the activation involved in simulation will be realized as an overt movement, such as a gesture: (1) the strength of activation of the simulated action; (2) the height of the speaker's current *gesture threshold* (described below); and (3) the simultaneous engagement of the motor system for speaking.

The first factor is the strength of activation of the simulated action. This activation must be sufficiently strong to spread from premotor to motor areas and to surpass the *gesture threshold*—the level of activation beyond which a speaker cannot inhibit the expression of simulated actions as gestures. Although the stance we take here is that language is based on simulated action, we recognize that this may not be true for all language produced in all circumstances. In some cases, speakers may produce speech based on verbal or propositional codes, and indeed, it may sometimes be advantageous to do so (e.g., when giving a speech or repeating a story in which it is desirable to reproduce precise words).

Some speakers may simply prefer to utilize verbal and propositional codes whenever possible, and these individuals may tend to spontaneously encode spatial information in a verbal or propositional form. Imagine a person who remembers her route to the grocery store as *a left on Washington Street, followed by a right on Milwaukee Avenue*. This strategy differs from that used by someone who remembers the same route in terms of the visual images seen along the way—the buildings marking the intersections, and the distances between turns. Both people could describe their routes accurately; however, the first might not simulate the experience of driving to the grocery store unless pressed for more details than the verbal representation encodes. In contrast, the second person would likely simulate the experience upon being asked for directions, and would extract the crucial pieces of information about turn directions and street names from the simulation. Although most people probably utilize a combination of these two strategies in everyday life, we suggest that individuals who typically rely on simulations of perception and action to encode and retrieve information should be especially likely to gesture while speaking. Indeed, Hostetter and Alibali (2007) have provided some evidence for this hypothesis, showing that speakers with stronger spatial skills tend to gesture more than speakers with weaker spatial skills.

In addition to whether a simulation underlies speech, the extent to which that simulation involves action is also important. Motor imagery necessarily involves simulated action; it is, by definition, the process of imagining the body in motion. Visual imagery, on the other hand, may primarily involve simulated perception. However, simulated action may be evoked when the task at hand involves imagining how a perceptual image would change over time or would look from another perspective, or when motion is part of the original perception being simulated. Additionally, visual imagery may involve simulated action when the simulated perception or its features are closely tied to action. For instance, a visual image of a pencil likely involves simulation of the actions typically performed with a pencil. Furthermore, in some instances, a visual image may be constructed

in order to extract relevant information about a perceptual feature of the simulated object, such as size, shape, or color. Some perceptual features are strongly related to action because perception of the feature depends on action (e.g., eye movements, tactile exploration) or because the feature determines the actions that are afforded by the object (e.g., grasping, carrying). For example, thinking about the size or shape of a particular cake may involve simulating a movement around the imagined cake's circumference or simulating the way in which the cake would be held, given its size and shape. Conversely, thinking about the color or smell of a particular cake may not involve simulated action because perception of these features is less strongly tied to action.

Whether a speaker utilizes motor or visual imagery may influence the viewpoint the speaker expresses in gesture. *Character-viewpoint* gestures are produced as though the speaker's body is the body of the character being described. For example, McNeill (1992) tells of a speaker who described Sylvester the cat climbing up a drainpipe by moving her hands up and down, as though she herself were climbing the drainpipe (p. 119). We propose that character-viewpoint gestures are produced as the result of simulated motor imagery, in which the speaker simulates the actions being described as though he or she is performing the actions him- or herself. In contrast, *observer-viewpoint* gestures depict characters or scenes as though the speaker is watching them as an outside observer. For example, a speaker might describe Sylvester the cat climbing up the drainpipe by moving his or her hand upward—not as though pulling his or her body upward, but as though the hand is Sylvester moving up the drainpipe (McNeill, 1992, p. 119). We propose that observer-viewpoint gestures result from simulated visual imagery. Instead of simulating the actions of the character, a speaker who produces an observer-viewpoint gesture simulates the motion involved in the visual image that he or she is describing. Whether a speaker simulates actions using motor imagery or animated visual imagery may depend on a number of factors, including the speaker's age and whether the speaker is describing his or her own actions.

The strength of simulated action may also depend on the extent to which the thoughts underlying speech concern concrete objects and events. When speaking about abstract ideas, speakers may rely on verbal and propositional representations more than on imagistic ones. However, as described above, there is evidence that our understanding of many abstract ideas is grounded in our understanding of the physical world (Gibbs, 2006; Johnson, 1987; Lakoff & Johnson, 1980). Basic image schemas (e.g., CONTAINER, OBJECT, ATTRACTION, BALANCE) about the characteristics of physical entities become metaphors for thinking about abstract concepts. For example, we think about ideas as objects that can behave in all of the ways that objects behave. They can be perceived (*I see what you're saying*), moved around in space (*Let's put that idea aside for a minute*), and balanced (*The report balanced both sides of the argument*).

The GSA framework contends that whenever ideas are being simulated in terms of perceptual and motor prop-

erties, regardless of whether the idea is about something physically spatial or only metaphorically spatial, a gesture can result. In McNeill's (1992) typology, gestures that do not directly or literally represent a corresponding word or phrase have their own category: *metaphoric gestures*. We contend that metaphoric gestures arise from perceptual and motor simulations of spatial image schemas on which metaphors are based. For example, a speaker who talks about *fairness* while moving her hands, palms up, up and down in an alternating fashion is simulating two opposing views as though they are objects being balanced on a scale. Similarly, a speaker who denotes the space on his left to refer to "the good guys" and the space on his right to refer to "the bad guys" may be simulating the abstract ideas of good and evil as though they are physical objects that can be located in space. Opposite concepts are particularly likely to be separated in space (one on the left, one on the right) because we have physical experience in separating and sorting objects according to their kind. These examples refer to spatial schemas that are associated with abstract ideas as objects. Along similar lines, Nuñez and Sweetser (2006) have provided a compelling analysis of how the spatial schemas associated with time can also be expressed in metaphoric gestures.

Although abstract ideas are often understood in terms of image schemas and associated action simulations, this is not to say that every abstract idea is always understood in this way. We recognize that some ideas are unlikely to be simulated in terms of action at all. For example, words like *lovely* and *beautiful* bear no clear relation to the metaphor of physical forces and attraction that underlie many other words about aesthetic attraction. Such words may involve a simulation of an emotional or static perceptual state (see Niedenthal, 2007; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005), but according to the GSA framework, unless the simulation involves action, a gesture is unlikely. Thus, the GSA framework cannot account for gestures that accompany such words. We believe that such gestures rarely occur, and we speculate that those that do are based on perceptual simulations of the objects to which the adjectives refer (e.g., a sweeping hand motion at the height of someone's face while he or she says "she looked beautiful").

Although our perspective is that gestures are a natural expression of the simulated actions that underlie speaking and thinking, speakers do not gesture in every speaking situation, or even in every speaking situation that clearly involves simulated action. The GSA framework holds that speakers can inhibit whether they express their simulated actions as gestures (the gesture threshold). A speaker's gesture threshold depends on the speaker's unique neurology, experiences, and beliefs, and aspects of the speaking situation. Regardless of how strongly action is activated during simulation, a gesture will not be produced unless the level of activation surpasses the speaker's current gesture threshold.

Here we consider three determinants of a speaker's gesture threshold, although there are surely others. First, a speaker's gesture threshold may depend on neural factors, such as the strength of connections between premotor and

motor areas. If connections are very strong, even small amounts of activation in premotor areas will spread to motor areas; if connections are weaker, greater activation may be needed. The strength of an individual's connections may be the result of genetics, experience, or individual differences in myelination or neural development.

The threshold may also depend on cognitive factors, such as the cognitive system's current level of effort. Each individual has a limited amount of cognitive resources to expend at any given time. We propose that inhibiting activation from spreading to a gesture requires more cognitive resources than does producing the gesture. Indeed, Goldin-Meadow, Nusbaum, Kelly, and Wagner (2001) found that participants who inhibited gesture while explaining a math task had fewer cognitive resources to devote to another task (i.e., remembering a list of words). This was true both when the experimenter imposed the gesture inhibition and when participants inhibited gestures on their own. Thus, it seems that maintaining a high gesture threshold (or inhibiting simulated action from spreading to gesture) requires cognitive effort. When the cognitive system is exerting effort on other fronts, the effort of maintaining a high threshold may be abandoned or lessened so those resources can be devoted to the current task. As would be expected, speakers gesture more when a verbal description is more challenging to plan or produce (e.g., Alibali et al., 2000; Chawla & Krauss, 1994; Hostetter et al., 2007; Melinger & Kita, 2007; Morsella & Krauss, 2004). In addition, in line with theories that gestures facilitate speech production (described in the previous section), lowering one's threshold in order to produce a gesture may also have benefits for the speech production process itself.

Finally, the gesture threshold may also depend on aspects of the social communicative situation. There is some evidence that speakers can alter their gesture production at will. For example, teachers can increase or decrease the number of gestures they produce during instruction when asked to do so (Hostetter, Bieda, Alibali, Nathan, & Knuth, 2006). The gesture threshold, then, may be thought of as being something that speakers can raise or lower depending on how beneficial they feel gesture would be to their communicative goals and to the listener's needs. Speakers may lower their threshold if they believe that their listeners may have difficulty understanding what is being said. In support of this idea, Alibali and Nathan (2007) found that a classroom teacher was more likely to gesture when she was explaining information that was particularly difficult, such as a new type of problem or in response to a student's question. On the other hand, speakers may also raise their threshold when they believe gesturing to be impolite or unprofessional, or when they do not wish to communicate effectively.

According to the GSA framework, then, a speaker's propensity to gesture is the product of (1) the amount of simulated action underlying his or her current thinking and (2) his or her current threshold or resistance to allowing this simulated action to be transferred into an overt motor plan. Neither of these factors, however, can fully explain why it is that the overwhelming majority of gestures are produced during speech production. It is rare (although certainly not unheard of) for one to gesture while thinking in the absence of speech. The GSA framework thus proposes a third factor that is important to whether simulated action is likely to be realized in gesture. This third factor is whether the motor system is simultaneously engaged in the motorically complex task of speech production.

Manipulating the vocal apparatus in the interest of producing speech is a motorically challenging task, since very specific movements must be made very quickly. To engage the motor system in such a complex task, it is important for activation to move quickly from premotor plans to motor production. As a message is conceptualized for speaking, premotor areas are activated, not only in planning the articulatory movements necessary to pronounce the appropriate words, but also in simulating the sensorimotor aspects of the ideas that are to be expressed. The mechanism that allows activation to quickly spread from premotor to motor areas in the process of speech production may make it difficult to inhibit other concurrent activity in premotor areas from spreading to motor areas. As an analogy, we might imagine activation spreading from premotor areas to motor areas through a gate. Once the gate is opened to allow motor activation for one task (speaking), it may be difficult to inhibit other premotor activation (that which supports gestures) from also spreading through the gate to motor areas. When the activation for the articulatory plan is sent to the motor cortex, the activation for the simulation "rides along" and may be manifested as a gesture. Speakers who do not wish to gesture must set their gesture threshold high enough to counteract this influx of motor activation—a practice that may be quite resource intensive.

The idea that the premotor areas are activated both for simulation and for articulatory plans suggests a reason for why speech-accompanying gestures are so often produced manually. Not all simulations involve a representation of the hands, but the hands are by far the most typical vehicles for gesture by adult speakers. The preference for superimposing simulations onto manual actions may derive from the close, dynamic coupling between oral and manual actions.

Iverson and Thelen (2000) described the hand and mouth as being *coupled oscillators*, such that vocal and manual behavior are mutually entrained through development in a dynamic system. Representations of the hand and mouth are adjacent in both the premotor and motor cortex, and from birth, infants demonstrate good coordination between manual and oral movements. For example, infants are quite good at timing manual and oral actions to put their hands (or objects grasped in their hands) into their mouths. During babbling, babies often move their hands in a rhythmical pattern very similar to the rhythm of babbling (Ejiri, 1998). Iverson and Thelen argued that this co-occurrence is evidence that the hands and the mouth influence each other in order to find a common pattern or rhythm. As the child begins using words and gestures to communicate, activity of the mouth and hands becomes further coordinated, and eventually, activation of the mouth for speaking automatically evokes activation of the hands (see also Kelly et al., 2002).

Hand and mouth may be closely coordinated because they are actually controlled by the same motor system.

Gentilucci and Dalla Volta (in press) provide evidence that syllable pronunciation is affected by simultaneous observations or executions of actions. The specifics of the relationship suggest that movement of the hands encourages similar movement of the tongue and oral articulators; manual grasping encourages preparation for oral grasping, which influences the aperture of the mouth and the pronunciation of a simultaneously produced word or syllable. Although it is doubtful that such isomorphism exists between every manual and oral action, these data suggest that the hands and the mouth are controlled by the same motor system. This suggestion lends plausibility to the idea that speech production automatically encourages the simultaneous production of manual movements.

Our claim, then, is that representational gestures are manifestations of both simulated action and concurrent activation of the oral–manual system. Representational gestures are produced when the simulated action involved in conceptualizing a message is superimposed on the activation of the hands that is coupled with oral activation during speech production (see Tuite, 1993, for a related claim).[2] This superimposition may require some amount of coordination, which may develop over time and with experience. Indeed, Butcher and Goldin-Meadow (2000) observed that gesture and speech become well integrated semantically and temporally in children's speech just before the onset of two-word speech.

As discussed so far, the GSA framework can account for many documented characteristics of speech-accompanying gestures, such as their co-occurrence with imagistic speech, their sensitivity to the cognitive and social situation, and their realization by the hands during speech production. We have not yet addressed the issue of discourse context, another documented factor that influences gesture and one that is at the center of McNeill's (1992, 2005) growth point theory. McNeill (2005) has provided extensive evidence that speakers are particularly likely to gesture about information that is particularly newsworthy or crucial to the ongoing discourse—information that is high in *communicative dynamism*, in his terms. In support of this idea, growing evidence indicates that speakers' gestures are influenced by what their listeners do and do not know (Gerwing & Bavelas, 2004; Holler & Stevens, 2007). According to McNeill (2005), the discourse context sets up a *field of oppositions*, and speakers are especially likely to gesture about information that contrasts with information already present in the field. For example, consider a speaker who has just described a scene from a Tweety and Sylvester cartoon in which Sylvester climbs up the *outside* of a drainpipe. In a later scene, Sylvester climbs up *inside* the drainpipe. The speaker will be especially likely to produce a gesture that represents the idea of interiority, because this idea contrasts with the previously mentioned approach of going up the outside of the drainpipe (McNeill, 2005, pp. 108–112).

According to the GSA framework, the discourse context could influence a speaker's gestures through any of three mechanisms. First, the discourse context could affect how strongly the speaker activates a particular aspect of a simulation (such as interiority in the example above). When speakers are not thinking about how two events are different (perhaps because they forgot that one of the events had happened at all), they are less likely to simulate the specific aspect of the event that would differentiate it from the previous event, and they are thus less likely to convey the distinction in gesture. However, when speakers contrast two events, they are more likely to strongly simulate the contrasting elements; thus they are more likely to express these contrasting elements in gesture. Note that this is a prediction about the form a particular gesture will take, as well as a prediction about its probability of occurring. A gesture that occurs in the absence of contextual differentiation may take any number of forms, depending on the nature of the speaker's simulation. However, a gesture that occurs in order to highlight a contextual differentiation will be particularly likely to take a form that captures the relevant difference, because this difference will be central to the speaker's simulation at that moment.

Second, the discourse context could influence a speaker's gesture threshold. A speaker may set his or her gesture threshold lower for information that is high in communicative dynamism, such as an event that is particularly important to a developing story. Such information is therefore more likely to be expressed in gesture.

Third, the ongoing discourse may influence the dynamic coordination of the oral and manual systems at a particular moment. When a speaker is engaged in gesturing, the oral and manual systems may become increasingly entrained, so that each subsequent utterance may increase in its likelihood of being accompanied by a gesture. Synchronization may also occur between speakers, so that if one speaker gestures at a high rate, the other speaker may be encouraged to do the same (Chartrand & Bargh, 1999; M. J. Richardson, Marsh, & Schmidt, 2005).

The GSA framework holds that gestures occur as the result of interactions among many factors. Although this makes the framework more comprehensive than some others, it also makes it difficult to predict when gestures will occur without knowing the precise state of many aspects of a speaker's cognitive and social situation at the moment of speaking. Nonetheless, we believe that the GSA framework is falsifiable and that it represents a valuable integration and extension of past research. In the following section, we compare the GSA framework with other contemporary frameworks for studying gestures. We conclude by discussing empirical predictions that derive from the GSA framework.

Comparison of the GSA Framework With Other Frameworks

There are many existing frameworks for studying gesture with which we could compare the GSA framework. Some influential frameworks, notably those of Kendon (2004) and Goodwin (2000), reflect an ethnographic, semiotic perspective, focusing on detailed analyses of face-to-face interactions. Others, like our own, are derived from a more psychological perspective, seeking to describe how gestures are produced and how they relate to language and cognition. Here we compare the GSA framework with five other frameworks, all of which are rooted in a psychological

or psycholinguistic tradition: the *sketch model* (de Ruiter, 2000), the *lexical gesture process model* (Krauss et al., 2000), the *interface model* (Kita & Özyürek, 2003), *growth point theory* (McNeill, 1992, 2005; McNeill & Duncan, 2000), and the *gesture-in-learning-and-development* (GLD) *framework* (Goldin-Meadow, 2003). The frameworks we consider differ in their perspectives on four key issues: (1) the nature of the representations that underlie gestures, (2) whether gestures are communicative, (3) how gestures facilitate speech production, and (4) how gesture and speech are integrated. The perspectives of each of the frameworks with respect to these four issues are summarized in Table 1 and are discussed below.

What is the nature of the representations that underlie gestures? As discussed earlier, a great deal of evidence suggests that gestures derive from spatial representations. This claim is not highly controversial, and at a general level, it is shared by all of the frameworks under discussion here. However, the frameworks make varying claims about two related points: (1) whether these spatial representations are stored as visuospatial images or as sets of spatial features, and (2) whether linguistic factors can also influence the production of gestures.

On the side of pure visuospatial imagery, the sketch model (de Ruiter, 2000) proposes that when a message is conceptualized, a *sketch generator* accesses visuospatial imagery in working memory and selects features of an image to be expressed in gesture. This information is then sent to a *gesture planner* that turns the sketch into a motor program. This model considers gestures as arising from visuospatial imagery, pure and simple. The purpose of the sketch generator is to select which aspects of a holistic image to express in gesture, and this information is sent to the gesture planner without input from linguistic factors.

The lexical gesture process model (Krauss et al., 2000) also posits that linguistic factors do not influence gesture production. However, in this model, gestures are based not on visuospatial images, but on sets of elementary spatial features. Relevant features are selected from working memory and turned directly into a motor plan. Thus, there is no need for something akin to the sketch generator that whittles a holistic image down to its relevant features. Instead, the features themselves are the starting point.

The interface model (Kita & Özyürek, 2003) disagrees with the lexical gesture process model (Krauss et al., 2000) on both points. The interface model suggests that gestures arise from visuospatial images and that linguistic factors can influence gesture production. Specifically, an *action generator* is responsible for planning the form of a gesture, and this generator accesses visuospatial images that are active in working memory. Furthermore, there is bidirectional communication between the action generator and the *message generator* responsible for speech. This communication allows for the form of a gesture to be influenced by linguistic constraints of the speaker's language, as well as by visuospatial images in working memory.

Growth point theory (McNeill, 1992, 2005) asserts that gestures emerge from *growth points*, or minimal units that combine imagery and linguistic categorical content. However, in growth point theory, the imagery that gives

Table 1
A Comparison of the GSA Framework With Other Models of Gesture Production

Issue	Sketch Model (de Ruiter, 2000)	Lexical Gesture Process Model (Krauss et al., 2000)	Interface Model (Kita & Özyürek, 2003)	Growth Point Theory (McNeill, 1992, 2005)	Gesture-in-Learning-and-Development Framework (Goldin-Meadow, 2003)	Gesture-as-Simulated-Action (GSA) Framework
Representations Underlying Gestures						
Are gestures based on visuospatial images?	Yes	No; elementary spatial features	Yes	No; imagery is linguistically categorized	Not addressed; but see Wagner et al. (2004)	Yes
Do linguistic factors influence gestures?	No	No	Yes	Yes	Not addressed	Yes
Communication						
Do gestures communicate?	Not addressed	Not usually	Not addressed	Yes	Yes	Not addressed
Are gestures communicatively intended?	Yes	Not usually	Yes	Yes	Yes	Yes
Speech Facilitation						
How do gestures facilitate speech production?	Not addressed; maintain working memory	Facilitate lexical access	Facilitate utterance planning	Drive discourse forward	Reduce cognitive load	Multiple mechanisms
Integration of Speech and Gesture						
Are speech and gesture one system or two?	Two systems with overlap during conceptualization	Two systems that interact during formulation and articulation	Two systems that interact during conceptualization	One system	Develop into one system	One system

rise to gestures is "categorized linguistically . . . and is never purely visuospatial" (McNeill, 2005, p. 131). Thus, according to growth point theory, gestures are not based solely on visuospatial imagery, and they are influenced by linguistic factors. This is a reversal of the claims made by the sketch model (de Ruiter, 2000).

Finally, the GLD framework (Goldin-Meadow, 2003) does not directly address either of these issues. This framework states that gestures "make use of visual imagery to convey meaning" (p. 24), but it does not take a stance on whether gestures arise from visual images or are based on spatial features.[3] Much of the data informing this framework come from studies of gesture–speech *mismatches*, in which speakers express one idea in gesture and another in speech. The fact that gestures can express information that differs from the information expressed in speech implies that at least some gestures are not influenced by linguistic factors; such gestures seem to be produced independently of speech.

With respect to these issues, the GSA framework is most closely aligned with the interface model (Kita & Özyürek, 2003). According to the GSA framework, gestures arise from simulation of the motor and perceptual components of visuospatial imagery. Unlike the interface model, the GSA framework does not explicitly propose bidirectional communication between gesture and speech. However, one assumption of the GSA framework is that linguistic planning involves simulation of visuospatial events, and this simulation ultimately gives rise to gestures. It is likely that the way in which speakers simulate visuospatial events is influenced by the constraints of the speakers' languages. Speakers may simulate actions and perceptions differently depending on the linguistic possibilities (and impossibilities) of their languages.

What role do gestures play in communication? This general issue subsumes two subsidiary issues. First, do listeners acquire information from speakers' gestures? Second, do speakers intend for their gestures to be communicative? Melinger and Levelt (2004) and de Ruiter (2000) have provided valuable discussions of the crucial distinctions that underpin these two questions.

The first question—whether listeners glean information from gestures—is not directly addressed by many of the frameworks under consideration here. Since they are primarily concerned with how and why gestures are produced, most frameworks consider the issue of whether gestures are communicative to be tangential. The two frameworks that most explicitly address this issue are the lexical gesture process model (Krauss et al., 2000) and the GLD framework (Goldin-Meadow, 2003), and they take opposite stances. Krauss and colleagues have argued that many gestures are ineffective at communicating, especially in typical communicative contexts, where both gesture and speech are available (Krauss, Dushay, Chen, & Rauscher, 1995; Krauss, Morrel-Samuels, & Colasante, 1991). Goldin-Meadow and colleagues, in contrast, have provided evidence that listeners do encode information from gesture (Alibali, Flevares, & Goldin-Meadow, 1997; Goldin-Meadow & Sandhofer, 1999; Goldin-Meadow, Wein, & Chang, 1992; Singer & Goldin-Meadow, 2005). The oppo-

site conclusions are likely due to differences in the types of data that the two groups of researchers considered. Goldin-Meadow and colleagues' evidence is based largely on cases in which gesture and speech express different information. It is quite possible that the communicative effectiveness of gestures depends on their redundancy with the accompanying speech. When gestures are redundant, they may communicate very little unique information. When gestures are not redundant, they can communicate rich information that is not available in speech alone.

The remaining four frameworks do not directly address the issue of whether gestures communicate information to listeners. However, none of the four frameworks excludes the possibility that such communication does occur in some situations. Growth point theory (McNeill, 1992, 2005), in particular, contends that considering gesture and speech together provides the most accurate insight into a speaker's thoughts. Thus, gestures must be able to provide information that is not available from speech alone.

The second question—whether speakers intend for their gestures to communicate—is more central to many of the frameworks considered here. The sketch model (de Ruiter, 2000), interface model (Kita & Özyürek, 2003), growth point theory (McNeill, 1992, 2005), and GLD framework (Goldin-Meadow, 2003) agree that gestures are communicatively intended, in the sense that they arise during speech conceptualization. For example, the sketch model (de Ruiter, 2000) contends that the communicative intention formed in the conceptualizer includes both a message (which becomes speech) and a sketch (which becomes gesture). In contrast, the lexical gesture process model (Krauss et al., 2000) holds that gestures should not be considered part of the speaker's communicative intention. According to the lexical gesture process model, most gestures originate outside the communicative process altogether, in visuospatial working memory.

According to the GSA framework, gestures are manifestations of the simulated actions and perceptions that underlie thinking. Thus, motor plans for gestures should be activated whenever simulation occurs, and it seems more accurate to think about these plans as originating in working memory than in the communication process. However, the GSA framework also proposes that not all simulations are manifested as gestures. Whether they are depends on the speaker's gesture threshold, which in turn depends on a complex set of factors, including the speaker's desire to communicate and his or her beliefs about whether gestures could be helpful toward that goal. According to the GSA framework, then, the plan for a gesture is initiated in working memory; whether the gesture is actually produced depends on additional factors, including the desire to communicate. The GSA framework thus contends that some gestures are communicatively intended, and, indeed, the gesture threshold may be lowered for communicative reasons. Other gestures may not be communicatively intended—for example, gestures produced when the threshold is high—but action is simulated strongly enough to pass the high threshold anyway.

Do gestures facilitate speech production? If so, how? This issue was discussed in the previous section on ges-

tures and mental imagery, and will therefore be considered only briefly here. Explaining how gestures facilitate speech production is the primary goal of the lexical gesture process model (Krauss et al., 2000). Specifically, Krauss et al. (2000) claimed that the motor program is sent to a *kinesic monitor* that influences the phonological encoder of speech production. Essentially, the movement of gesture influences the speech production process at the stage of word finding (this was referred to earlier as the *lexical access hypothesis*).

According to the interface model (Kita & Özyürek, 2003), the plan for a gesture that is formed in the *action generator* influences the planning for speech in the *message generator*, and vice versa. In this model, which is in line with the information packaging hypothesis discussed earlier, gestures facilitate speech production by helping to organize spatial information in a way that is compatible with the linear demands of speaking.

The GLD framework (Goldin-Meadow, 2003) proposes that gestures are beneficial to the cognitive system as a whole, rather than being specific to speech production. According to this framework, gestures reduce cognitive load so that more effort may be devoted to speech production (or to some other task). For example, gesturing during explanations of math problems enables speakers to maintain more unrelated information in memory than they can when they do not gesture (Goldin-Meadow et al., 2001; Wagner et al., 2004). This finding holds true not only when the to-be-remembered information is verbal, but also when it is spatial (Wagner et al., 2004), suggesting that gestures do not merely shift load from verbal working memory to visuospatial working memory. Instead, gestures seem to reduce the load on the entire cognitive system. Goldin-Meadow (2003) has suggested several possibilities for how this may occur, including boosting activation of the system overall, structuring spatial information, and indexing words to the surrounding environment.

The sketch model (de Ruiter, 2000) and growth point theory (McNeill, 1992, 2005) are less concerned with describing how gestures facilitate speech production than with describing how gestures arise from speaking and thinking. Nonetheless, the implicit claim in both frameworks is that gestures can facilitate communication (if not speech production, specifically) by expressing information in gestures that need not be expressed verbally.[4] McNeill (2005) has argued that gestures work together with speech to drive discourse forward. By expressing new ideas or growth points, gestures highlight the crucial features of a speaker's message—that is, the features that then become part of the field of oppositions that structures the ongoing discourse.

As is the case for the sketch model (de Ruiter, 2000) and growth point theory (McNeill, 1992, 2005), the main goal of the GSA framework is to describe how gestures arise from speaking and thinking, rather than to describe how they might facilitate speech production. However, the issue of speech facilitation is relevant to the GSA framework, since the gesture threshold can be affected by how beneficial a gesture would be to the current state of the cognitive system or to the current social interaction. Thus, according to the GSA framework, speech facilitation may be a by-product and, in some cases, a goal of gesture production. However, characterizing how gestures might benefit the speech production system is not the aim of the framework. Indeed, within the GSA framework, all of the facilitative effects suggested by other models are possible.

How are gesture and speech integrated? Many of the models discussed here start with Levelt's (1989) model of speech production and then propose that gesture production is a separate, but sometimes interacting, system. The sketch model (de Ruiter, 2000) views gesture as being a system separate from that of speech production, with the two systems overlapping only during the early stage of conceptualization, when the communicative intention is formed. The lexical gesture process model (Krauss et al., 2000) also treats gesture production as being a system separate from speech production and holds that the two systems can interact in two places. Gestures can inform speech production during the formulation stage, when the words for the verbal utterance are being planned. Speech can inform gesture production following the articulation of the words, when an *auditory monitor* can inform the motor planner to discontinue the gesture. The interface model (Kita & Özyürek, 2003) also proposes that gesture and speech are two separate systems but allows for intense interaction between the two during the conceptualization stage.

In contrast, growth point theory (McNeill, 1992, 2005) proposes that gesture and speech are two parts of a single system and cannot be separated. According to McNeill (2005), both gesture and speech are needed to express a growth point; they cannot be considered separately.

The GLD framework (Goldin-Meadow, 2003) also considers gesture and speech to be two components of a single system in mature speakers; a speaker's meaning can be understood best by considering both gesture and speech, particularly when there is a mismatch between the meaning of the gesture and the meaning of the speech. However, the integration of gesture and speech into one system is something that occurs over the course of a child's development. Gesture and speech are originally two separate systems that develop into a single, integrated system in the early years (Butcher & Goldin-Meadow, 2000).

In this regard, the GSA framework is similar to growth point theory (McNeill, 1992, 2005) and to the GLD framework (Goldin-Meadow, 2003). The GSA framework also considers gesture and speech to be two parts of the same cognitive and communicative system. Speaking involves simulations of perception and action; forming a simulation evokes a motor plan that can come to be expressed in gesture. Thus, gestures are a natural by-product of the cognitive processes that underlie speaking, and it is difficult to consider the two separately because both are expressions of the same simulation. In some cases, gesture and speech may express different aspects of that simulation, as in gesture–speech mismatches, but they derive from a single simulation; thus, they are part of the same system.

Predictions of the GSA Framework

The central claims of the GSA framework should be manifested in observable behavior. The central tenet of the

framework is that gestures reflect embodied simulations of action and perception that underlie thinking. This implies that when speakers simulate actions or perceptions while speaking, they should gesture more than when they rely on stored verbal or propositional codes.

Further, even when speakers make simulations while speaking, those simulations may differ in the extent to which they are tied to action. One way to conceptualize variations in imagery is proposed by Cornoldi and Vecchi (2003). They consider images activated in visuospatial working memory to fall along a continuum, and our adaptation of this continuum is shown in Figure 3. The strength of action involved in a particular simulation is represented with shades of black, with darker shades representing a stronger action component. As one moves from visual images through spatial images to motor images, the amount of action simulation increases, and, according to the GSA framework, so does the likelihood of gesture. Indeed, previous research has shown that gestures occur more frequently with descriptions of motor imagery than of visual imagery (Feyereisen & Havard, 1999; Hostetter & Alibali, 2007); however, in these studies, the information being described in each condition was quite different. Future research should manipulate the type of imagery that underlies speech while controlling the content of the imagery. For example, speakers should gesture more when describing a pattern that they have constructed manually (motor imagery) than when describing the same pattern that they have simply viewed (visual imagery). Additionally, speakers describing motor imagery should produce more character-viewpoint gestures than should speakers describing visual imagery. This difference between gestures based on motor imagery and those based on visual imagery is predicted by the GSA framework. Although the other frameworks discussed here could presumably explain such a finding, it is not directly predicted by any of them.

In addition to the horizontal continuum of visuo-spatio-motor imagery, Figure 3 depicts activation as varying along the vertical dimension. Following Cornoldi and Vecchi (2003), the vertical dimension represents the amount of active processing of the image, with relatively passive processes, such as storage and maintenance, falling near the base, and more active processes, such as rotation and synthesis, falling closer to the peak. As seen in Figure 3, the amount of action involved in a simulation increases with the amount of active processing needed. Thus, according to the GSA framework, tasks that require a high amount of active processing should elicit more gestures than should tasks that require passive processing. This is in line with a finding reported by Trafton et al. (2005): Gestures occur more often with speech about mental transformations, which involve simulated movement, than with speech about any other spatial property (e.g., size, shape). Future research could test this claim more directly by asking participants to describe information either as it was originally viewed or as it would look after undergoing a mental transformation (e.g., rotation). The GSA framework predicts that gestures should occur more often when images are being transformed and recalled than when they are only being recalled. Again, although other frameworks

might be able to explain such a finding, they do not make this prediction explicitly a priori.

The GSA framework also makes some predictions about the form a particular gesture will take, in addition to the probability of its occurring at all. A single concept is often gestured differently by different speakers in different situations. For example, a speaker may gesture the concept *small* in one situation by placing the thumb and index finger of one hand a few inches apart, and in another situation may gesture *small* by placing both hands a few inches apart. We contend that the form of a gesture is related to the action component of a given simulation. Although both of the *small* gestures described above represent actions (grasping a small object), one represents performing this action unimanually, and the other, bimanually. We predict that this difference in form should be predictable by the characteristics of the object being referred to. If speakers are talking about an object that they would likely pick up with a single hand, the first *small* gesture should be more likely. If speakers are talking about an object that they would grasp bimanually, the second *small* gesture should be more likely.
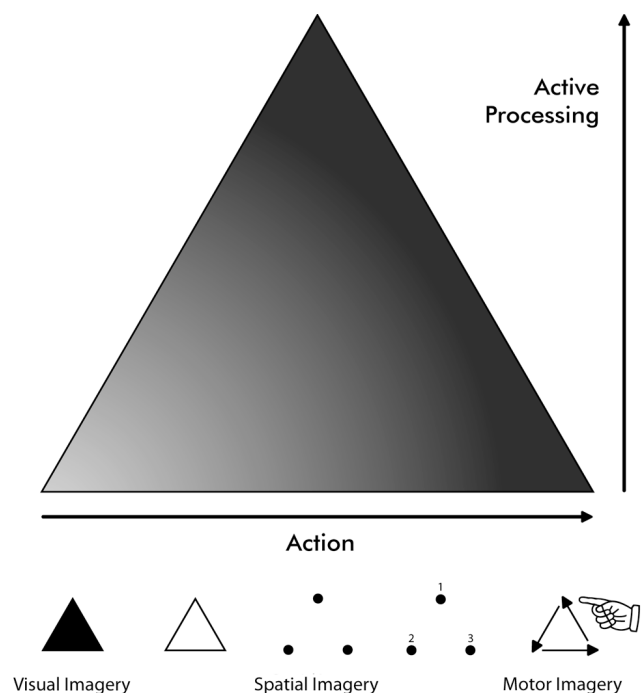

Figure 3. A schematic depiction of how the action component involved in a particular representation varies as a function of type of imagery and of amount of active processing (adapted from Cornoldi & Vecchi, 2003). Darker shades represent stronger amounts of action activation. The amount of action involved in imagery varies along the horizontal continuum, with spatial sequential (e.g., each corner dot of a triangle being represented in succession) and motor (e.g., representation of how one's finger would move in order to create the outline of a triangle) representations relying heavily on action. Visual imagery, however, may primarily evoke action representations when it is being imagined in motion or when it is being actively processed (represented by the vertical continuum). Active processes (e.g., mental rotation, transformation, etc.) evoke action more strongly than do passive processes (e.g., maintenance, rehearsal, etc.).

The effects of the gesture threshold should also be observable in behavior. The gesture threshold can account for individual differences in gesture production across speakers, and in this sense, it is a strength of the GSA framework. Other models predict that a particular situation should elicit similar amounts of gesture from all individuals, but there is wide variability in how much individuals gesture. At least some of this variability can be attributed to individual differences in cognitive variables (Hostetter & Alibali, 2007) and in social variables (Campbell & Rushton, 1978). However, the full array of variables that likely matter remains to be determined. We propose that one factor that affects gesture rate is the speaker's motivation to communicate. A speaker who is in a situation where the stakes of communicating are particularly high should gesture more than a speaker who is in a situation where the stakes are not as high. This suggests that the gestures of individuals in groups with a strong interest in being communicatively effective, such as teachers, advertisers, or doctors, might be particularly revealing. The stakes of communicating could also be manipulated experimentally, perhaps by offering participants a higher reward for communicating effectively. Furthermore, it is possible that motivation to communicate might influence gesture production only in speakers who believe that gestures are helpful to successful communicating. Thus, it may be valuable to devise a way of measuring individual speakers' attitudes about gestures, such as whether gestures are helpful and in what circumstances they are appropriate.

As conceptualized within the GSA framework, the gesture threshold implies that gesturing is the norm, and inhibiting gesture is an effortful task. This is in line with findings showing that inhibiting gesture during an explanation task diverts resources from a secondary task (Goldin-Meadow et al., 2001; Wagner et al., 2004). It is unclear from these findings, however, whether gestures relieve cognitive demands or whether inhibiting gestures increases cognitive demands. Most frameworks that emphasize the facilitative role of gesture emphasize that gestures relieve cognitive demands, without ruling out the possibility that inhibition may also utilize cognitive resources. The GSA framework, however, emphasizes that inhibition utilizes cognitive resources, without ruling out the possibility that gestures might also relieve cognitive demands. These alternatives could be tested by comparing the effects of physically restraining gesture with the effects of asking participants to inhibit their own gesture. Facilitative frameworks predict no difference in performance; in both cases, participants are unable to benefit from producing gestures. The GSA framework, however, predicts that asking participants to inhibit their own gestures should be more detrimental than physically restraining the participants' hands, because inhibiting one's own gestures—that is, maintaining a higher gesture threshold—requires more cognitive effort.

Conclusion

In this article, we have sought to integrate ideas about embodied cognition, language processing, mental imagery, and gesture production, with the goal of formulating a theory about how gestures arise during speech production. In short, our claim is that gestures occur as the result of simulated action and perception, which are the bases of mental imagery and language production. Whether embodied simulations are executed as gestures depends on three factors: the strength of the simulation's action component, the speaker's current gesture threshold, and whether the system is concurrently engaged in speaking.

The GSA framework has its roots in many of the assumptions of previous frameworks, but it offers a unique perspective on how gestures arise from an embodied cognitive system. Specifically, the GSA framework differs from other frameworks in the claims it makes about the source of gestures. Whereas other models and frameworks focus on the role of gestures in the speaking process, the GSA framework situates gestures in the larger cognitive system, which includes mental imagery, embodied simulations of perception and action, and language production. The GSA framework also takes a more dynamic view of gestures than do many of the other frameworks. According to the GSA framework, whether a gesture is produced is the product of many interacting factors, such as the strength of active simulation and the current cognitive and social situation—factors that can change from moment to moment during speaking.

In conclusion, we believe that thinking about gestures from an embodied perspective is valuable for two reasons. First, for researchers interested in gesture's roles in communication and cognition, the embodied approach can inform thinking about how gestures arise and how they represent the speaker's active thoughts at the moment of speaking. Second, for researchers who wish to use gestures as evidence that thinking is embodied, their claims will be strengthened by an account of how gestures actually arise from embodied thinking. The GSA framework offers an account of how gestures make embodiment visible.

REFERENCES

ADOLPH, K. E. (1997). Learning in the development of infant locomotion. *Monographs of the Society for Research in Child Development*, **62**, 1-140.

ALIBALI, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition & Computation*, **5**, 307-331.

ALIBALI, M. W., BASSOK, M., SOLOMON, K. D., SYC, S. E., & GOLDIN-MEADOW, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Science*, **10**, 327-333.

ALIBALI, M. W., FLEVARES, L. M., & GOLDIN-MEADOW, S. (1997). As-

sessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, **89**, 183-193.

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 & Language*, **44**, 169-188.

ALIBALI, M. W., KITA, S., & YOUNG, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language & 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.

BARSALOU, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, **22**, 577-660.

BARSALOU, L. W., & WIEMER-HASTINGS, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 129-163). Cambridge: Cambridge University Press.

BEATTIE, G., & COUGHLAN, J. (1998). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology*, **90**, 35-56.

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

BEGG, I. (1972). Recall of meaningful phrases. *Journal of Verbal Learning & Verbal Behavior*, **11**, 431-439.

BERTENTHAL, B. I., CAMPOS, J. J., & KERMOIAN, R. (1994). An epigenetic perspective on the development of self-produced locomotion and its consequences. *Current Directions in Psychological Science*, **5**, 140-145.

BORGHI, A. M. (2005). Object concepts and action. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 8-34). Cambridge: Cambridge University Press.

BORGHI, A. M., GLENBERG, A. M., & KASCHAK, M. P. (2004). Putting words in perspective. *Memory & Cognition*, **32**, 863-873.

BUTCHER, C., & GOLDIN-MEADOW, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235-258). Cambridge: Cambridge University Press.

CAMPBELL, A., & RUSHTON, J. P. (1978). Bodily communication and personality. *British Journal of Social & Clinical Psychology*, **17**, 31-36.

CAMPOS, J. J., ANDERSON, D. I., BARBU-ROTH, M. A., HUBBARD, E. M., HERTENSTEIN, M. J., & WITHERINGTON, D. (2000). Travel broadens the mind. *Infancy*, **1**, 149-219.

CAMPOS, J. J., BERTENTHAL, B. I., & KERMOIAN, R. (1992). Early experience and emotional development: The emergence of wariness of heights. *Psychological Science*, **3**, 61-64.

CHAMBERS, C. G., TANENHAUS, M. K., EBERHARD, K. M., FILIP, H., & CARLSON, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory & Language*, **47**, 30-49.

CHAMBERS, C. G., TANENHAUS, M. K., & MAGNUSON, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 687-696.

CHARTRAND, T. L., & BARGH, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality & Social Psychology*, **76**, 893-910.

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

CHURCH, R. B., & GOLDIN-MEADOW, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, **23**, 43-71.

COOPER, L. A. (1976). Demonstration of a mental analog of an external rotation. *Perception & Psychophysics*, **19**, 296-302.

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

DE RUITER, J. P. (1998). *Gesture and speech production*. Unpublished doctoral dissertation, Katholieke Universiteit Nijmegen.

DE RUITER, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284-311). Cambridge: Cambridge University Press.

DECETY, J., & JEANNEROD, M. (1995). Mentally simulated movements in virtual reality: Does Fitts's law hold in mental imagery? *Behavioural Brain Research*, **72**, 127-134.

DECETY, J., JEANNEROD, M., & PRABLANC, C. (1989). The timing of mentally represented actions. *Behavioural Brain Research*, **34**, 35-42.

DELL, G. S., CHANG, F., & GRIFFIN, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, **23**, 517-542.

DENIS, M., GONÇALVES, M.-R., & MEMMI, D. (1995). Mental scanning of visual images generated from verbal descriptions: Towards a model of image accuracy. *Neuropsychologia*, **33**, 1511-1530.

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

DI PELLEGRINO, G., FADIGA, L., FOGASSI, L., GALLESE, V., & RIZZOLATTI, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, **91**, 176-180.

EFRON, D. (1972). *Gesture, race, and culture.* The Hague: Mouton.

EJIRI, K. (1998). Synchronization between preverbal vocalizations and motor actions in early infancy: I. Pre-canonical babbling vocalizations synchronize with rhythmic body movements before the onset of canonical babbling. *Japanese Journal of Psychology*, **68**, 433-440.

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

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

FREYD, J. J., & FINKE, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 126-132.

FRICK-HORBURY, D., & GUTTENTAG, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *American Journal of Psychology*, **111**, 43-62.

GALLAGHER, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.

GALLESE, V., & GOLDMAN, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, **2**, 493-501.

GANIS, G., THOMPSON, W. L., & KOSSLYN, S. M. (2004). Brain areas underlying visual imagery and visual perception: An fMRI study. *Cognitive Brain Research*, **20**, 226-241.

GENTILUCCI, M., & DALLA VOLTA, R. (in press). Spoken language and arm gestures are controlled by the same motor control system. *Quarterly Journal of Experimental Psychology*.

GERLACH, C., LAW, I., & PAULSON, O. B. (2002). When action turns into words: Activation of motor-based knowledge during categorization of manipulable objects. *Journal of Cognitive Neuroscience*, **14**, 1230-1239.

GERWING, J., & BAVELAS, J. B. (2004). Linguistic influences on gesture's form. *Gesture*, **4**, 157-195.

GIBBS, R. W., JR. (2003). Embodied experience and linguistic meaning. *Brain & Language*, **84**, 1-15.

GIBBS, R. W., JR. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

GIBBS, R. W., JR., & BERG, E. A. (2002). Mental imagery and embodied activity. *Journal of Mental Imagery*, **26**, 1-30.

GIBSON, E. J., & WALK, R. D. (1960). The "visual cliff." *Scientific American*, **202**, 64-71.

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

GLENBERG, A. M. (1997). What memory is for. *Behavioral & Brain Sciences*, **20**, 1-55.

GLENBERG, A. M., HAVAS, D., BECHER, R., & RINCK, M. (2005). Grounding language in bodily states: The case for emotion. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 115-128). Cambridge: Cambridge University Press.

GLENBERG, A. M., & KASCHAK, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, **9**, 558-565.

GLENBERG, A. M., & ROBERTSON, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, **43**, 379-401.

GOLDIN-MEADOW, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

GOLDIN-MEADOW, S., NUSBAUM, H., KELLY, S. D., & WAGNER, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, **12**, 516-522.

GOLDIN-MEADOW, S., & SANDHOFER, C. M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, **2**, 67-74.

GOLDIN-MEADOW, S., WEIN, D., & CHANG, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition & Instruction*, **9**, 201-219.

GOODWIN, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, **32**, 1489-1522.

GRAFTON, S. T., FADIGA, L., ARBIB, M. A., & RIZZOLATTI, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *NeuroImage*, **6**, 231-236.

GRUSH, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral & Brain Sciences*, **27**, 377-396.

HADAR, U., BURSTEIN, A., KRAUSS, R. M., & SOROKER, N. (1998). Ideational gestures and speech in brain-damaged subjects. *Language & Cognitive Processes*, **13**, 59-76.

HARNAD, S. (1990). The symbol grounding problem. *Physica D*, **42**, 335-346.

HAUEISEN, J., & KNÖSCHE, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, **13**, 786-792.

HAUK, O., JOHNSRUDE, I., & PULVERMÜLLER, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, **41**, 301-307.

HECHT, H., VOGT, S., & PRINZ, W. (2001). Motor learning enhances perceptual judgment: A case for action–perception transfer. *Psychological Research*, **65**, 3-14.

HELSTRUP, T., & ANDERSEN, R. E. (1991). Imagery in mental construction and decomposition tasks. In R. H. Logie & M. Denis (Eds.), *Mental images in human cognition* (pp. 229-240). New York: Elsevier, North-Holland.

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

HOSTETTER, A. B., & ALIBALI, M. W. (2004). On the tip of the mind: Gesture as a key to conceptualization. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 589-594). Mahwah, NJ: Erlbaum.

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

HOSTETTER, A. B., ALIBALI, M. W., & KITA, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language & Cognitive Processes*, **22**, 313-336.

HOSTETTER, A. B., BIEDA, K., ALIBALI, M. W., NATHAN, M., & KNUTH, E. J. (2006). Don't just tell them, show them! Teachers can intentionally alter their instructional gestures. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1523-1528). Mahwah, NJ: Erlbaum.

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

IVERSON, J. M., & THELEN, E. (2000). Hand, mouth, and brain: The dynamic emergence of speech and gesture. In R. Nuñez & W. J. Freeman (Eds.), *Reclaiming cognition: The primacy of action, intention, and emotion* (pp. 19-40). Charlottesville, VA: Imprint Academic.

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

JEANNEROD, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, **33**, 1419-1432.

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

JOHNSON, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

KASCHAK, M. P., MADDEN, C. J., THERRIAULT, D. J., YAXLEY, R. H., AVEYARD, M., BLANCHARD, A. A., & ZWAAN, R. A. (2005). Perception of motion affects language processing. *Cognition*, **94**, B79-B89.

KASCHAK, M. P., ZWAAN, R. A., AVEYARD, M., & YAXLEY, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science*, **30**, 733-744.

KELLY, S. D., IVERSON, J. M., TERRANOVA, J., NIEGO, J., HOPKINS, M., & GOLDSMITH, L. (2002). Putting language back in the body: Speech and gesture on three time frames. *Developmental Neuropsychology*, **22**, 323-349.

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

KIRSH, D., & MAGLIO, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, **18**, 513-549.

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

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 & Language*, **48**, 16-32.

KOSSLYN, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics*, **14**, 90-94.

KOSSLYN, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, **22**, 333-347.

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

KRAUSS, R. M., CHEN, Y., & CHAWLA, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, 389-450). San Diego: Academic Press.

KRAUSS, R. M., CHEN, Y., & GOTTESMAN, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261-283). Cambridge: Cambridge University Press.

KRAUSS, R. M., DUSHAY, R. A., CHEN, Y., & RAUSCHER, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, **31**, 533-552.

KRAUSS, R. M., MORREL-SAMUELS, P., & COLASANTE, C. (1991). Do conversational hand gestures communicate? *Journal of Personality & Social Psychology*, **61**, 743-754.

LAKOFF, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

LAKOFF, G., & JOHNSON, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

LAKOFF, G., & NUÑEZ, R. E. (2001). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books.

LEVELT, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

MARKMAN, A. B., & DIETRICH, E. (2000). Extending the classical view of representation. *Trends in Cognitive Sciences*, **4**, 470-475.

MAST, F. W., & KOSSLYN, S. M. (2002). Visual mental images can be ambiguous: Insights from individual differences in transformation abilities. *Cognition*, **86**, 57-70.

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

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

MCNEILL, D., & DUNCAN, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141-161). Cambridge: Cambridge University Press.

MELINGER, A., & KITA, S. (2007). Conceptualisation load triggers gesture production. *Language & Cognitive Processes*, **22**, 473-500.

MELINGER, A., & LEVELT, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, **4**, 119-141.

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

NIEDENTHAL, P. M. (2007). Embodying emotion. *Science*, **316**, 1002-1005.

NIEDENTHAL, P. M., BARSALOU, L. W., WINKIELMAN, P., KRAUTH-GRUBER, S., & RIC, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality & Social Psychology Review*, **9**, 184-211.

NUÑEZ, R. [E.] (2005). Do real numbers really move? Language, thought, and gesture: The embodied cognitive foundations of mathematics. In F. Ilda, R. P. Pfeifer, L. Steels, & Y. Kuniyoshi (Eds.), *Embodied artificial intelligence* (pp. 54-73). Berlin: Springer.

NUÑEZ, R., & SWEETSER, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, **30**, 401-450.

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

PAIVIO, A. (1963). Learning of adjective–noun paired associates as a function of adjective–noun word order and noun abstractness. *Canadian Journal of Psychology*, **17**, 370-379.

PAIVIO, A. (1965). Abstractness, imagery, and meaningulness in paired-associate learning. *Journal of Verbal Learning & Verbal Behavior*, **4**, 32-38.

PAIVIO, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning & Verbal Behavior*, **7**, 41-49.

PAIVIO, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

PAIVIO, A. (1975). Coding distinctions and repetition effects in memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 179-214). New York: Academic Press.

PAIVIO, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, **45**, 255-287.

PECHER, D., ZEELENBERG, R., & BARSALOU, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, **14**, 119-124.

PETERSON, M. A., KIHLSTROM, J. F., ROSE, P. M., & GLISKY, M. L. (1992). Mental images can be ambiguous: Reconstruals and reference-frame reversals. *Memory & Cognition*, **20**, 107-123.

PULVERMÜLLER, F. (1999). Words in the brain's language. *Behavioral & Brain Sciences*, **22**, 253-336.

PULVERMÜLLER, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, **6**, 576-582.

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

RICHARDSON, D. C., SPIVEY, M. J., BARSALOU, L. W., & MCRAE, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, **27**, 767-780.

RICHARDSON, D. C., SPIVEY, M. J., EDELMAN, S., & NAPLES, A. J. (2001). "Language is spatial": Experimental evidence for image schemas of concrete and abstract verbs. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 873-878). Mahwah, NJ: Erlbaum.

RICHARDSON, M. J., MARSH, K. L., & SCHMIDT, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception & Performance*, **31**, 62-79.

RIZZOLATTI, G., FADIGA, L., GALLESE, V., & FOGASSI, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, **3**, 131-141.

ROGERS, T. T., & MCCLELLAND, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

SCHWARTZ, D. L., & BLACK, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 116-136.

SHAPIRO, L. (2007). The embodied cognition research programme. *Philosophy Compass*, **2**, 338-346.

SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

SINGER, M. A., & GOLDIN-MEADOW, S. (2005). Children learn when their teachers' gestures and speech differ. *Psychological Science*, **16**, 85-89.

SOLOMON, K. O., & BARSALOU, L. W. (2001). Representing properties locally. *Cognitive Psychology*, **43**, 129-169.

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

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

STANFIELD, R. A., & ZWAAN, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, **12**, 153-156.

TRAFTON, J. G., TRICKETT, S. B., STITZLEIN, C. A., SANER, L., SCHUNN, C. D., & KIRSCHENBAUM, S. S. (2005). The relationship between spatial transformations and iconic gestures. *Spatial Cognition & Computation*, **6**, 1-29.

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

TUITE, K. (1993). The production of gesture. *Semiotica*, **93**, 83-105.

WAGNER, S. M., NUSBAUM, H., & GOLDIN-MEADOW, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory & Language*, **50**, 395-407.

WESP, R., HESSE, J., KEUTMANN, D., & WHEATON, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, **114**, 591-600.

WEXLER, M., KOSSLYN, S. M., & BERTHOZ, A. (1998). Motor processes in mental rotation. *Cognition*, **68**, 77-94.

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

WILSON, M., & KNOBLICH, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, **131**, 460-473.

WILSON, N. L., & GIBBS, R. W., JR. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science*, **31**, 721-731.

WOHLSCHLÄGER, A., & WOHLSCHLÄGER, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 397-412.

ZWAAN, R. A., & MADDEN, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224-246). Cambridge: Cambridge University Press.

ZWAAN, R. A., STANFIELD, R. A., & YAXLEY, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, **13**, 168-171.

ZWAAN, R. A., & YAXLEY, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, **10**, 954-958.

NOTES

1. The term *premotor* should not be taken to imply *premotor cortex* specifically. Rather, these premotor action states may be the result of activity in any of the areas that are involved in action planning and preparation, including the premotor cortex, but also the cerebellum and other subcortical areas.

2. Coordination between the manual and oral systems can also occur in the absence of any action simulation by the conceptual system. We contend that this situation gives rise to beat gestures, or those movements that correspond to the prosody of speech without conveying semantic content.

3. In at least one study (Wagner et al., 2004), Goldin-Meadow and colleagues seem to suggest that at least some gestures arise from propositional representations rather than imagistic ones.

4. De Ruiter (1998) has argued in favor of the image maintenance theory, that gestures maintain spatial information in working memory so that it can be encoded verbally. However, this view is not an explicit part of his sketch model.