# Sensorimotor Simulation in Speaking, Gesturing, and Understanding

In (Eds.) Müller, C., Cienki, A. J., Fricke, E., Ladewig, S. H., McNeill, D., & Teßendorf, S. (2013). *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*

Marcus Perlman

Raymond W. Gibbs, Jr.

Department of Psychology

University of California, Santa Cruz

Santa Cruz, CA 95064

U.S.A

## Abstract

This chapter examines empirical studies and theory related to the hypothesis that embodied simulations of sensorimotor imagery play a crucial role in the construction of meaning during conversation. We review some of the substantial body of work implicating sensorimotor simulations in language comprehension and also some more recent work suggesting that simulations are similarly involved in the production of gestures. Furthermore, we consider two less standard lines of research that are argued to be relevant to the sensorimotor simulation hypothesis. The first is new empirical evidence demonstrating the pervasive use of vocal gesture in communication. Because of their direct integration with speech, these gestures offer a unique window into the immediate conceptualizations that motivate the online production of spoken language. Second, we take an evolutionary perspective on simulations and examine the use of iconic gestures by the great apes. These reports reveal the phylogenetic emergence of simulation as the basis for the production of iconic gestures. Overall, this chapter offers a more unified understanding of how sensorimotor simulations figure into productive and receptive communication and across the spectrum from language to gesture.

## Sensorimotor Simulation in Speaking, Gesturing, and Understanding

Communicating thoughts requires that people do something with their bodies, whether it be waving their hands, flashing their eyebrows, moving their tongues and lips, or nodding their heads. Each of these bodily activities is traditionally understood as an outward translation of inner thought processes. We speak or gesture in particular ways as a code to signal to others what we are thinking, with particular systematic spoken and bodily gestures typically viewed as having arbitrary connections with the messages we hope others to understand.

This view of language and communication, where the body serves as the conduit for the expression of thought and meaning, does not address why we have the words and gestures we employ nor how these express the meanings they typically communicate. Our argument in this chapter is that spontaneous iconic gestures and even highly conventionalized systems of language and gesture are continually linked to ongoing experiential simulations of sensorimotor imagery. Under this view, articulatory and manual movements associated with language and gesture are produced as the physical manifestations of these simulation processes. Furthermore, one's ability to interpret both speech and gesture as being specifically meaningful in context includes partly simulating what others' must be doing by their use of speech and gesture. A crucial part of this view is that the human body serves not as a mere conduit for the transmission of meaning, but plays an essential role in how we imaginatively simulate meaningful ideas. Humans conceptualize all kinds of imagery through the multimodal actions of their bodies, and we argue that this ability lies at the foundation of linguistic and gestural meaning. We refer to this claim as the "sensorimotor simulation hypothesis" and in this chapter describe some of the theoretical debate and empirical research related to the role that sensorimotor simulation plays in speaking, gesturing, and understanding.

In addition to reviewing the more standard literature on embodied simulation, the chapter also considers two additional lines of research that we argue are relevant to the sensorimotor simulation hypothesis. First we look at recent evidence for the pervasive use of vocal gesture in communication. These gestures, because of their incorporation into the articulatory gestures of speech, offer a unique window into the immediate conceptualizations that motivate the online production of spoken language. Second, we take an evolutionary perspective on simulations and examine reports of the use of iconic gestures by the great apes. It is argued that these reports, spanning almost 100 years, reveal the phylogenetic emergence of simulation as the basis for the production of iconic gestures.

## Mental Simulation and the Body

Mental simulation is an idea that has attracted the attention of many contemporary cognitive scientists in their study of perception, thought, and language. Consider, for example, a single act of visual perception in which a person looks at a coffee cup sitting on a table. One proposal for how people perceive the cup suggests that our perceptual systems have evolved to facilitate our interaction with a real, three-dimensional world. Perception does not take place solely in the eyes or brain of the perceiver, but rather is an act of the whole animal, the act of perceptually guided exploration of the environment, the function of which is to keep the perceiver in touch with the environment and to guide action, not to produce inner experiences and representations (Gibson, 1979). More specifically, at any given moment, the environment affords a range of possibilities, called affordances. A person looking at a coffee cup may, for instance, imagine different ways of interacting with the cup, including reaching to grasp the cup, picking it up, and tilting it in just the right way to sip its contents. One could also do other things while observing a coffee cup, such as imagining it as something to pick up and throw, something to sit on, balance on one's head and so on. Most generally, perception and action are deeply intertwined, such that perceiving anything in the world is very much a matter of anticipating different embodied interactions with these things, including other people. We perceive by simulating embodied possibilities, and then act on them given our adaptive needs in context (Noe & O'Regan, 2002).

Mental simulation is often understood as the "reenactment of perceptual, motor, and introspective states acquired during interactions with world, body, and mind" (Barsalou, 2008: 618). Much behavioral research demonstrates that even higher-order conceptual processing involves sensorimotor simulations. For example, people's categorization behaviors (e.g., recalling as many instances of fruit as possible) reflect their sensorimotor imagining of themselves in different real-world situations (e.g., walking down the produce aisle in a grocery store), rather than their simply retrieving lists of exemplars from a taxonomic list in one's conceptual knowledge (e.g., recalling all fruits stored in memory) (Vallee-Tourangeau, Anthony, & Austin, 1998). People also appear to simulate how different concepts may be combined (e.g., a half smile is different from a half watermelon), such that concepts are better understood as being created on the fly through sensorimotor simulation as opposed to being passively stored as a set of abstract features in memory (Wu & Barsalou, 2001).

Our argument is that both language and gesture are produced and understood in terms of highly imagistic simulation processes, yet scholars differ in their characterization of simulation processes in human cognition. Some psycholinguists adopt the idea that simulation in language use involves neural processes ordinarily having nonlinguistic functions (e.g., perception and action) (Havas, Glenberg, & Rinck, 2007), a view that closely links simulation to brain states. Others embrace an ecological, embodied view in which mental simulations emerge from the dynamical interactions of brains, bodies and world (Gibbs, 2006a). Under this latter perspective, simulation is not purely mental or neural, but a process that involves and effects many full-bodied sensations. For instance, a classic social psychological study asked people to unscramble mixed up wording in individual sentences that contained words referring to elderly people (e.g., grey, old, wrinkled, tired, Florida, wise) (Chartand & Bargh, 1999). After completing this task, participants were videotaped as they left the experimental room. The result of interest is that people walked significantly slower after reading words referring to the elderly than in other conditions. Thus, reading words with certain meanings affected people's subsequent walking behaviors to be similar to the multi-modal simulation processes required to understand those earlier read words. As Zwaan (2004: 38) argued, "comprehension is not the manipulation of abstract, arbitrary and amodal symbols, a language of thought. Rather, comprehension is the generation of vicarious experiences making up the comprehender's experiential repertoire."

## Sensorimotor Simulation in Speaking and Understanding

One of the earliest theories on sensorimotor influences on language understanding is seen in work on the motor theory of speech perception (Galantucci, Fowler & Turvey, 2006). This theory maintains that listeners use their own articulatory motor programs to interpret spoken language. Thus, considerable evidence shows that perception of phonemes is accomplished not simply by analysis of physical acoustic patterns but through their articulatory events, such as movements of the lip, tongue, and so on. People hear speech sounds by imagining producing the stimuli they hear. One analysis of different speech perception and production tasks proposed that listeners "focus on acoustic change, because changing regions of the sound spectrum best reveal the gestural constituency of talker's utterances (Fowler, 1987: 577). More recent research claims that phonetic primitives are gestural, and not abstract features (Browman & Goldstein, 1995).

Articulatory gestures are unified primitives characterizing phonological patterns, in addition to capturing something about the activity of the vocal articulators. Our repository of words in the mental lexicon is more specifically composed of dynamically specified gestures.

In recent years, experimental psycholinguists have studied various ways that the sensorimotor system functions in the higher order structures of language involved in understanding meaningful utterances. One trend in this work is to explicitly explore how overt and covert bodily movement shapes people's interpretation of meaning. Thus, some research demonstrates that appropriate bodily actions (e.g., making a hand gesture with the thumb and fingers touching) facilitate semantic judgments for action phrases such as "aim a dart" (Klatzky, Pelligrino, McCloskey, & Doherty, 1989). In addition, Glenberg and Kaschak (2002) demonstrate what they call the action-sentence compatibility effect. In one experiment, participants made speeded sensibility judgments for sentences that implied action either toward or away from the body (e.g. "Close the drawer" implies action of pushing something away from the body). Participants indicated their judgment by use of a button box which contained a line of three buttons perpendicular to their body. Presentation of the sentence was initiated when the participant pressed the center button, and yes or no responses (i.e., sensible or not sensible) were made with the two remaining buttons, requiring action either away from or toward the body. Glenberg and Kaschak found an interference effect, such that comprehension of a sentence implying action in one direction interfered with a sensibility response made in the opposing direction. This effect was interpreted as evidence that understanding language referring to action recruits the same cognitive resources needed to actually perform the action.

Another study investigated whether people mentally represent the orientation of a referent object when comprehending a sentence (Stanfield & Zwaan, 2001). Participants were presented with sentences that implicitly referred to the orientation of various objects (e.g. "Put the pencil in the cup" implies a vertical orientation of the pencil). After each sentence, a picture was presented, to which participants answered whether the pictured object had been in the previous sentence. For pictures that were contained in the previous sentence, the picture's orientation varied as to whether or not it matched the orientation implied by the sentence (e.g., a pencil was presented in either a vertical or horizontal orientation). Overall, participants responded faster to pictures that matched the implied orientation than to mismatched pictures and sentences. This empirical finding suggests that people form analogue representations of objects during ordinary sentence comprehension, which is consistent with the simulation view of linguistic processing.

Studies also suggest that sensorimotor simulations during language comprehension can be shaped by emotion. Earlier brain imaging experiments showed that observing (perception) and imitating (action) emotional facial expressions activate the same neural areas of emotion, as well as relevant motor parts of the mirror neuron system (Carr, Iacoboni, Dubeau, Mazzitto, & Lenzi, 2003). Generating facial expressions also primes people's recognition of others' facially conveyed emotions (Niedenthal, Braver, Halberstadt, & Innes-Ker, 2001). These findings on simulation and emotion recognition have been extended to language comprehension. For instance, one set of studies evoked positive or negative emotions in participants by having them either hold a pen with their teeth (i.e., producing a smile) or gripping a pen with their lips (i.e., producing a frown) (Havas, Glenberg, & Rinck, 2007). As people were either smiling or frowning, they made speeded judgments as to whether different sentences were either pleasant (e.g., "You and your lover embrace after a long separation)" or unpleasant (e.g., "The police car rapidly pulls up behind you, siren blaring"). Response times to make these judgments showed that people made pleasant judgments faster when smiling than when frowning and made unpleasant judgments faster when frowning than when smiling. Subsequent experiments revealed that these effects could not be replicated when people make speeded lexical decisions to isolated words that were either pleasant (e.g., "embrace") or unpleasant (e.g., "police"), suggesting that emotional simulation during language comprehension operates most strongly at the level of full phrases or sentences. Overall, sensorimotor simulations appear to be critical parts of language understanding as people create meaningful and emotional interpretations of linguistic expressions.

One concern with the psycholinguistic research on word and sentence processing is that simulations may be important in understanding concrete actions and objects, but not necessarily abstract ideas, such as "justice" or "democracy." Yet there is much work in cognitive linguistics showing that people understand at least some abstract concepts in embodied metaphorical terms (Gibbs, 2006a; Lakoff & Johnson, 1999). More specifically, abstract ideas, such as "justice" are structured in terms of metaphorical mappings where the source domains are deeply rooted in recurring aspects of embodied experiences (i.e., ACHIEVING JUSTICE IS ACHIEVING PHYSICAL BALANCE BETWEEN TWO ENTITIES). Many abstract concepts are presumably structured via embodied metaphors (e.g., time, causation, spatial orientation, political and mathematical ideas, emotions, the self, concepts about cognition, morality) across many spoken and signed languages (see Gibbs, 2008). Systematic analysis of conventional expressions, novel extensions, patterns of polysemy, semantic change, and gesture all illustrate how abstract ideas are grounded in embodied source domains. Thus, the metaphorical expression "John ran through his lecture" is motivated by the embodied metaphor of MENTAL ACHIEVEMENT IS PHYSICAL MOTION TOWARD A GOAL (a submetaphor derived from CHANGE IS

MOTION). These cognitive linguistic findings provide relevant evidence showing that abstract concepts are partly structured through embodied simulation processes (Gibbs, 2006b).

A second response to the concern about simulation and abstract concepts comes from different psycholinguistic research demonstrating that embodied conceptual metaphors motivate people's use and understanding of metaphoric language related to various abstract concepts (Gibbs, 2006a, b). These experimental studies indicate that people's recurring embodied experiences often play a role in how people tacitly make sense of many metaphoric words and expressions. In fact, some of these studies extend the work on simulative understanding of non-metaphorical language to processing of metaphorical speech. Gibbs et al. (2006) demonstrated how people's mental imagery for metaphorical phrases, such as "tear apart the argument," exhibit significant embodied qualities of the actions referred to by these phrases (e.g., people conceive of the "argument" as a physical object that when torn apart no longer persists). Wilson and Gibbs (2007) showed that people's speeded comprehension of metaphorical phrases like "grasp the concept" are facilitated when they first make, or imagine making, in this case, a grasping movement. Bodily processes appear to enhance the construction of simulation activities to speed up metaphor processing, an idea that is completely contrary to the traditional notion that bodily processes and physical meanings are to be ignored or rejected in understanding verbal metaphors. Furthermore, hearing fictive motion expressions implying metaphorical motion, such as "The road goes through the desert," influences people's subsequent eye-movement patterns while looking at a scene of the sentence depicted (Richardson & Matlock, 2007). This suggests that the simulations used to understand the sentence involve a particular motion movement of what the roads does, which interacts with people's eye movements.

Experimental findings like these emphasize that people may be creating partial, but not necessarily complete, sensorimotor simulations of speakers' metaphorical messages that involve moment-by-moment "what must it be like" processes, such as grasping, that make use of ongoing tactile-kinesthetic experiences (Gibbs, 2006b). These simulation processes operate even when people encounter language that is abstract, or refers to actions that are physically impossible to perform, such as "grasping a concept," because people can metaphorically conceive of a "concept" as an object that can be grasped. One implication of this work is that people do not just access passively encoded conceptual metaphors from long-term memory during online metaphor understanding, but perform online simulations of what these actions may be like to create detailed understandings of speakers' metaphorical messages (Gibbs, 2006b).

## Gesture as Sensorimotor Simulation

During an 1880 conference held in Milan to advocate oral language education for the deaf, one proponent Guiulo Tarra argued,

> Gesture is not the true language of man which suits the dignity of his nature. Gesture, instead of addressing the mind, addresses the imagination and the senses…Thus, for us it is an absolute necessity to prohibit that language and to replace it with living speech, the only instrument of human thought…Oral speech is the sole power that can rekindle the light God breathed into man when, giving him a soul in a corporeal body, he gave him also a means of understanding, of conceiving, and of expressing himself…mimic signs…enhance and glorify fantasy and all the faculties of the imagination (Lane, 1984: 391; from Wilcox, 2004: 120-121).

The empirical research reviewed above on the active involvement of the sensorimotor system in online language understanding reveals that Tarra was not so much mistaken about the nature of gesture, but rather, the nature of language. Language, through the imaginative processes of simulation, is deeply grounded in movement and the senses, which form the basis for the construction of linguistic meaning. Yet, ironically, as evidence has accumulated for this view on language, empirical research on gesture has been slow to adopt a distinctly simulation perspective, despite its intuitive appeal. One reason for this imbalance may be found in the asymmetry between different approaches to studying meaning in language compared to gesture. While language researchers interested in meaning have tended to focus on comprehension, gesture researchers have tended to pay more attention to production, which does not so easily lend itself to the traditional dependent measures used to examine the online processes involved in meaningful communication. Thus, when gesture comprehension has been measured, this has largely been in the way of offline, targeted assessments of information uptake.

The intuitive basis for a simulation theory of gesture stems from its clear correspondence to sensorimotor imagery, a point that has also received substantial empirical attention over the years (McNeill, 1992, 2005). One line of research has documented a positive correlation between spatial processing and gesture production. For example, Lavergne and Kimura (1987) found that people gesture more frequently when conversing about spatial topics compared to neutral or verbal topics. Another study asked participants to describe animated action cartoons, and found that speech-accompanying gestures were nearly five times as likely to occur with phrases containing spatial prepositions than those without spatial content (Rauscher, Krauss, & Chen, 1996). A study by Hostetter & Alibali (2007) investigated the influence of individual differences in spatial skills on gesture production, with the finding that people with strong spatial skills and weak verbal skills gestured most frequently. Finally, neuropsychological research shows that stroke patients who suffered visuospatial deficits gestured less than matched controls (Hadar, Burstein, Krauss, & Soroker, 1998).

Other studies have revealed a positive relationship between gesture and imagery more broadly. In one experiment, Rime and colleagues had participants engage in a 50 minute conversation while seated in an armchair that restrained the movement of their head, arms, hands, legs, and feet (Rime, Schiaraptura, Hupet and Ghysselinkckx, 1984). Analysis of their dialogue revealed that while people's movement was restricted, the content of their speech showed a significant decrease in the amount of vivid imagery compared to freely moving controls. Another study asked participants to describe a cartoon after, in one condition, watching it, or in a second condition, reading it in narrative form (Hostetter & Hopkins, 2002). People gestured more frequently after watching the cartoon, presumably because of the richer imagery in the non-verbal presentation. Last, Beattie and Shovelton (2002) investigated the properties of verbal clauses that were likely to occur with iconic gestures. Participants narrated cartoons, and afterwards, the clauses in these narrations were rated by other subjects for their imageability. Clauses associated with gestures were rated as more highly imageable.

These various studies demonstrate that gesture is highly correlated with communication about imagery-laden topics that span a range of motor, visual, and spatial imagery. Indeed, these findings are not at all surprising as one considers the highly imagistic nature of gestures themselves, which are inherently visible, spatially-oriented motor actions. Yet, the studies also beg the question of how sensorimotor imagery might relate to gesture within the momentary processes involved in communicative interaction. One recent proposal—the Gesture as Simulated Action (GSA) hypothesis—offers a potential answer, arguing for the idea that gestures arise from the simulation of motor imagery (Hostetter & Alibali, 2008). According to this view, simulations of action-related thoughts lead to the activation of neural premotor action states, which then has the potential to spread to motor areas. This spreading activation comes to be realized as the overt action of representative gesture.

Hostetter and Alibali's GSA hypothesis is clearly related to our claim about the importance of sensorimotor simulation in speaking, gesturing, and understanding. Yet the GSA may have some limitations. The first is not critical, but concerns the scope of the gestural behaviors the GSA actively attempts to explain. Explicitly the theory concerns "representative gestures," which are said to include iconic and deictic gestures. However, deictics receive minimal consideration, and more broadly, the theory does not address language-related gestures, including spoken prosody, manual beat gestures, or iconic "vocal" gestures (see below), nor the conventionalized gestures of speech and sign. Given the tight temporal coordination and semantic coherence between spoken language and gesture (McNeill, 1992, 2005; Perlman, in press), we emphasize the importance of developing a comprehensive theory of gesture aimed to account for the range of these gestural behaviors. This point gains weight as researchers increasingly observe a more graded distinction between modality-independent notions of gesture and language that rests largely on a continuous quality of conventionalization.

A second limitation of the GSA account is in its exclusive emphasis on simulated *action* as the foundation of gesture. The theory argues that simulated motor imagery is the critical conceptual ingredient for gesture production, and that gestures related to non-motor, perceptual imagery might arise, but only through co-activation with afforded actions or by simulated perceptual actions like eye movements or tactile exploration. Though one cannot deny that action is an essential element of many if not all gestures (though consider a gestural depiction of stillness and expand from there), the fact that gestures are composed of substance and context is likewise essential to the simulation process. To fully appreciate the simulative processes involved in the production of gesture, it is critical to attend to the full process of how the body is imaginatively and often metaphorically used to constitute an endless variety of simulated entities and qualities. As McNeill (1992) observed, “The hand can represent a character’s hand, the character as a whole, a ball, a streetcar, or anything else…In other words, the gesture is capable of expressing the full range of meanings that arise from the speaker” (p. 105). Moreover, as we describe below, the cross-modal sorts of representations involved in common vocal gestures demonstrate a complexity of simulation that is just not adequately captured by the simplifying notion of “simulated action.”

This capacity is easily demonstrated even within a conceptually simple context. For example, consider a study by Lausberg and Kita (2003), which investigated hand preference for observer-viewpoint iconic gestures. Participants watched and described animations of two shapes interacting with each other, with one shape oriented on the left side and the other on the right. In one condition, participants verbally described the event as they also produced many speech-accompanying gestures, and in a second condition, they depicted the event silently. Lausberg and Kita were interested in whether the verbal condition would elicit, through left hemispheric activation associated with speech, more right-handed gestures. Indeed, the findings showed only a minimal effect induced by speech, and instead, by far the most important factor influencing hand choice was whether the represented block was oriented on the left or the right side of the animation. Not surprisingly, when the represented block was on the left, speakers tended to use their left hand, and when the block was on the right, speakers tended to use their right hand. In addition, participants also produced numerous bimanual gestures in which the hands worked together to embody the spatial relationship between the blocks. This empirical finding provides a nice and simple illustration of the essential role played by substance and context, in addition to action, in how simulative processes give rise to gesture.

These difficulties with the GSA hypothesis do not damper our enthusiasm for the general idea that gesture is both produced and understood in terms of ongoing sensorimotor simulations. By providing a detailed formulation of their hypothesis, Hostetter and Alibali open the way for direct, empirical testing of a simulation-based theory of gesture, helping to incorporate gesture into an important theory for the comprehension of meaning in language. Notably, by linking the sensorimotor simulations of language comprehension to gesture production, the hypothesis brings together the meaning-making processes involved in language and gesture, as well as comprehension and production.

The value of their proposal can be seen in one recent study that examined the degree of detail incorporated into the production of a gesture, and in turn, how much of this detail is conveyed to the listener (Cook & Tanenhaus, 2009). Participants solved the Tower of Hanoi problem, either with a stack of weights or on a computer, and afterwards described the solution to a listener. (In this problem, a stack of disks is arranged bottom up from largest to smallest on the leftmost of three pegs. The goal is to move all of the disks to the rightmost peg, moving only one disk at a time and without ever placing a larger disk on top of a smaller one.) Analysis of the trajectory of speakers' gestures revealed that these were finely tuned to the actual trajectory involved in solving the problem, with differences reflecting the differently afforded constraints of the real-weight versus computer version of the task. Moreover, listeners were sensitive to this information, which transferred into matching trajectories when listeners later performed the computer version of the problem themselves. Cook and Tanenhaus interpret these analog differences in gesture production and comprehension as evidence for the activation of perceptual-motor information that is involved in the actual performance of the task. More generally, they suggest that this finding is consistent with a sensorimotor simulation account of gesture production and comprehension.

Cook and Tanenhaus provide a compelling interpretation of their results, and we anticipate much future empirical research focused in this direction. Yet, returning to the current formulation of the GSA hypothesis, we also reiterate the need for a more comprehensive account that emphasizes the full embodied and contextualized complexity of these simulation processes and the meaningful actions that result. Given evidence for the multimodal nature of language and gesture, it is crucial to understand, not just the action aspect of gesture, but how the body is incorporated into the construction of meaning during these activities. To illustrate this point, we next describe some recent research and examples related to the production of gestures within the vocal modality.

## Vocal Gesture

When people talk, they commonly pattern their voice in a variety of ways to iconically depict an aspect of their subject matter. Many times these iconic correspondences are created within the domain of sound, with the sounds of our voice imitating the sounds of our environment. One familiar example is when quoting another person's speech, we often imitate certain characteristics of that person's voice, such as their emotional state, accent, and tonal quality (Clark & Gerrig, 1990). Nonhuman animals can also be "quoted," such as when imitating the high-pitched barks of a yapping dog (perhaps in combination with a one-handed gesture of the dog's mouth opening and closing). And we are similarly inclined to imitate the sounds of inanimate subjects too, an ability that often proves useful when taking a malfunctioning car to the mechanic. It seems that people produce and comprehend these sorts of iconic vocalizations so naturally that we hardly even notice as they are seamlessly woven into our speech.

Close observation reveals further that iconic vocalizations go well beyond simple pantomimic imitations of other sounds. Indeed, spoken words and phrases often take on prosodic patterns that reflect aspects of their so-called semantic meaning, often extending these iconic correspondences across modalities through processes of abstraction and metaphor. For example, one might describe “a *looong* snake,” expressing the snake’s physical length and size by iconically accenting the adjective with extended duration and low pitch. Or in contrast, the phrase “a *quicklittlebug*” might be uttered with a fast tempo and high pitch to convey the bug’s rapid movement and size. Building from such observations, scholars have recently begun to argue that these co-speech vocalizations are, in fact, the same qualitative sort of behavior as manual gestures (Emmorey, 1999; Liddell, 2003; McNeill, 2005; Okrent, 2002; Perlman, in press). Below we describe some of the empirical research on vocal gesture and discuss its implications for an integrative, simulative account of gesture and language.

Bolinger (1983, 1986) was one of the first to recognize an iconic quality to spoken prosody and to point to the close relationship between intonation and gesture. He saw intonation as “part of a gestural complex whose primitive and still surviving function is the signaling of emotion” (p. 195). Ohala (1994) too observes a link between intonation and the more primitive emotional vocalizations shared with other mammals, which is expressed in what he calls the “frequency code.” According to Ohala, high-frequency vocalizations signal apparent smallness and, by extension, non-threatening, submissive, or subordinate attitude, and low-frequency vocalizations signal apparent largeness and thus threat, dominance, and self-confidence.

Although both scholars stress the expansion of this iconic intonational system through processes of ritualization and metaphor, recent research nevertheless indicates that this view of intonation and gesture is too narrow. By focusing on the iconic expression of emotion, Bolinger and Ohala neglect intonation's incorporation into more imagistic, representational gestures.

Recent empirical research has begun to document the prevalent use of representative vocal gesture within various experimental and more naturalistic contexts. An early series of studies investigated people's production of vocal gestures, or what the authors called "analog acoustic expressions" (Shintel, Nusbaum, & Okrent, 2006). One group of participants described the movement of an animated dot moving up or down with the phrases, "It is going up" or "It is going down." Another group of participants simply read these same sentences as they were presented on a computer screen. Analysis of people's speech revealed that the final word of these phrases, "up" and "down," were spoken with a higher or lower fundamental frequency both when spoken to describe the dot or just simply read.

A second study in this series had participants describe animated dots as they moved at either a fast or slow rate to the left or right. Participants were instructed to use the phrases "It is going left" or "It is going right" to describe the dot, without explicit mention of its speed. Participants nevertheless spoke the phrase with an overall shorter duration for fast-moving dots and longer duration for the slow-moving dots. Moreover, when these descriptions were replayed for listeners to guess whether the utterance had been spoken in description of a fast or slow moving dot, their accuracy was significantly correlated with phrase duration, indicating sensitivity to the prosodic information.

A different series of experiments examined whether modulations in speech rate contribute to a speaker's mental representation of a described event as in motion or still (Shintel

& Nusbaum, 2007). Building on an experiment by Zwaan and Yaxley (2002), participants listened to sentences (e.g., "The horse is brown") spoken at a fast or slow rate and then indicated whether a pictured object had been mentioned in the sentence. Critically, the picture presented the object either in a stationary position or in motion (e.g., a horse standing still or running), with the idea that fast-spoken sentences would contribute a sense of movement in the listener's representation and facilitate responses to in-motion pictures. This prediction was confirmed as participants were faster in responding to compatible trials (fast rate-moving picture or slow rate-still picture). According to Shintel and Nusbaum, this finding suggests that speech rate (an "analog acoustic expression") can contribute to a listener's analog perceptual representation of a described event. Analogically conveyed motion information can influence listeners' representations about described objects, even when information is conveyed exclusively in the prosodic properties and not the propositional content of the sentence.

The regular use of vocal gesture has also been demonstrated in a more naturalistic setting. Perlman (in press) investigated the spontaneous use of iconic speech rate by having participants watch and describe a series of short video clips showing fast or slow-paced events. For each description that made explicit mention of speed, speech rate measurements were made for the full utterance, as well as for speed-related adverbial phrases. The analysis showed that speakers generally spoke faster or slower in their full descriptions of fast or slow events, respectively, and additionally, they spoke adverbial phrases about 'fast' events faster than adverbials about 'slow' ones.

Perlman suggests that these two separate speech rate effects may arise as the manifestation of two different sorts of simulation-related processes. The more general shift in speech rate is suggested to arise from a background simulation of the event, reflecting speakers' imaginative engagement with the tempo of the action as they proceed through their description, scanning and profiling specific details to highlight different aspects of the message. On the other hand, the adverbial phrase-specific effect is qualitatively compared to the more commonly observed manual gestures that are produced concurrently with speech. The vocal gesture emerges precisely as the speaker is conceptualizing and communicating about speed as the profiled aspect of the message. We propose here that this simulative focus is the motivating force that drives both the conventional articulatory gestures of the adverbial phrase and simultaneously, the iconic increase in the rate with which the words are spoken. These conventional and iconic forms of gesture are dynamically integrated together as they are simultaneously activated by the focused concept of speed as it is contextualized within the simulation of the event.

Yet, ongoing work in this paradigm indicates that vocal gesture is only part of the story, and that the notion of multimodal iconic gesture is probably more apt (Perlman, in preparation). In a current study, subjects come into the lab in pairs and take turns watching and describing to each other short video clips of various animals engaging in different activities. Preliminary analysis of audio and video shows that vocal gestures, including the spontaneous rhythms and intonational patterns of speech, are often performed in precise temporal and semantic coordination with iconic manual gestures. Importantly, in many instances, gestures may not be arbitrarily synchronized with speech, but rather, both gesture and speech are performed in iconic synchrony with an ongoing simulation of the event being described.

To illustrate this synchrony, consider the following excerpt from the description of a video clip of a large fish floating around in an aquarium. The fish drifts up to the surface of the water and then suddenly gulps down a bug.

(1) it was this big fish [kind of hanging out, **he was floating slowly up to the top** and he **ate some**…thing…]
(1) (2)

*Preparation: raises and holds right hand with thumb and fingers pinched together as a fish and its mouth.*
*Manual iconic (1): right hand rises slowly upward and pauses*
*Manual iconic (2): on "ate" the right hand thrusts forward, thumb and fingers spreading open and closing like the mouth of the attacking fish. On "some" the fingers and hand retract back and are held to the end of the utterance.*
*Vocal iconic (1): speech is slowed down and low in intensity reflecting the fish's manner of floating to the surface. The slowing is most marked in extended duration of the vowel /o/ in the adverb "slowly." Pitch steadily rises, peaking on the word "top."*
*Vocal iconic (2): speech suddenly increases in tempo and intensity in synchrony with the stressed syllable of "ate."*

This person's description demonstrates the multimodal nature of gesture as a seemingly single unified expression incorporates iconic manual and vocal gestural elements with the conventional articulatory gestures of speech. It is additionally interesting to note that the sudden nature of the fish's attack, although it is quite apparent in the corresponding manual and vocal gestures, is not entirely discernible from just the semantics of the words. A clue, however, is provided by the grammar of the utterance, in which the temporally extended 'float,' expressed by the progressive aspect, is contrasted to the punctuated 'eat,' expressed in the past tense. Thus, corresponding elements of the fish's manner of motion, the temporal contour of the motion, and even perhaps its upward direction are all manifested synchronously within iconic manual gesture, iconic prosody, lexical semantics, and even higher-order syntactical structures.

Unlike manual gestures (with the exception of when they are integrated with signing), vocal gestures are, in a sense, parasitic on the phonological form of an utterance. That is to say, these gestures must be interwoven within the phonological material provided by the forms of spoken words. How is it that particular segments, during the online moments of speech production, are modified to accentuate certain iconic qualities? One possibility is that phonological aspects of a word form maintain some latent potential to take on a quality of iconicity, which may, in some instances become activated in relation to the contextualized dynamics of an utterance. This idea borrows from Wilcox's (2004) theory of cognitive iconicity and Mueller's (2008) notion of activation as it plays out in the triadic structure of metaphor.

Cognitive iconicity adopts Langacker's (1987) claim that semantic and phonological poles (i.e., semantic meaning and phonological form) each reside within semantic space, which is itself a subset of the full expanse of conceptual space. Wilcox describes that, "The phonological pole reflects our conceptualizations of pronunciations, which range from the specific pronunciation of actual words in all their contextual richness to more schematic conceptions, such as a common phonological shape shared by all verbs, or a subset of verbs, in a particular language" (p. 122). Wilcox explains that cognitive iconicity is not an objective similarity relation between a form and its signified referent, but rather a constructed relation between two structures in a multidimensional conceptual space. He also notes that metaphor can act as a "worm hole" through this space, functioning to shorten the distance between the phonological and semantic poles.

A similar schema exists in the triadic structure of metaphor, which involves two meaning structures of source and target and a relation between them. Considering a traditional view distinguishing dead and living metaphors, Mueller (2008) proposes a dynamical alternative in which the metaphorical relation between two concepts does not need to be fully active or fully opaque, but instead can be activated to a greater or lesser degree with the dynamics of each instantiated usage. Applying this framework to the triadic structure of cognitive iconicity, it follows that the iconic relation between the semantic and phonological poles may lie dormant and become more or less activated dependent on the dynamics of usage. When activated, these iconic relations become accentuated and take form as vocal gesture.

Consider for example, the saliently extended duration of /o/ in the pronunciation of the word "slooowly" in the above example (1). In this case, the duration of the /o/ was particularly activated, observable through its exaggeration. However, for comparison, consider how one might articulate the word "slowly" in the phrase, "a ssloowly sslitherinng sssnake." In this case the /o/ is still articulated with some extended duration, but in addition, the frication of the /s/ is extended too, in part because of the alliteration, but surely also in part to the onomatopoeic hiss that we associate with snakes. Thus one can see how iconic relations between a phonological form and an aspect of its meaning might become differentially activated within each dynamical usage.

Vocal gestures seen in examples like (1) above offer a special window into the full extent to which speech and gesture are integrated together as they manifest from the same simulative processes. Furthermore, these gestures imply that the embodied representations that arise during these simulative processes are profoundly multimodal. Various concepts, such as those related to speed, manner of movement, size, and verticality, are spontaneously embodied in the movements of our hands and also in the movements of our vocal tract, and indeed, in the right context, body parts ranging across much of our anatomy. The frequent and casual use of iconic vocal gestures in particular suggests that humans have a special knack for conceiving of their experience in terms of iconic movements of the vocal tract, often through cross-modal abstractions and metaphor. As McNeill (1992: 12), puts it, “Gestures are like thoughts themselves,” pointing to the idea that the embodiment of thoughts through gesture is an essential aspect of the very nature of how humans think. Vocal gestures may also be thoughts with conventional linguistic gestures reflecting conventionalized aspects of embodied thought patterns

**Evolutionary Perspective on Gesture as Simulation**

We now change gears slightly to consider, from the perspective of the sensorimotor simulation hypothesis, the evolutionary root of language and gesture. Our goal with presenting this evolutionary perspective is to try and gain insight into the more fundamental processes that might be involved with sensorimotor simulations and the bodily activations that arise from them.

If, as we argue, language and gesture are produced and understood through the bodily movements arising from sensorimotor simulations, then what might be the evolutionary origin of this communicative ability? Can we observe in our great ape cousins the precursor of representative gesture as sensorimotor simulation? Starting from the simulation hypothesis, one might begin by making some predictions about the characteristics and qualities of the gestures we would expect to find. Generally one might expect the most phylogenetically primitive representative gestures to be those that manifest from the most imaginatively simple of simulations. Thus, for example, the simulation ought to involve the gesturer’s own perspective and body, and not someone else’s, let alone the gesturer’s body imagined as something entirely different. In addition, one would expect the contextual and afforded elements of the simulation to be largely present in the gesturer's immediate perceptual experience-- the more distant the element, presumably the more difficult it would be to imagine. In sum, primitive representative gestures ought to be tightly connected—in form, meaning, and context—to the presently afforded instrumental and attentive actions that are available to the gesturer. Our following review of some of the existing reports of spontaneous iconic gestures by the great apes shows that, by and large, these gestures do in fact tend to exhibit these sorts of properties (Crawford, 1937; Köhler, 1925; Savage-Rumbaugh, Wilkerson & Bakeman, 1977; Tanner, Patterson, & Byrne, 2006; Tanner & Byrne, 1996).

Kendon (1991), in his examination of the origins of representative action, begins by directing his reader to the classic work of Köhler (1925) and his observations of the partial enactments of imagined actions by chimpanzees. Kendon points to this work, as we follow suit, to emphasize the tight relationship between an instrumental action, the mental imagining of that action, and the partially enacted gestures that manifest during the imagined act. Almost a century ago, Köhler observed that some chimpanzee gestures originate from a process that we might in present terms refer to as simulated action. He describes several instances in which partial actions arise in non-communicative contexts when the chimpanzee is merely imagining his participation in an action, such as when he observes another performing a well-rehearsed act.

Köhler captures one interesting example of this in a photograph (Plate IV), in which a chimpanzee, Sultan, exhibits what Köhler calls a "sympathetic left hand." The chimp had earlier mastered the skill of stacking up a series of wooden boxes to acquire a highly desired banana hanging from the ceiling. In the photograph, Sultan is caught observing another chimpanzee accomplishing the same task just at the moment when the second chimp has reached the top and is grabbing for the banana. Sultan is seen staring intently up at the scene, his left hand reaching into the air. Though we of course cannot know for sure what Sultan was thinking about, it is certainly plausible to consider that the chimp was imagining himself up on top of the box grabbing for that banana, just as he had practiced many times before. Relevant to the simulation hypothesis, notice that the task was both highly familiar to Sultan as well as fully present perceptually. Moreover, one can only presume that the highly desirable nature of the food served to entice Sultan's imagination into action.

Importantly, Köhler also observed that partial actions sometimes arise from imagined actions within the context of social interaction. He writes,

> [A] considerable proportion of all desires is naturally expressed by slight initiation of the actions which are desired. Thus, one chimpanzee who wishes to be accompanied by another, gives the latter a nudge, or pulls his hand, looking at him and making the movements of 'walking' in the direction desired. One who wishes to receive bananas from another initiates the movement of snatching or grasping, accompanied by intensely pleading glances and pouts. … In all cases their mimetic actions are characteristic enough to be distinctly understood by their comrades" (Köhler, 1925: 307-308).

In each of these circumstances, it appears that the gesturing chimp desires an interactive outcome and partially enacts a gesture that, if it were carried out to instrumental completion, would function to bring the outcome into being. Indeed, this ability to imagine the performance of an instrumental act upon a social interactant, but then to partially inhibit that act's performance, appears crucial in the origin of these possible precursors of representative gesture. And as Köhler notes, comprehension of such partial actions comes naturally, facilitated by their mimetic resemblance and presumably their contextual relevance to an afforded instrumental action.

More recent studies have also reported the use of spontaneously produced iconic gestures by great apes, often occurring in contexts of tactful social engagements in which one ape is trying to influence the movement and position of an interlocutor (Savage-Rumbaugh et al., 1977; Tanner & Byrne, 1996). For example, Savage-Rumbaugh and colleagues documented the gestures uses by bonobo chimpanzees as they coordinated copulatory positions with one another (promiscuous sex is a common behavior of bonobos). They observed that many of the gestures that immediately preceded copulatory bouts bore an iconic quality and could generally be placed into three categories: positioning motions (actual physical, gentle movements to move the recipient's body or limbs), touch plus iconic hand motions (limb or body part is lightly touched and then movement is indicated by an iconic hand motion), and iconic hand motions (simply indicates via an iconic hand movement, without touch). Interestingly, this ordering of increasing abstraction correlated negatively with the gestures' frequency of occurrence, suggesting that those gestures closest to instrumental action were the easiest to perform. One could reason that the more abstractly iconic gestures, further removed from the immediate context of instrumental action, would place a greater load on the imagination.

Another study by Tanner and Byrne (1996) observed the use of iconic gestures between two captive gorillas at the San Francisco Zoo. In this case, a 13-year-old adult male Kubie was recorded using a variety of iconic tactile and visual gestures to encourage play and direct interactive movement with Zura, a 7 year-old-female. Of particular interest here, Tanner and Byrne note how special conditions in the zoo enclosure made coercion ineffective and propose that these conditions were a motivating force for the production of the gestures. For instance, a door to an indoor pen was opened wide enough to allow Zura to fit through but not her larger companion, permitting Zura to escape if Kubie was too forceful. Additionally, a second, older silverback male was part of the troop, which meant that Kubie had to be especially charming so as to engage Zura without drawing the other silverback's attention. Thus, again, we find that iconic gestures arise in contexts closely connected with instrumental action, in which one apes desires to bring about a certain outcome with an interactant, but must exercise restraint for purposes of social tact.

Evidence shows also that with rich human social interaction and enculturation, apes develop a markedly expanded capacity to produce more abstracted iconic gestures (Tanner, Patterson, Byrne, 2006). In many of these cases, the expansion of the imagination and its distinct role within particular gestures is obvious. For example, consider some of the iconic gestures produced by the language-trained bonobos Kanzi and Mulika in which they "made twisting motions toward containers when they needed help in opening twist-top lids" and "hitting motions toward nuts they wanted others to crack for them" (Savage-Rumbaugh et al., 1986: 218). Both of these gestures differ from those described above in how they incorporate the imagined physical manipulation of objects that are not available to immediate tactile experience. Notably though, the simulated objects *are* available to visual perception, and as Kanzi and Mulika perform these gestures, it sounds from the preposition "through" that their visual attention is clearly drawn towards these objects. One might wonder whether they could produce such sophisticated iconic gestures with the same degree of facility if the objects were located outside of their perceptual purview. (See Kohler 1925 for interesting examples from the domain of problem solving in which direct perceptual access to the elements involved in a solution is necessary for a chimp to conceive of the solution.)

Finally, a simulation-based account of gesture has implications for more domain-general cognitive processes, the evolution of which is typically assumed a necessary precondition for the use of iconic gestures. A traditional view on iconic gesture assumes that their production and comprehension depends on highly developed cognitive abilities related to imitation and theory of mind. For example, Tomasello reasons, “To use an iconic gesture one must first be able to enact actions in simulated form [a more deliberate notion of simulation than we intend in our usage], outside their normal instrumental context—which would seem to require skills of imitation, if not pretense. But even more importantly, to comprehend an iconic action as a communicative gesture, one must first understand to some degree the Gricean communicative intention; otherwise the recipient will suppose that the communicator is simply acting bizarrely, trying to run like an antelope or to dig a hole for real when the context is clearly not appropriate” (2008: 203). This line of reasoning, in combination with empirical research, has led researchers such as Tomasello and his colleagues, to dismiss on a priori grounds the possibility that the great apes use iconic gestures (Call & Tomasello, 2007; Pika, 2007; Tomasello, 2008). They argue that the great apes have only minimal abilities to imitate and to share communicative intentions, and thus they simply cannot use iconic gestures.

However, from the perspective of the simulation hypothesis, the use of iconic gestures, although spontaneous, creative, and socially-minded, does not require an ability for deliberate imitation and pantomime or hard Gricean social-cognitive reasoning. As we have seen above, reports of iconic gestures by the great apes describe them as used, not “outside their normal instrumental context,” but directly within it. Moreover, these iconic gestures are comprehended without necessarily a reflective understanding of the gesturer’s “communicative intention,” but more directly through an activated sense of the full action within a context that is rife with expectation of exactly that sort of action. We suggest that these gestures appear to reflect an emerging ability to perform increasingly imaginative, sensorimotor simulations and to modulate their iconic motor activations towards communicative expression. According to the simulation hypothesis, this capacity is interwoven at the core of the cognitive skills leading to the origin of our own, dramatically more sophisticated system of socially-tuned simulations which are foundational in the motivation of our gesture and language.

## Conclusion

Our main hypothesis is that experiential simulations of sensorimotor imagery are fundamental to the conceptual processes that underlie the use of gesture and language and the construction of meaning during conversational interaction. According to this view, one's ability to interpret meaning during conversation resides largely in the ability to simulate the thoughts and ideas of one's interlocutor through the expressive movements of their speech and gesture. These simulative processes are also involved in the production of language and gesture. Articulatory movements, whether of the vocal tract, the hands, or potentially any other part of the body, are produced by the activations that arise during a sensorimotor simulation. Indeed, these bodily activations are, in the sense of Vygotsky's (1986) notion of a "material carrier," an essential aspect of the thought itself. An important implication of this idea is that our embodied, sensorimotor experience plays a crucial role in the formation of the concepts and the meanings we construct and express during the online moments of conversation. Critically, these concepts and meanings appear to be interwoven across modalities, and often involve the creation of schematic and metaphorical cross-modal correspondences.

Although we do not, as yet, have a firm idea on all of the constraints on simulation processes, and the extent to which they create simplified, as opposed to complex, meanings, we suggest that these processes include aspects of full-bodied experiences, and are critical to understanding the minds of others. People are likely to be quite flexible in the level of details they create during sensorimotor simulation, depending on their immediate motivations and goals, the social context, the linguistic material to be understood, and the task. The process of constructing sensorimotor simulations is constrained similarly as are other fundamental cognitive operations in the pursuit of meaning. People will create simulations rich enough to enable them to infer sufficiently relevant meanings and impressions, while also trying to minimize the cognitive effort needed to produce meaningful effects. In some cases, the meanings, emotions, and impressions one infers when understanding a speaker's utterance will be relatively crude, primarily because this set of products will be "good enough" for the purposes at hand. At other times, people may engage in more elaborate, even highly strategic simulation processes as they tease out numerous meanings and impressions from an utterance in context, such as when reading novel metaphors in poetry. Interestingly, these more elaborate instances of communication often seem to be ones that invoke more richly iconic expression and interpretation.

Finally, there is one important challenge that must be addressed to achieve a more unified understanding of how sensorimotor simulations figure into the production and comprehension of language and gesture. This challenge rests in the need to resolve the qualitative distinctions that are often assumed to distinguish linguistic communication from gesture and other so-called paralinguistic forms of expression. On the surface, language appears to be a completely different sort of behavior from gesture, and language scholars have long believed that its use depends on specialized cognitive processes. More specifically, language is typically characterized as a conventional system of discrete, arbitrary forms that are strung together by the phonological and syntactic rules that comprise duality of patterning (e.g., Hockett, 1960; Jackendoff, 1994; Pinker, 1994). In contrast, gesture appears be almost diametrically opposite. Gestures are idiosyncratic, iconic, and analog in form; they lack syntactic combinatorial rules and cannot be analyzed into anything resembling phonological components (McNeill, 1992). Gestures seem naturally molded to our thoughts and it is intuitive how they might manifest directly from the bodily activations of sensorimotor simulations. Language, on the other hand, is generally thought to be a symbolic code for thought and thus to involve processes of encoding/decoding or often “mapping” (e.g., Glenberg & Kaschak, 2002) between thought and linguistic form.

Though on the surface these distinct properties may appear to reflect differences of quality, a new framework has emerged that instead considers them as a set of continua (McNeill, 1992; 2005). The properties of language and gesture occupy opposing prototypical ends of these continua and the intermediate properties of various other forms of communication, such as pantomimes and emblems, are positioned in between. Building on this perspective, evidence suggests that the most critical difference between language and gesture may relate more to conventionality than to any particular formal differences intrinsic to language *per se*. Under this view, the formal properties associated with language—duality of patterning, arbitrary symbolism, and categorical form—may simply be emergent properties of functional constraints on a conventional communication system.

Crucial to this idea are various empirical observations demonstrating how a functional need to establish conventional communication appears to lead quite naturally to the properties characteristically associated with language. For example, deaf children raised without access to a system of sign language naturally create their own linguistic homesign system (Goldin-Meadow & Feldman, 1977). Over a child's development, initially idiosyncratic iconic gestures become conventionalized into more discrete word-like forms, which the child combines together by simple syntactic rules. In fact, this pattern is so robust that it can be induced within just a few minutes in a laboratory setting. McNeill (1992) describes a study in which speakers told fairytales to a partner, but were permitted to use only manual gestures and no words. McNeill notes how "Within 15 or 20 minutes a system has emerged that includes segmentation, compositionality, a lexicon, a syntax with paradigmatic oppositions, arbitrariness, distinctiveness, standards of well-formedness, and a high degree of fluency" (p. 66).

Other evidence for this natural pathway between language and gesture comes from studies documenting the residual iconicity found in sign languages. One study examining 1,944 signs in Italian Sign Language found that fifty percent of handshape occurrences and sixty-seven percent of body locations had clear iconic motivations (Pietrandrea, 2002). Indeed, sign language scholars have traced some of the historical routes by which conventional signs originate from iconic gesture (Wilcox, 2004). Furthermore, the fact that a substantial amount of this iconicity persists even in mature languages suggests that it continues to play an active role in their ongoing development. Although it is less transparent, there is also evidence to suggest that vocally-based iconicity similarly persists to some degree in spoken languages (Hinton, Nichols, & Ohala, 1994).

Perhaps the most crucial difference between gesture and language may turn out mostly to be a matter of convention. (This is not to say, of course, that humans do not have a special knack for acquiring the conventional; clearly, conventional actions and behavioral routines of all sorts abound in human culture.) As we have seen, this quality of convention is fluid and, importantly, bidirectional. Gestures can become more linguistic under certain functional constraints, and likewise, in certain contexts, linguistic forms can become more like gesture. Such is the case with poetry, both spoken and signed, and more mundanely, it is the case with prosodic vocal gestures. Given this slippery distinction, it seems perhaps less plausible to view language from the standard notion as a symbolic, arbitrary code, into which thoughts are encoded, transmitted, and then decoded back. A more parsimonious account might consider instead that language is produced and understood in the same way as gesture, with the only significant difference being that linguistic gestures arise from the conventionalized aspects of simulated sensorimotor imagery. Moreover, it takes only casual observation to witness the completely ordinary way in which language and gesture are seamlessly integrated with pantomimes and emblems, as well as music, dance and everything in between. Underlying it all, we argue, is the ability to interactively engage our minds and bodies in imaginative simulations of sensorimotor imagery.

# References

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology, 71*, 230-244.

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

Bolinger, D. (1983). Where does intonation belong? *Journal of Semantics, 2,* 101-120.

Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Palo Alto, CA: Stanford University Press.

Browman, C.,, & Goldstein, L. (1995). Dynamics and articulatory phonology. In T. van Gelder & R. Port (Eds.), *Mind as Motion* (pp. 175-193). Cambridge, MA: MIT Press.

Call, J. & Tomasello, M. (2007). *The gestural communication of apes and monkeys.* London: Lawrence Erlbaum Associates, Publishers.

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J.C., & Lenzi G.L (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Science USA, 100*:5497-5502.

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

Clark, H. H. & Gerrig, R. J. (1990). Quotations as demonstrations. *Language, 66,* 764-805.

Cook, S. W. & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition, 113*, 98-104.

Crawford, M. P. (1937). The cooperative solving of problems by young chimpanzees. *Comparative Psychology Monograph, 14,* 1-88.

Emmorey, K. 1999. Do Signers Gesture? In L. Messing & R. Cambell (eds.). *Gesture, Speech,* *and Sign* (pp. 133-159). New York: Oxford University Press.

Fowler, C. (1987). Perceivers as realists; talkers too. *Journal of Memory & Language, 26,* 574-587.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*, 361-377.

Gibbs, R.W., Jr. (2006a). *Embodiment and cognitive science.* New York: Cambridge University Press.

Gibbs, R. (2006b). Metaphor interpretation as embodied simulation. *Mind & Language, 21*, 434-458.

Gibbs, R. (Ed.) (2008). *Cambridge handbook of metaphor and thought*. New York: Cambridge University Press.

Gibbs, R, Gould, J., & Andric, M. (2006). Imagining metaphorical actions: Embodied simulations make the impossible plausible. *Imagination, Cognition, & Personality, 25*, 221-238.

Gibson, J. (1979). *An ecological approach to visual perception*. Houghton Mifflin: Boston.

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

Goldin-Meadow, S. & Feldman, H. (1977). The development of language-like communication without a language model. *Science, 197,* 401-403.

Hadar, U., Burstein, A., Krauss, R. M., & Soroker, N. (1998). Ideational gestures and speech: A neurolinguistic investigation. *Language and Cognitive Processes, 13,* 59-76.

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

Hinton, L., Nichols, J. & Ohala, J. J. (1994). *Sound symbolism*. Cambridge: Cambridge Univesity Press.

Hockett, C. (1960). The origin of speech. *Scientific American, 203*, 89-97.

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

Hostetter, A. B. & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15,* 495-514.

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

Jackendoff, R. (1994). *Patterns in the mind: Language and human nature.* USA: Basic books.

Kendon, A. (1991). Some considerations for a theory of language origins. *Man, 26*, 602-619.

Klatzky, R. L., Pellegrino, J.W., McCloskey, B.P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language, 28*, 56-77.

Kohler (1925). *The mentality of apes.* London: Routledge and Kegan Paul.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh. The embodied mind and its challenge to Western thought*. New York: Basic Books.

Lane, H. (1984). *Where the mind hears: A history of the deaf.* New York: Random House.

Langacker, R. W. (1987). *Foundations of cognitive grammar, vol. 1: Theoretical foundations.* Stanford: Stanford University Press.

Lausberg, H. & Kita, S. (2003). The content of the message influences the hand choice in co-speech gestures and in gesturing without speaking. *Brain and Language, 86,* 57-69.

Lavergne & Kimura (1987). Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia, 25,* 689-693.

Liddell, S. (2003). Sources of meaning in ASL classifier predicates. In K. Emmorey (ed.), *Perspectives on Classifier Contsructions in Sign Languages* (pp. 199-219). Mahway, NJ: Erlbaum.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition, 32*, 1389-1400.

McNeill, D. (1992). *Hand in mind*. Chicago: University of Chicago Press.

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

Mueller, C. (2008). *Metaphors dead and alive, sleeping and waking.* Chicago: University of Chicago Press.

Niedenthal, P.M., Brauer, M., Halberstadt, J.B., & Innes-Ker, A.H. (2001). When did her smile drop? Contrast effects in the influence of emotional state on the detection of change in emotional expression. *Cognition and Emotion, 15*, 853-864.

Noe, A., & O'Regan, K. (2002).On the brain-basis of visual consciousness: A sensorimotor account. In A. Noe and E. Thompson (Eds.), *Vision and mind* (pp. 367-398). Cambridge, MA: MIT Press.

Ohala, J.J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In Hinton, Nichols, and Ohala (Eds.), *Sound Symbolism* (pp. 325-347). Cambridge University Press: Cambridge.

Okrent, A. (2002). A modality-free notion of gesture and how it can help us with the morpheme vs. gesture in question in sign language linguistics. In R. P. Meier, K. Kormier, & D.

Quinto-Pozos (eds.) *Modality and Structure in Signed and Spoken Language* (pp. 175-198). Cambridge: Cambridge University Press.

Perlman, M. (in press). Talking fast: The use of speech rate as iconic gesture. In F. Perrill, V. Tobin, & M. Turner (eds.) *Meaning, form, and body.* Stanford: CSLI Publications.

Pietrandrea, P. (2002). Iconicity and arbitrariness in Italian Sign Language. *Sign Language Studies, 2,* 296-321.

Pika, R. (2007). Gestures in subadult gorillas (*Gorilla gorilla*). In J. Call & M. Tomasello (eds.) *The gestural communication of apes and monkeys.* (pp. 41-67). New York: Lawrence Erlbaum Associates.

Pinker, S. (1994). *The Language Instinct.* New York: William Morrow.

Rauscher, F. B., 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. & Matlock, T. (2007) The integration of figurative language and static depictions: An eye movement study of fictive motion, *Cognition*, *102*, 129-138.

Rime, B., Schiarapture, L., Hupet, M., & Ghysselinkckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation & Emotions, 8,* 311-325.

Savage-Rumbaugh, E. S., Wilkerson, B. J., & Bakeman, R. (1977). Spontaneous gestural communication among conspecifics in the pygmy chimpanzee (Pan paniscus). In G. H. Bourne (Ed.) *Progress in Ape Research* (pp. 97-116). New York: Academic Press Inc.

Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rupert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus). Journal of Experimental Psychology: General, 115,* 211-235.

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition*, 105, 681-690.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*, 167-177.

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

Tanner, J. E. & Byrne, R. W. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Current Anthropology, 37,* 162-173.

Tanner, J. E., Patterson, F. G. and Byrne, R. W. (2006). The development of spontaneous gestures in zoo-living gorillas and sign-taught gorillas: from action and location to object representation. *Journal of Developmental Processes, 1*, 69-102.

Tomasello, M. (2008). *Origins of human communication.* Cambridge: MIT Press.

Vallee-Tourangeau, F., Anthony, S.H. & Austin, N.G. (1998). Strategies for generating multiple instances of common and ad hoc categories. *Memory, 6,* 555-592.

Vygotsky, L. S. (1986). *Thought and Language.* Edited and translated by E. Hanfmann and G. Vakar, revised and edited by A. Kozulin. Cambridge: MIT Press.

Wilcox, S. (2004). Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed languages. *Cognitive Linguistics, 15,* 119-147.

Wilson, N., & Gibbs, R. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science, 31,* 721-731.

Wu, L., & Barsalou, L. (2001). Grounding concepts in perceptual simulation: I. Evidence from property generation.Unpublished manuscript.

Zwaan, R. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B.H. Ross (Ed.) *The psychology of learning and motivation* (pp 35-62). Academic Press, New York.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science, 13,* 168-171.