Gesture as simulated action: Revisiting the framework

Autumn B. Hostetter¹ · Martha W. Alibali²

Published online: 3 December 2018
© Psychonomic Society, Inc. 2018

Abstract

The Gesture as Simulated Action (GSA) framework was proposed to explain how gestures arise from embodied simulations of the motor and perceptual states that occur during speaking and thinking (Hostetter & Alibali, *Psychonomic Bulletin & Review*, 15, 495–514, 2008). In this review, we revisit the framework's six main predictions regarding gesture rates, gesture form, and the cognitive cost of inhibiting gesture. We find that the available evidence largely supports the main predictions of the framework. We also consider several challenges to the framework that have been raised, as well as several of the framework's limitations as it was originally proposed. We offer additional elaborations of the framework to address those challenges that fall within the framework's scope, and we conclude by identifying key directions for future work on how gestures arise from an embodied mind.

Keywords Gesture · Embodied cognition · Imagery · Action · Language production

A growing number of scholars are emphasizing the critical role of the body in cognitive processes (e.g., Glenberg, Witt, & Metcalfe, 2013). Although the embodied-cognition framework is not without its critics (e.g., Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016), it provides an account of cognition and behavior in many domains, including language (e.g., Glenberg & Gallese, 2012), mathematical reasoning (e.g., Hall & Nemirovsky, 2012; Lakoff & Núñez, 2000), and music perception (e.g., Leman & Maes, 2014). One source of evidence for the embodiment of cognition in these domains is the way that people use their bodies as they describe their thinking; that is, when people describe their mathematical thinking or their musical perception, they use their hands and arms to depict the ideas they express. Such movements (hereafter, *gestures*) seem to indicate that the hands naturally reflect what the mind is doing. Indeed, decades of research have shown that gestures are intricately tied to language and thought (e.g., Goldin-Meadow, 2003; McNeill, 1992).

Hostetter and Alibali (2008) proposed the Gesture as Simulated Action (GSA) framework as a theoretical account

of how gestures arise from an embodied cognitive system. Drawing on research from language, mental imagery, and the relations between action and perception, the GSA framework proposes that gestures reflect the motor activity that occurs automatically when people think about and speak about mental simulations of motor actions and perceptual states. The framework has been influential since its publication. It has been applied to explain a range of findings about gesture (e.g., Wartenburger et al., 2010), and it has inspired models that explain the embodiment of other sorts of behaviors (e.g., Cervasco & Ramos, 2013; Chisholm, Risko, & Kingstone, 2014; Perlman, Clark, & Johansson Falck, 2015). Furthermore, the central idea proposed in the GSA framework—that gestures reflect embodied sensorimotor simulations—has been taken as a warrant for using gestures as evidence about the nature of underlying cognitive processes or representations in a wide range of tasks and domains (e.g., Eigsti, 2013; Gerofsky, 2010; Gu, Mol, Hoetjes, & Swerts, 2017; Perlman & Gibbs, 2013; Sassenberg, Foth, Wartenburger, & van der Meer, 2011; Yannier, Hudson, Wiese, & Koedinger, 2016). But, are such claims warranted? What is the evidence that gestures actually do express sensorimotor simulations in the ways specified in the GSA framework?

The purpose of this review is to revisit the GSA framework, with an eye toward both examining the evidence to date for its predictions and identifying its limitations. In the sections that follow, we first summarize the central tenets of the GSA framework. We then review the evidence for and against each of its

✉ Autumn B. Hostetter
autumn.hostetter@kzoo.edu

¹ Department of Psychology, Kalamazoo College, Kalamazoo, MI, USA

² Department of Psychology, University of Wisconsin, Madison, WI, USA

predictions. In light of this groundwork, we address the primary challenges that have been raised against the framework. We conclude with a discussion of limitations and future directions.

Overview of the Gesture as Simulated Action framework

The GSA framework situates gesture production within a larger embodied cognitive system. Put simply, speakers gesture because they simulate actions and perceptual states as they think, and these simulations involve motor plans that are the building blocks of gestures. The GSA framework was developed specifically to describe gestures that occur with speech, although the basic tenets also apply to gestures that occur in the absence of speech (termed *co-thought* gestures; Chu & Kita, 2011).

The term *simulation* has been used in a variety of contexts within cognitive science and neuroscience. Simulation has been contrasted with motor imagery (e.g., O'Shea & Moran, 2017; Willems, Toni, Hagoort, & Casasanto, 2009) and with emulation (e.g., Grush, 2004), and there is debate about the roles of conscious representations and internal models in simulation (e.g., Hesslow, 2012; Pezzulo, Candidi, Dindo, & Barca, 2013). When we use the term *simulation*, we are not taking a stance regarding these issues; by *simulation* we simply mean the activation of motor and perceptual systems in the absence of external input. We do assume that simulations are predictive, in that they activate the corresponding sensory experiences that result from particular actions (see Hesslow, 2012; Pouw & Hostetter, 2016), but the framework does not hinge on either whether simulations are conscious or whether they are neurally distinct from mental imagery.

The basic architecture of the GSA framework is depicted in Fig. 1. According to this framework, the likelihood of a gesture at a particular moment depends on three factors: the producer's mental simulation of an action or perceptual state, the activation of the motor system for speech production, and the height of the producer's current gesture threshold. We consider each of these factors in turn.

First, if a speaker is not actively simulating actions or perceptual states as she speaks, a gesture is very unlikely to occur. This tenet is based on the idea that information can be represented in either a symbolic, verbal/propositional form or in a grounded, visuospatial or imagistic form (e.g., Paivio, 1991; Pecher, Boot, & Van Dantzig, 2011; Zwaan, 2014). The GSA framework contends that thinking that involves visuospatial or motor imagery requires activation of the motor system. Just as perception and action are intricately linked in cognition more generally (e.g., Gibson, 1979; Prinz, 1997; Proffitt, 2006; Witt, 2011), people automatically engage the motor system when they activate mental images—for example, when thinking about how they would interact with imagined objects or imagining themselves moving in ways that embody particular

actions. These simulations involve motor plans that can come to be expressed as gestures. In contrast, when speakers rely instead on a stored verbal or propositional code as the basis for speaking or thinking about an idea, their motor systems are activated to pronounce those codes as words, but the speakers are unlikely to produce gestures, because their motor systems are not actively engaged in simulating the spatial or motoric properties of the idea.

Under this view, the motor plan that gives rise to gesture is formed any time a thought involves the simulation of motor or perceptual information. However, people are most likely to produce gestures when they also express their thoughts in speech. Thus, the second factor that influences gesture production is the concurrent activation of the motor system for speech production. Although co-thought gestures can and do occur (e.g., Chu & Kita, 2016), the motor activation that accompanies simulations of actions and perceptual states is more likely to be expressed overtly in gestures when the motor system is also engaged in producing speech. Because the hands and the mouth are linked, both developmentally (e.g., Iverson & Thelen, 1999) and neurally (e.g., Rizzolatti et al., 1988), it may be difficult to initiate movements of the mouth and vocal articulators in the interest of speech production without also initiating movements of the hands and arms. When there is no motor plan that enacts a simulation of what is being described, this manual movement might take the form of a beat gesture (Hostetter, 2008), or a simple rhythmic movement that adds emphasis to a particular word or phrase. However, when there is an active, prepotent motor plan from the simulation of an action or perceptual state, this motor plan may be enacted by the hands and become a representational gesture. Thus, the simultaneous activation of the speech production system influences the likelihood of a gesture. Although gestures can occur in the absence of speech, they should be more prevalent when people activate simulations along with speech. This idea that it is difficult to inhibit the motor activity involved in simulation while simultaneously allowing motor activity for articulation has since been integrated into several other models that seek to explain language production more broadly (e.g., Glenberg & Gallese, 2012; Pickering & Garrod, 2013).

Third, even when the basic motor plan for a gesture has been formed, the plan will only be realized as gesture if the motor system has been activated with enough strength to surpass the speaker's current *gesture threshold*. According to the GSA framework, the gesture threshold reflects a speaker's own resistance to overtly producing a gesture. The height of this threshold can depend on a variety of dispositional and situational factors. For example, some people may hold strong beliefs that gestures are impolite or distracting, leading them to maintain a consistently high threshold. For such individuals, only very strongly activated motor plans will surpass the threshold and be realized as gesture. In contrast, in some

Fig. 1 Architecture of the GSA framework

situations speakers may believe that a gesture is likely to be particularly effective for the communicative exchange or for managing their own cognitive load, leading them to maintain a lower threshold temporarily, so that even a small amount of motor activation is strong enough to surpass the threshold. In sum, whether an individual will produce a gesture along with a particular utterance or thought depends on the dynamic relationship between how strongly the underlying simulation activates the individual's motor system and how high the individual's gesture threshold is at that moment.

These mechanisms give rise to several predictions regarding the rate, form, and cognitive cost of gestures. We next describe the six predictions made in the original statement of the framework (Hostetter & Alibali, 2008), and we review the relevant evidence to date for each.

Review of evidence

Prediction 1: Speakers gesture at higher rates when they activate visuospatial or motor simulations in service of language production than when they do not activate such simulations

The GSA framework contends that speakers form both imagistic representations and verbal/propositional representations, and that either can form the basis of an utterance. In some

situations, speakers activate *images*, or grounded modal representations that resemble the experiences they represent. In other situations, speakers may activate abstract, amodal representations that symbolically represent the message, or they may activate verbal codes that instantiate the message.¹ When imagistic representations are activated, the motor and perceptual systems become activated, because sensorimotor properties of the image are simulated. By this, we mean that the same neural areas that are involved in producing the action, interacting with the object, or experiencing the perceptual state are activated. In the case of motor imagery, the motor system is activated as if to produce the action. In the case of visual imagery, the visual system is activated, but that activation can spread to motor areas as well, because of the close ties between perception and action (e.g., Prinz, 1997).

The key prediction stemming from this claim is that speakers should gesture more often when their speech is based on imagery than when their speech is based on verbal or propositional information. At the time the GSA framework was formulated, there was already some evidence for this claim (e.g., Hostetter & Hopkins, 2002). In the years since, additional studies have tested the claim more definitively. Hostetter and Skirving (2011) compared the gestures produced by speakers as they described vignettes that they had learned in one of two different ways: by listening to a description twice, or by listening to a description once and then watching the events in an animated cartoon. They found higher rates of gesture when speakers had seen the cartoon than when they had only heard the verbal description. Importantly, because speakers in both conditions heard words describing the vignette, the evidence favors the interpretation that speakers gesture more in the presence of a visuospatial image, rather than in the absence of easy lexical access to the words.

¹ Note that verbal representations may also be “simulated,” by activating the sensorimotor system as if to orally pronounce the words (e.g., Topolinski & Strack, 2009). For example, one might have the motor or auditory experience of pronouncing or hearing the word as if it were spoken. To the extent that such simulations activate the motor system with enough strength to pass threshold, they may also be produced as gesture. For example, thinking about how to pronounce the German *ii* could be accompanied by pursing of the lips.

However, not all studies have supported the prediction. Parrill, Bullen, and Hoburg (2010) found that speakers both gestured and spoke more when they had seen a cartoon than when they had only read a description of its events, and speakers' gesture rates (controlling for the amount of speech) was no higher when describing the cartoon than when describing the verbal text. This is in direct contrast to the results of Hostetter and Skirving (2011), who found a difference in gesture rate, even when controlling for the amount of speech.

Several methodological differences between the studies could account for this difference. Most notably, the audience in Parrill et al. (2010) was a friend of the speaker who was truly naïve about the events in the cartoon, whereas the audience in Hostetter and Skirving (2011) was the experimenter. There are likely motivational differences between communicating with a naïve friend and an unknown (and assumedly knowledgeable) experimenter. For example, increased motivation to be interesting and helpful to a friend who needs to understand the story could encourage speakers to generate rich images on the basis of the words they hear, thereby reducing the strength of the manipulation of words versus images. Indeed, Parrill et al. (2010) conclude that their results suggest that speakers can simulate images based on a verbal text, and when they do, their gestures look very similar to those that are produced when the events have been seen directly.

Although speculative, there is some empirical support for this explanation that the effect of stimulus presentation modality can depend on aspects of the communicative situation. Masson-Carro, Goudbeek, and Krahmer (2017) found that participants gestured more when describing an object to someone else when the object was indicated with a picture than when it was indicated with a verbal label. The effect disappeared, however, when the task was made more difficult by preventing participants from naming the object, perhaps because participants adopted a very different strategy when not allowed to use the name. Specifically, when participants could not name the object, they produced longer descriptions that included more detail about the appearance and function of the object. We argue that, to create these richer descriptions, people simulated the objects' properties in more detail, and they then expressed these rich simulations as gestures (regardless of whether they had seen a picture of the object or a verbal label)—just as being more motivated to be clear in the Parrill et al. (2010) study may have encouraged people to form more richly detailed images, even in the text condition.

Thus far, the finding that people gesture at a higher rate when they have seen images than when they have read text could also be explained by arguing that people gesture at a higher rate when the speaking task is more difficult. Indeed, there is evidence that speakers gesture more when a task is more cognitively or linguistically complex (e.g., Kita & Davies, 2009) and that gestures can help manage increased cognitive load (e.g., Goldin-Meadow, Nusbaum, Kelly, &

Wagner, 2001). Perhaps it is simply more difficult to describe information that was presented as images than to describe the same information when it was presented as words, and this increased difficulty leads to more gestures. Certainly, prohibiting speakers from naming the object they are describing (as in Masson-Carro et al., 2017) imposes an additional cognitive load that could result in more gestures.

However, converging evidence suggests that difficulty cannot fully account for speakers' high rates of gesture when describing images. Sassenberg and van der Meer (2010) observed the gestures produced by participants as they described a route. They found that steps that were highly activated in speakers' thinking (because they had been described before) were expressed with higher rates of gesture than steps that were newly activated, even though the highly activated steps should have been easier to describe because they had been described previously. Furthermore, Hostetter and Sullivan (2011) found that speakers gestured more on a task that involved manipulating images (i.e., how to assemble shapes to make a picture) than on a task that involved manipulating verbal information (i.e., how to assemble words to make a sentence), even though the verbal task was arguably more difficult for participants and resulted in more errors. Finally, Nicoladis, Marentette, and Navarro (2016) found that the best predictor of children's gesture rates in a narrative task was the overall length of their narratives, and they argued that greater length indicated greater reliance on imagery. When children formed rich images of the story they were telling, they told longer stories and also gestured at higher rates. In contrast, narrative complexity—indexed by the number of discourse connectors (e.g., *so*, *when*, *since*)—did not predict gesture rates. Children who told more difficult and complex narratives did not necessarily gesture more, as would be expected if gestures were produced in order to manage cognitive difficulty. In all these cases, higher gesture rates did not coincide with higher difficulty, but with greater visuospatial imagery.

Individual differences in speakers' likelihood of activating visuospatial imagery may also explain some dimensions of individual differences in gesture production. Speakers whose spatial skills outstrip their verbal skills tend to gesture more than speakers with other cognitive profiles (e.g., Hostetter & Alibali, 2007). One interpretation of this finding is that speakers who are especially likely to form an imagistic simulation of the information they are speaking about are also especially likely to produce gestures.

People also gesture more when visuospatial imagery is more strongly activated. In support of this view, Smithson and Nicoladis (2014) found that speakers gestured more as they retold a cartoon story while also watching an unrelated, complex visuospatial array than while watching a simpler array. The researchers suggested that watching the complex visuospatial array interfered with participants' imagery of the cartoon, and therefore required them to activate that imagery

more strongly. This additional activation resulted in more gesture about that imagery. Note that this explanation hinges on the fact that the cartoon retell task requires participants to activate imagery. In a different sort of task in which participants could shift to a non-imagery-based approach (i.e., a symbolic or propositional approach), participants might shift their approach to the task when confronted with visuospatial interference from the complex visuospatial array, rather than activating the imagery more strongly. In that case, they would gesture *less* with the complex array than with the simple array.

Whereas Smithson and Nicoladis (2014) showed participants motion unrelated to the story they were describing, an alternative approach would be to show participants a visual array that is related to what they are describing. If gestures emerge from activation in the sensorimotor system, then it may be possible to strengthen that activation through a concurrent perceptual task as participants are speaking. Hostetter, Boneff, and Alibali (2018) tested this hypothesis by asking speakers to describe vignettes about vertical motion (e.g., a rocket ship blasting off; an anchor dropping) while simultaneously watching a perceptual display of circles moving vertically across the computer screen. They found that speakers gestured about the vertical motion events at a higher rate when they were watching a perceptual display that moved in a congruent direction (e.g., watching circles moving up while describing a rocket moving up) than when watching a perceptual display that moved in an incongruent direction. Moreover, the difference could not be explained by differences in what speakers talked about, as they did not use more words to describe the vertical events when watching the congruent versus the incongruent display. It appears that the representations that give rise to gestures can also be activated through extraneous activity in the perceptual system.

In sum, the available evidence largely favors the claim that gesture rates increase when speakers describe information that is based on visuospatial images. Although this effect may be sensitive to aspects of the communicative situation, the increase in gesture rate that comes with imagery does not seem to be due to increased cognitive or lexical difficulty.

It is worth noting that, across studies, researchers operationalize gesture *rate* in several different ways (e.g., gestures per word, gestures per second, gestures per semantic attribute). These measures of rate are typically highly correlated, though they can also diverge (e.g., Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts, 2015). If speakers produce more words in response to images than in response to verbal texts (as in Parrill et al., 2010), it may be particularly important for future research on this issue to examine gesture rates per semantic attribute rather than per word. Speakers may gesture more about each attribute they describe if they have encoded that attribute imaginistically than if they have encoded it verbally, regardless of how many words they use to describe that attribute.

This important methodological issue aside, studies have yielded evidence consistent with the claim that gestures emerge with greater frequency when speakers are describing a visuospatial image than when describing a verbal text. We next explore the possibility that gesture rates are affected by the specific type of imagery involved in a simulation.

Prediction 2: Gesture rates vary depending on the type of mental imagery in which the gesturer engages

The GSA framework draws a distinction between visual imagery, which is the re-instantiation of a visual experience, and motor imagery, which is the re-instantiation of an action experience. Visual and motor imagery are conceptualized as end-points of a continuum that varies in the strength of motor system involvement, with spatial imagery, which is thinking about the spatial relationships between points (e.g., orientation, location), at the midpoint on the continuum (see also Cornoldi & Vecchi, 2003). Importantly, all three types of imagery involve simulation of sensorimotor states, and all can activate motor plans and be expressed in gesture. In the case of motor imagery, the link to gesture is straightforward, as the simulation of the action forms the basis of the motor plan for the gesture. When speakers think about a particular action (that they performed, observed, or imagined), they create a neural experience of actually performing that action, which involves activation of the premotor and motor systems.

In the cases of visual and spatial imagery, the activation of a motor plan that can be expressed in gesture warrants more explanation. We contend that spatial imagery has close ties to action because imagining where points are located in space involves simulating how to move from one point to another. In support of this view, the ability to plan and produce movements to various locations is critical for encoding and maintaining information in spatial working memory (e.g., Lawrence, Myerson, Oonk, & Abrams, 2001; Pearson, Ball, & Smith, 2014). When someone thinks specifically about the spatial properties of an image (e.g., its location, its orientation), the motor system may be activated so as to reach and grasp the object. This activation could be manifest in gestures that indicate the general location of an object. Furthermore, when speakers imagine spatial relations between objects in a mental map or between parts of an object in a visual image, they shift their attention between locations (Kosslyn, Thompson, & Ganis, 2006). In our view, such attentional shifts among the components of an image can give rise to gestures that trace the relevant trajectories.

In contrast to spatial images that encode information about the orientation, location, and size of objects, visual images encode information about object identity (Kosslyn et al., 2006). When speakers imagine a particular visual experience, the neural experience is visual, but it can also activate the

motor system because of the close ties between perception and action. There is ample evidence that perception involves automatic activation of corresponding action states—such as how to grasp, hold, or use the object that is perceived (e.g., Bub, Masson, & Cree, 2008; Masson, Bub, & Warren, 2008; Tucker & Ellis, 2004). Our claim is that these same links between perception and action also exist in the realm of mental simulation (see Hostetter & Boncoddio, 2017, for more discussion of this point).

Specifically, visual images can give rise to gesture when either the motor or spatial characteristics of the visual image are salient in the image or important for the task at hand. First, thinking about a visual image can activate *motor* imagery about how to interact with the imagined object or scene. For example, thinking about a hammer may activate a motor simulation of how to use a hammer, which would result in a gesture that demonstrates the pounding action one would perform with a hammer. Second, visual imagery may evoke *spatial* imagery about how the parts of the objects or scene are arranged in space. For example, thinking about the components of a particular hammer could evoke spatial simulations of how the various parts of the hammer are positioned relative to one another. As we described above, such spatial simulations rely on imagined movement between the locations and could result in a gesture that traces the hammer's shape. Thus, visual images can give rise to gestures when the spatial or motor components of the image are salient, because thinking about the spatial or motor components of a visual image relies on simulations of movement. In contrast, thinking about nonspatial visual properties (e.g., color) of an imagined object should lead to very few gestures, because action is not a fundamental part of how we simulate those properties.

Under this view, the likelihood that a description is accompanied by gesture should depend on the extent to which the description evokes thoughts of action. Speakers should gesture more when their imagery is primarily motoric or primarily spatial than when it is primarily visual. Indeed, children gesture more when describing verbs than when describing nouns (Lavelli & Majorano, 2016). Moreover, speakers gesture more when describing manipulable objects (e.g., hammer) than when describing nonmanipulable ones (Hostetter, 2014; Masson-Carro, Goudbeek, & Krahmer, 2016a; Pine, Gurney, & Fletcher, 2010).

Furthermore, there is evidence that speakers' gesture rates depend on their actual or imagined experience with the object they are describing. Hostetter and Alibali (2010) found that speakers gestured more about visual patterns when they had motor experience making those patterns (by arranging wooden disks) before describing them than when they had only viewed the patterns. Similarly, Kamermans et al. (2018) found that participants gestured more about a figure they had learned through haptic exploration than about a pattern they had learned visually. In both cases, participants gestured more

when describing information they had learned through manual experience—that was thus more evocative of motor imagery—than when describing something they had learned through visual experience alone. Along similar lines, Chu and Kita (2016) found that speakers gestured less when they described the orientation of a mug that had spikes along its surface (so as to prevent grasping) than when they described the orientation of a mug without such spikes. Even *imagined* manual action (or the lack thereof) can affect the activation of motor imagery and, as a consequence, gesture.

Thus, it appears that visual images are particularly likely to evoke gestures when the action-relevant or spatial characteristics of the object that is seen or imagined are salient. To date, the studies testing this idea have manipulated the extent to which the motor affordances of a visual image are salient. To our knowledge, no studies have yet manipulated the salience of the spatial characteristics (such as size or location) of a visual image. We predict that gesture rates should increase when the spatial characteristics of a visual image are emphasized, just as gesture rates increase when motor affordances are emphasized. For example, learning a pattern by seeing its component parts appear sequentially might result in more gesture than would learning the pattern by viewing it holistically, because the sequential presentation would highlight the spatial relations among the parts. However, we suspect that the very act of describing the parts of a visual image likely emphasizes the spatial characteristics of the image, making it difficult to truly manipulate how strongly spatial characteristics are activated in a simulation without also changing what participants talk about.

In sum, the GSA framework explains the likelihood of gesture as being partially due to how strongly individuals activate action simulations as they think about images. People may activate action simulations more or less strongly depending on how they encode an image (i.e., as a motor, spatial, or purely visual image) and on how they think about the image when speaking or solving a problem (i.e., whether motor or spatial components are particularly salient).

What about other, nonvisual forms of imagery, such as auditory or tactile imagery? According to the GSA framework, if such imagery involves simulated actions, it should also give rise to gestures. Imagining a rabbit's soft fur might evoke the action of petting the rabbit, and this action simulation could give rise to a gesture. Likewise, imagining the quiet sound that a gerbil makes may evoke a simulation of orienting the head or body so as to better hear the (imagined) sound. This action simulation might also give rise to a gesture. According to the GSA framework, any form of imagery that evokes action simulation is likely to be manifested in gesture.

Even when no direct action simulation is evoked by a particular image, the GSA framework proposes that imagining the image in motion may also give rise to gestures. We turn next to this issue.

Prediction 3: People gesture at higher rates when describing images that they mentally transform or manipulate than when describing static images

People sometimes imagine images in motion, such as when they mentally manipulate or transform an imagined object. This imagined motion may involve simulations of acting on the object or of the object moving, and the GSA framework proposes that this imagined manipulation or action can give rise to gesture.

One task that involves imagined manipulations of objects is mental rotation. Chu and Kita (2008, 2011, 2016) provide in-depth explorations of the gestures that people produce during mental rotation. They have found that people initially produce gestures that closely resemble actual actions on the objects, using grasping hand shapes as if physically grasping and rotating the objects. As participants gain experience with the task, their gestures shift to representing the objects themselves, with the hand acting as the object that is being rotated (Chu & Kita, 2008; see also Schwartz & Black, 1996). Furthermore, speakers who are better at the task of mental rotation rely on gestures less than speakers who are worse at it, and preventing gesture during mental rotation significantly increases reaction time (Chu & Kita, 2011). Thus, it appears that gestures play a functional role in mental rotation; enacting the mental simulation of rotation in gesture seems to improve people's ability to perform the task.

According to the GSA framework, gestures occur with mental rotation because speakers mentally simulate the actions needed to rotate the object. As they gain experience with the task, the nature of their simulation changes, such that they no longer need to simulate the action of manipulating, and instead simulate the object as it moves. As a result, the form of the gesture changes from one that depicts acting on the object to one that depicts the object's trajectory as it moves. In addition, as the internal computation of mental rotation becomes easier (e.g., with experience on the task), people activate the exact motion of the object and the mechanics of how to move it less strongly in their mental simulations, and they produce fewer gestures. Speakers who have extensive experience with the task may be able to imagine the end state of a given rotation, without imagining the motion or action involved in getting the object to that position. Indeed, the correlation between amount of rotation and reaction time disappears with repeated practice (Provost, Johnson, Karayanidis, Brown, & Heathcote, 2013), suggesting that mental simulation of the rotation is no longer involved.

The hypothesis that speakers gesture more when imagining how something moves than when imagining it in a static state has received some support. Hostetter, Alibali, and Bartholomew (2011) asked speakers to describe arrows either in their presented orientation or as they would look rotated a specified amount (e.g., 90 deg). Importantly, in the no-rotation

condition, the arrays were presented in the orientation they would be at the end of the rotation in the rotation condition. Thus, speakers described the same information in both conditions, but in the rotation condition, they had to imagine mentally rotating the arrows in order to arrive at the correct orientation. Gesture rates were higher when describing the mentally rotated images than when describing the same images presented in final orientation, supporting the claim that gesture rates are higher when images are simulated in motion than when they are not.

Speakers sometimes need to mentally rotate their mental maps when they give directions. Sassenberg and van der Meer (2010) showed participants a map depicting several possible routes and then examined the gestures produced by speakers as they described the routes. Speakers were told to describe the routes from a "route perspective," that is, from the point of view of a person walking through the route (e.g., using words like "left" and "right" rather than "up" or "down" to describe a turn that takes one vertically up or down the page). Sassenberg and van der Meer compared the frequency of gestures accompanying the descriptions of left and right turns in the routes. They found that gestures were more prevalent when the step had been shown on the map as a vertical move (e.g., up vs. down)—requiring mental rotation to fit the first-person perspective—than when it had been shown as a horizontal turn. This finding aligns with the prediction that gesture rates should be higher when images are mentally rotated than when they are not.

Thus far, we have considered predictions made by the GSA framework that apply exclusively to gesture *rates*, or the likelihood and prevalence of gestures in various situations. However, the framework also yields predictions about the form particular gestures should take. Specifically, the framework predicts that the form of a gesture should reflect particular aspects of the mental simulation that is being described.

Prediction 4: The form of a gesture should mirror the form of the underlying mental simulation

One hallmark of iconic co-speech gestures is that there is not a one-to-one correspondence between a particular representation and a particular gesture (McNeill, 1992). Gestures for the same idea can take many different forms: For example, gestures can be produced with one hand or two, they can take the perspective of a character or of an outside observer, and they can depict the shape of an object or how to use it. According to the GSA framework, the particular form a gesture takes should not be random, but instead should be a consequence of what is salient in the speaker's simulation at the moment of speaking. As a result, there should be observable correspondences between the forms of speakers' gestures and the specific experiences they describe or think about.

There is much evidence that the forms of speakers' gestures reflect their experiences, and thus presumably the specifics of the simulations they have in mind as they are speaking. For example, the shape of the arc trajectory in a particular lifting gesture is related to whether the lifting action being described is a physical lifting action or a mouse movement; speakers produce more pronounced arcs in their gesture when they have had experience physically lifting objects rather than sliding them with a mouse (Cook & Tanenhaus, 2009). Furthermore, speakers often gesture about lifting small objects that are light in weight with one hand rather than two (Beilock & Goldin-Meadow, 2010), and speakers who estimate the weights of objects as heavier gesture about lifting them with a slower velocity, as though the objects are more difficult to lift (Pouw, Wassenburg, Hostetter, de Koning, & Paas, 2018). When speakers describe the motion of objects, the speed of their hands reflects the speed with which they saw the objects move (Hilliard & Cook, 2017). Thus, the specific experiences that speakers have with objects are often apparent in the forms of their gestures about those objects.

The form of gesture presumably depends on the nature of the underlying imagery. For example, when speakers form a spatial image of an object that emphasizes how the parts of the object are arranged in space, they should be particularly likely to produce gestures that move between the various parts of the shape. Indeed, Masson-Carro et al. (2017) found that such tracing or molding gestures were particularly likely to occur when speakers were describing nonmanipulable objects (e.g., a traffic light), particularly when they had seen a picture (rather than read the name) of the object. It seems likely the spatial characteristics of the objects were particularly salient after seeing the objects depicted. In contrast, Masson-Carro et al. (2017) found that participants were more likely to produce gestures that mimicked holding or interacting with objects when they described objects that are readily manipulable (e.g., a tennis racket), and they produced such gestures most often when they had read the name of the object rather than seen it depicted. It seems likely that manipulable objects activate motor imagery more strongly than spatial imagery, particularly when the objects' spatial characteristics have not just been viewed.

The way events are imagined or simulated can also be affected by the way those events are verbally described. Reading about events in a particular way can encourage speakers to simulate those events accordingly, and thus affect the way they gesture about those events. Parrill, Bergen, and Lichtenstein (2013) presented participants with written stories that described events written either in the past perfect tense, which implied that the actions were complete (e.g., the squirrel *had scampered* across the path), or in the past progressive tense, which implied that the actions were ongoing (e.g., the squirrel *was scampering* across the path). When events were presented in the past progressive tense,

speakers were more likely to use this tense in their speech than when the events were presented in past perfect tense. More to the current point, when speakers reproduced the past progressive tense in speech, they produced gestures that were longer in duration and more complex (i.e., contained repeated movements) than when they used the past perfect tense in speech. The authors argued that the past progressive tense signals a mental representation of the event that is focused on the details and internal structure of the event. This focus on the internal details leads to more detailed (i.e., complex and longer lasting) gestures. Thus, the specifics of how an individual simulates an event are reflected in the form of the corresponding gesture.

Experiences in the world often support multiple possible simulations during speaking. For example, when describing movements, speakers often describe both the path taken and the manner of motion along that trajectory. Languages differ in how they package this information syntactically (see Talmy, 1983). Satellite-framed languages, such as English, tend to express manner in the main verb and path in a satellite (e.g., *he rolled down*), whereas verb-framed languages, such as Turkish, tend to express path in the main verb and manner in the satellite (e.g., *he descended by rolling*). Investigations of how speakers of different languages gesture about path and manner suggest that adult speakers typically gesture in ways that mirror their speech (e.g., Özyürek, Kita, Allen, Furman, & Brown, 2005). If their language tends to use separate verbs to express manner and path (as in verb-framed languages), speakers typically segment path and manner into two separate gestures or leave manner information out of their gestures altogether. In contrast, if their language tends to express path and manner in a single clause with only one verb (as in satellite-framed languages), speakers typically produce gestures that express both path and manner simultaneously. This tendency to segment or conflate gestures for path and manner in the same way they are packaged in speech is present even among speakers who have been blind from birth, suggesting that experience seeing speakers gesture in a particular way is not necessary for the pattern to emerge (Özçalışkan, Lucero, & Goldin-Meadow, 2016b).

We suggest that the language spoken can bias speakers to simulate path and manner either separately or together as they are speaking. According to this account, speakers of satellite-framed languages tend to form a single simulation that involves both path and manner, and their gestures reflect this. In contrast, speakers of verb-framed languages tend to form a simulation of path followed by a simulation of manner, and they produce two separate gestures as a result. However, these biases also depend on how closely path and manner are causally related in the event itself. Kita et al. (2007) distinguished between events in which the manner of motion was inherent to the path (e.g., jumping caused the character to ascend) and events in which the manner of motion was incidental to the path (e.g., the character fell and just happened to be spinning as he went). They

found that English speakers produced single-clause utterances most often for events in which manner was inherent to the path; when manner was incidental, English speakers often expressed manner and path in separate clauses. Moreover, speakers' gestures coincided with their syntactic framing. People produced separate gestures for manner and path most often when they produced separate clauses, whereas they produced conflated gestures that expressed both path and manner most often when they produced single clauses that expressed path and manner together. Thus, whether path and manner are simulated together or separately—and thus whether they are expressed in speech and gesture together or separately—can depend on the nature of the event. Both features of the event itself and language-specific syntactic patterns can affect the ways that speakers simulate path and manner, and consequently, the gestures they produce.

It is also possible for a particular simulation to differentially emphasize path or manner, in which case gesture should reflect the dimension that is most salient. Akhavan, Nozari, and Göksun (2017) found that speakers of Farsi often gestured about path alone, regardless of whether or how they mentioned manner in speech. They argued that this is because path is a particularly salient aspect of the representation of motion events for Farsi speakers. Yeo and Alibali (2018) directly manipulated the salience of manner in motion events, by making a spider move in either a large, fast zigzag motion or a smaller, slower zigzag motion. English speakers produced more gestures that expressed the manner of the spider's motion when that motion was more salient (i.e., when the zigzags were large and fast), and this difference was not attributable to differences in their mentioning manner in speech. Thus, how path and manner are expressed in gesture depends on what is most salient in the speaker's simulation at the moment of speaking, which is shaped both by properties of the observed motion event and by syntactic patterns typical in the speaker's language.

The relation between the form of a gesture and the form of the underlying simulation is also apparent in the *perspective* that speakers take when expressing manner information in gesture. Specifically, character viewpoint gestures are those in which the speaker's hands and/or body represent the hands and/or body of the character they are describing. For example, a speaker who describes a character throwing a ball could produce a character viewpoint gesture by moving his hand as though throwing the ball. The GSA framework contends that such gestures arise from motor imagery, in which speakers imagine themselves performing the actions they describe. In contrast, observer viewpoint gestures are those that depict an action or object from the perspective of an outside observer. For example, a speaker who describes a character throwing a ball could produce an observer viewpoint gesture by moving her hand across her body to show the path of the ball as it moved. The GSA framework proposes that such gestures arise from action simulations activated during spatial imagery—for example, simulations that trace the trajectory of a motion or the outline of an object.

Whether a gesture is produced with a character or observer viewpoint thus depends on how the event is simulated, which may in turn depend on how the event was experienced and on what information about the event is particularly salient. For example, Parrill and Stec (2018) found that participants who saw a picture of an event from a first-person perspective were more likely to produce character viewpoint gestures about the event than participants who saw a picture of the same event from a third-person perspective. Moreover, certain kinds of events are particularly likely to be simulated from the point of view of a character. When a character performs an action with his hands, speakers very frequently gesture about such events from a character viewpoint (Parrill, 2010). Likewise, speakers tend to gesture about objects that are used with the hands by depicting how to handle the objects (character viewpoint), rather than by tracing their outlines (observer viewpoint; Masson-Carro et al., 2016a, 2017).

Of course, the relation between how an event is experienced and how that event is simulated as it is described is not absolute. Speakers can adjust their mental images of events, leading to differences in how frequently they adopt a character viewpoint in gesture. For example, speakers who have Parkinson's disease have motor impairments that may prevent them from forming motor images (Helmich, de Lange, Bloem, & Toni, 2007). On this basis, one might expect that patients with Parkinson's disease would be unlikely to produce character viewpoint gestures, and indeed, this is the case. One study reported that patients with Parkinson's disease produced primarily observer viewpoint gestures when they described common actions, whereas age-matched controls produced primarily character viewpoint gestures (Humphries, Holler, Crawford, Herrera, & Poliakoff, 2016). The speakers with Parkinson's disease also demonstrated significantly worse performance on a separate motor imagery task than the age-matched controls. The researchers concluded that, because patients with Parkinson's disease are unable to form motor images that take a first-person perspective, they rely instead on visual images in third-person perspective, and these visual images result in gestures produced primarily from an observer viewpoint.

There is evidence of a similar relation between character viewpoint gesture and the ability to take a first-person perspective among children. Demir, Levine, and Goldin-Meadow (2015) observed variations in whether 5-year-old children adopted character viewpoint in their gestures as they narrated a cartoon. Interestingly, children who used character viewpoint at age 5 produced narratives that were more well-structured around the protagonist's goals at ages 6, 7, and 8 than children who did not use character viewpoint gestures at age 5. The authors argued that the use of character viewpoint at age 5 indicated that the children took the character's perspective in their mental imagery, and this ability to take another's perspective is important for producing well-structured narratives.

In sum, much evidence has suggested that the form of people's gestures about an event or object is related to how they simulate the event or object in question. Differences in gesture form, gesture content, gesture segmentation, and gesture viewpoint correspond (at least in part) to differences in how people have experienced what they are speaking or thinking about—a point we return to below.

One noteworthy complication in studies that examine gesture form is that it is not uncommon to find participants who do not gesture at all on a given task, and such participants cannot be included in studies examining factors that influence gesture form. However, some recent work has suggested that asking participants to gesture does not significantly alter the type or form of the gestures that they produce (Parrill, Cabot, Kent, Chen, & Payneau, 2016). Thus, we suggest that researchers who are primarily interested in questions about gesture form take steps to increase the probability that each participant produces at least some gestures, whether by giving them highly imagistic or motoric tasks to complete, instructing them to gesture, or creating a communicative context in which gestures are invaluable.

The idea that speakers adjust how much they gesture depending on the communicative context is realized in the GSA framework in terms of adjustments in the height of the gesture threshold. Next, we consider evidence for the claim that speakers can adjust their use of gesture depending on the cognitive and communicative situation.

Prediction 5: Gesture production varies dynamically as a function of both stable individual differences and more temporary aspects of the communicative and cognitive situation

To account for these variations, the framework incorporates a *gesture threshold*, which is the minimum level of activation that an action simulation must have in order to give rise to a gesture. If the threshold is high, even very strongly activated action simulations may not be expressed as gestures; if the threshold is low, even simulations that are weakly activated may be expressed as gestures. Some people may have a set point for their gesture threshold that is higher than other people, due to individual differences in stable characteristics, such as cognitive skills or personality, differences in cortical connectivity from premotor to motor areas, or differences in cultural beliefs about the appropriateness of gesture. In addition to these individual differences, the gesture threshold can also fluctuate around its set point in response to more temporary, situational factors, including the importance of the message, the potential benefit of the gesture for the listeners' understanding, and aspects of the communicative situation or the cognitive task.

There are persistent individual differences in how much people gesture. According to the GSA framework, these

individual differences are due in part to differences in the set points of people's gesture thresholds. People's set points for their gesture thresholds depend on a range of factors, including cognitive skills and personality characteristics. For example, people who tend to have difficulty maintaining visual images (e.g., Chu, Meyer, Foulkes, & Kita, 2014) or verbal information (e.g., Gillespie, James, Federmeier, & Watson, 2014) may maintain low set points for their gesture thresholds, so that they can regularly experience the cognitive benefits of gestures. Likewise, more extroverted people may tend to maintain lower set points, and consequently have higher gesture rates than less extroverted people (Hostetter & Potthoff, 2012; O'Carroll, Nicoladis, & Smithson, 2015), because extroverted people may believe (or have experienced) that gestures are generally engaging for their listeners.

Individual differences in gesture thresholds may also be evident in patterns of neural activation. Speakers who have low gesture thresholds may have more difficulty inhibiting the premotor plans involved in simulation from spreading to motor cortex, where they become expressed as gestures, and this may be manifested in neurological differences. In the only study to date to examine this issue, Wartenburger et al. (2010) found that cortical thickness in the left hemisphere of the temporal cortex and in Broca's area was positively correlated with gesture production on a reasoning task that required imagining how geometric shapes would be repositioned. They argued that to succeed on the geometric task, speakers imagined moving the pieces, and therefore activated motor simulations. These motor simulations were particularly likely to be expressed as gesture for speakers who had high cortical thickness, because for these speakers, the motor activation involved in simulating the movement of the pieces spread more readily to motor areas. However, it is impossible to know from this study alone whether cortical thickness led to increased gesture, or whether a lifetime of increased gesturing led to increased cortical thickness. Nonetheless, the idea that neurological differences may relate to individual differences in gesture production is compatible with the predictions of the GSA framework.

There are also cultural differences in gesture rates (Kita, 2009), suggesting that speakers may have different set points for their gesture thresholds depending on norms regarding gesture in their culture. For example, monolingual speakers of Chinese gesture less than monolingual speakers of American English (So, 2010). This may be because Chinese speakers have different beliefs about the appropriateness of gesture, leading to relatively high set points for their gesture thresholds in comparison to American English speakers. Interestingly, however, learning a second language with a different cultural norm regarding gesture may lead to shifts in speakers' thresholds; speakers who are bilingual in Chinese and American English gesture more when speaking Chinese than do monolingual Chinese speakers (So, 2010), suggesting

that their experiences speaking English have led to lower thresholds across languages. Moreover, experience learning American Sign Language (ASL) may also lead to generally lower gesture thresholds; people who have a year of experience with ASL produce more speech-accompanying gestures when speaking English than do people who have a year of experience learning a second spoken language (Casey, Emmorey, & Larrabee, 2012).

Thus, there may be stable differences in the set points of people's gesture thresholds, based on factors such as cognitive skills, personality, culture, and experience learning other languages. In addition to these more stable differences, people's gesture thresholds also vary depending on temporary aspects of the cognitive or communicative situation. For example, bilingual French–English speakers gesture more when speaking in French, a high-gesture language, than when speaking in English (Laurent & Nicoladis, 2015), suggesting that speakers' thresholds may shift depending on the communicative context.

Some aspects of the communicative context besides the language spoken also matter. For example, when speakers perceive the information they are communicating about to be particularly important, they gesture more than when the information is less important. This can occur across an entire conversation, with speakers gesturing more when they think the information will be highly relevant and useful to their audience than when the information has no obvious utility (Kelly, Byrne, & Holler, 2011). It can also occur within a conversation, with speakers being particularly likely to gesture about aspects of their message that are particularly important. For example, mothers gesture more when speaking to their children about safety information that is relevant to situations that the mothers perceive to be particularly unsafe (Hilliard, O'Neal, Plumert, & Cook, 2015). Likewise, teachers gesture more when describing information that is new for their students than when describing information that is being reviewed (Alibali & Nathan, 2007; Alibali, Nathan, et al., 2014). Furthermore, although speakers sometimes gesture less when they repeat old information to the same listener (e.g., Galati & Brennan, 2013; Jacobs & Garnham, 2007), speakers' use of gesture increases with repeated information if they believe that their listeners did not understand the message the first time (e.g., Hoetjes, Krahmer, & Swerts, 2015). All of these phenomena can be conceptualized in terms of a temporary shift in speakers' gesture thresholds.

Of course, for speakers' gestures to be beneficial to listeners, the listeners must be able to see the gestures. Many studies have examined the effects of audience visibility on speakers' gestures (see Bavelas & Healing, 2013, for a review), and several of these have reported that speakers gesture at higher rates when speaking to listeners who can see the gestures than when speaking to listeners who cannot see—and thus cannot benefit from—the gestures (e.g., Alibali, Heath, &

Myers, 2001; Bavelas, Gerwing, Sutton, & Prevost, 2008; Hoetjes, Koolen, et al., 2015). From the perspective of the GSA framework, this effect occurs because speakers' gesture thresholds are lower when their listeners are visible, thereby allowing more simulations to exceed threshold and be expressed as gestures. When listeners are not visible, speakers' thresholds remain high, so that they produce gestures only for simulations that are very strongly activated. Indeed, the effect of visibility on speakers' gestures seems to depend on speaking topic, with gestures persisting during descriptions of manipulable objects (which presumably include very strong simulations of action), even when listeners are not visible (Hostetter, 2014; Pine et al., 2010).

Although listener visibility effects are clearly predicted by the GSA framework, some recent studies have found no differences in gesture as a function of listener visibility (e.g., de Ruiter, Bangerter, & Dings, 2012; Hoetjes, Krahmer, & Swerts, 2015) or have found that speakers gesture more when their listeners cannot see them than when they can (e.g., O'Carroll et al., 2015). As Bavelas and Healing (2013) pointed out in their review, it is possible that these differences across studies can be understood by considering the tasks in more detail.

In experiments with confederates or experimenters as the audience, speakers generally receive no verbal feedback about the effectiveness of their communication. Without any feedback, speakers likely assume that their communication is effective. However, when speakers can see the audience, they may receive nonverbal feedback about the effectiveness of their communication, and this nonverbal feedback may prompt speakers to adjust their speech and gestures. As a result, studies that use confederates or experimenters as listeners are particularly likely to find audience visibility effects on gesture rate (Bavelas & Healing, 2013).

In contrast, when the audience in an experiment is a naïve participant who is allowed to engage in dialogue with the speaker, speakers receive feedback about listeners' understanding in both the visible and the nonvisible conditions. If speakers perceive that listeners are not understanding, they may lower their thresholds—not for the listeners, who cannot see the gestures anyway—but for themselves, in order to reduce their own cognitive load as they attempt to give better, clearer verbal descriptions. The result is that gesture rates may remain high in such situations, even when speakers cannot see their listeners. Indeed, O'Carroll et al. (2015) found that speakers actually gestured *more* when their listeners were not visible, and they suggest that speakers increased their gesture rates in the nonvisible listener condition in order to help themselves describe the highly spatial task more effectively in speech.

The GSA framework proposes that another determinant of the height of the gesture threshold is the speaker's own cognitive load. It takes effort to maintain a high threshold (as we

discuss in detail in the following section), and actually producing gestures appears to have a number of cognitive benefits (see Kita, Alibali, & Chu, 2017). Thus, the gesture threshold may be lowered in situations in which externalizing a simulation as gesture can facilitate the speaker's own thinking and speaking. In support of this view, speakers gesture more in description tasks that are more cognitively demanding (e.g., Hostetter, Alibali, & Kita, 2007), and they gesture more when they are under extraneous cognitive load (e.g., Hoetjes & Masson-Carro, 2017). Moreover, there is evidence that speakers gesture more when ideas are difficult to describe, either because the ideas are not easily lexicalized in their language (e.g., Morsella & Krauss, 2004; but see de Ruiter et al., 2012), because the speakers are bilingual (e.g., Nicoladis, Pika, & Marentette, 2009), or because they have a brain injury (e.g., Göksun, Lehet, Malykhina, & Chatterjee, 2015; Kim, Stierwalt, LaPointe, & Bourgeois, 2015). Although task difficulty does not guarantee an increase in gesture rates—as there must first be an underlying imagistic simulation that is being described—the GSA framework contends that increased cognitive load can result in lower thresholds and higher gesture rates.

This discussion raises the question of whether adjusting the gesture threshold is a conscious mechanism (one that speakers are aware of) or an unconscious one. The GSA framework is neutral with regard to this point. There is evidence that speakers *can* consciously adjust their gesture thresholds; for example, teachers can gesture more when instructed to do so (Alibali, Young, et al., 2013). Speakers also sometimes behave as though they intend for their gestures to communicate—for example, by omitting necessary information from their speech (e.g., Melinger & Levelt, 2004) or by gazing at their own gestures (e.g., Gullberg & Holmqvist, 2006; Gullberg & Kita, 2009). At the same time, speakers are often unaware of their gestures, and we propose that the gesture threshold often operates outside of conscious awareness. According to the GSA framework, gestures emerge automatically on the basis of the strength of action simulation and on the current level of the gesture threshold at the moment of speaking. Conscious intention to gesture can bring speakers' attention to this process, but the process occurs even without conscious awareness. We generally agree with the position, articulated by Campisi and Mazzone (2016) and by Cooperrider (2018), that gestures can be produced either intentionally or unintentionally, depending on the situation. The unique contribution of the GSA framework to this discussion is that it provides a possible explanation for how general cognitive processes (action simulation during speaking) can give rise to gestures, regardless of intentionality.

In sum, the GSA framework posits that gestures are determined, in part, by the height of a dynamic gesture threshold that fluctuates around a particular set point. This gesture threshold helps to explain both stable individual differences

in gesture rates across speakers and situational differences in gesture rates that depend on the communicative situation or on cognitive demands. Regardless of the set point, a speaker *can* maintain a high threshold and inhibit simulations from being expressed as gestures. For this reason, in studies testing how different factors affect the gesture threshold, nongesturers should be treated as having a gesture rate of zero, rather than being excluded from analysis. This is because the complete absence of any gesture in a highly imagistic task is meaningful, as it suggests that the gesture threshold of that individual was particularly high. However, under our view, maintaining a high threshold even while describing a strongly activated action simulation should be relatively uncommon, because it requires cognitive effort, a claim we turn to next.

Prediction 6: Gesture inhibition is cognitively costly

According to the GSA framework, action simulations are an automatic byproduct of thinking about visual, spatial, and motor images during speaking. These simulations involve activation of the premotor and motor systems, and this activation is expressed as gesture. Thus, gestures are the natural consequences of a mind that is actively engaged with modal, imagistic representations, and it takes effort to inhibit gestures from being expressed. People can maintain a high gesture threshold, thereby effectively inhibiting simulations from being expressed as gestures, but the GSA framework predicts that there should be an associated cognitive cost.

The cognitive effort involved in gesture production has been investigated using the dual-task paradigm (e.g., Cook, Yip, & Goldin-Meadow, 2012a; Goldin-Meadow et al., 2001; Marstaller & Burianova, 2013; Ping & Goldin-Meadow, 2010; Wagner, Nusbaum, & Goldin-Meadow, 2004). In this paradigm, speakers are presented with extraneous information to remember (e.g., digits, letters, or positions in a grid). They are then asked to speak about something, either with gesture or without. Finally, they are asked to recall as much of the extraneous information as they can. The general finding is that speakers have more cognitive resources available to devote to the memory task when they gesture during the speaking task than when they do not gesture during the speaking task. The interpretation is that producing gestures during speaking “frees up” cognitive resources that can then be devoted to remembering the information. Gestures reduce cognitive load only when they are meaningful; producing circular gestures that are temporally but not semantically coordinated with speech does not have the same effect (Cook et al., 2012a).

It is difficult to discern whether the results of these studies reflect a cognitive benefit of producing gestures, a cognitive cost to *not* producing gestures, or a combination of both effects. The argument in favor of facilitation has come from studies including a condition in which speakers were given no instructions about gesture, but they nonetheless

spontaneously chose not to gesture on some trials (e.g., Goldin-Meadow et al., 2001). The cognitive load observed in these spontaneous nongesture trials is similar to that in instructed nongesture trials, casting doubt on the possibility that the increased cognitive load imposed by gesture inhibition is the result of consciously remembering not to gesture. According to the GSA framework, the cognitive cost associated with gesture inhibition should be present whenever a highly activated simulation underlies language or thinking but gesture is inhibited—regardless of whether gesture is inhibited consciously or not. Nonetheless, the dominant interpretation in the published literature using the dual-task paradigm has been that gesture production reduces cognitive load, rather than that gesture inhibition increases cognitive load (e.g., Goldin-Meadow et al., 2001; Ping & Goldin-Meadow, 2010).

One study has specifically favored the interpretation that gesture inhibition results in a cognitive cost, rather than that gesture production results in a cognitive benefit. Marstaller and Burianova (2013) found that the difference in cognitive load associated with gesture production versus inhibition was greater for participants with low working memory capacity than for those with high working memory capacity. They argued that individuals with low working memory capacity are less able to inhibit prepotent motor responses, such as gesture, and they are thus strongly affected by instructions to inhibit gesture. In contrast, individuals with high working memory capacity are better able to inhibit responses in general, and they are therefore able to inhibit gesture without negative effects on performance.

Although this is the interpretation offered by Marstaller and Burianova (2013), it is also possible to explain their results as stemming from a reduction of cognitive load when gestures are produced. Under this account, individuals with high working memory capacity experienced a manageable load on the speaking task, and they were therefore unable to be further helped by gesture; for this reason, only the individuals with low working memory capacity displayed benefits from gesture (see also Eielts et al., 2018). Indeed, Goldin-Meadow et al. (2001) found that adults who were under a very low working memory load (remembering only two letters) while speaking were not affected by whether or not they gestured, lending some support to the possibility that when cognitive load is very low, there is no “room” for gesture to further reduce load in a way that affects performance. However, it should be noted that the participants in the Marstaller and Burianova study were not at ceiling on the memory task; even the individuals with high working memory capacity recalled fewer than half of the letters presented. Thus, it is difficult to make the case that individuals with high working memory capacity could not have been further helped by gesture, as they did have room to improve. Regardless, with only a gesture and a no gesture condition included in the design, it is

impossible to know whether any difference between conditions resulted from decreased cognitive load when gestures were produced, increased load when gestures were inhibited, or a combination of both effects. To disentangle these effects in the dual-task paradigm, studies are needed that include baseline measures of performance on the memory task.

If a cognitive cost is associated with inhibiting a simulation from being expressed as gesture, as we claim, then speakers who are asked to refrain from gesturing may adopt a cognitive strategy on the task, if possible, that does not involve active simulation. Alibali, Spencer, Knox, and Kita (2011) presented participants with gear movement prediction problems, in which they were asked to view a configuration of gears and to predict how the final gear would move, if the initial gear were turned in a particular direction. Participants who were prohibited from gesturing during the task were more likely to generate an abstract rule (i.e., the parity rule, which holds that all odd-numbered gears turn in the same direction as the initial gear) than participants who were allowed to gesture. Inhibiting gesture seemed to discourage participants from using the simulation-based strategy of thinking about the direction of movement of each gear in order, a strategy that was favored by participants who could gesture. One explanation for this finding is that simulating the direction of each gear in order without being able to express that simulation as gesture was too effortful, so participants adopted an easier, non-simulation-based strategy when they could not gesture.

In many cases, another strategy is not readily available for use on a particular task. In such situations, when gestures are prohibited, speakers may struggle to completely inhibit their simulations from being expressed as gestures, and they may overtly express them in other bodily movements. Indeed, when people are prevented from gesturing with their hands in a cartoon-retell task, they are more likely to produce gestures with other body parts, including their heads, trunks and feet, than when they are free to gesture (Alibali & Hostetter, 2018). Along similar lines, Pouw, Mavilidi, van Gog, and Paas (2016) found that participants who were not allowed to gesture as they explained the Tower of Hanoi task made more eye movements to the positions where the pieces should be placed than participants who were allowed to gesture. The effect was particularly pronounced for participants who had low visual working memory capacity, suggesting that individuals who experience high cognitive load on a task have difficulty inhibiting their simulations from being expressed overtly. Pouw et al. argue that this is because participants with low visual working memory capacity need the extra scaffolding that gestures or eye movements can provide, but an equally plausible interpretation is that they do not have the cognitive resources needed to inhibit their simulations from being expressed as motor plans. Because they have been told not to move their hands, they express their simulations in their eye movements instead.

It should be clear, then, that there is some ambiguity in the interpretation of results that compare the cognitive effects of producing versus inhibiting gestures. The interpretation usually proffered is that there is a cognitive benefit to gesture production (e.g., Goldin-Meadow et al., 2001), but the results are equally compatible with the explanation that a cost is associated with gesture inhibition. Importantly, however, the GSA framework does not discount that there might be both cognitive and communicative benefits to producing gestures (a point to which we return below); indeed, anticipating that gestures might make one's speaking task easier is one factor that could affect a speaker's gesture threshold—and rate of gesturing—in a particular situation. At the same time, the GSA framework predicts that inhibiting a simulation from being expressed as gesture should have a discernible cognitive cost (in addition to any benefit that may come from producing the gesture). More studies are needed to dissociate the benefits of gesture production from the costs of gesture inhibition, for example by including control conditions that can provide baseline estimates of cognitive effort that are independent of gesture production. Studies are also needed that compare how much speakers gesture on tasks as a function of whether they are under an extraneous cognitive load or not; if speakers are under extraneous load, they may not have cognitive resources available to devote to inhibiting gesture (see, e.g., Masson-Carro, Goudbeek, & Krahmer, 2016b).

Finally, it bears mentioning that our claim regarding the cognitive costs of inhibiting gesture is specifically about gestures that emerge automatically from simulations of actions and perceptual states. If a speaker is producing speech based on a verbal or propositional representation (e.g., reciting a memorized speech) and wishes to produce a gesture along with the speech (perhaps for communicative purposes), cognitive effort might be needed to generate that gesture (rather than to inhibit it). In this case, according to our view, the speaker must form an imagistic simulation of the event in order to produce a gesture, and generating this simulation may be effortful. Similarly, there may be situations in which speakers deliberately exaggerate or adjust the form or size of their gestures for communicative purposes, and the GSA framework is neutral about whether such additional crafting of gesture might require more cognitive resources than not gesturing at all (see Mol, Krahmer, Maes, & Swerts, 2009).

Conclusion

According to the GSA framework, the likelihood that a person will gesture on a given task is determined by the dynamic relation between the level of activation of the relevant action simulations and the current height of the gesture threshold, which depends on stable individual differences, as well as the communicative situation and the cognitive demands of the speaking task. Furthermore, the form of the gestures that

are produced is dependent on the content of the underlying simulation. As such, many dynamic factors are relevant to understanding gesture, and each of these factors is difficult to measure with high certainty at any particular moment. Nonetheless, the framework does yield concrete predictions about when gestures should be most likely to occur and about the forms gestures should take.

In the decade since the GSA framework was formulated, each of its six predictions has been tested, either directly or indirectly. Although it is necessary to consider the entire cognitive and communicative situation that participants experienced when interpreting the results of any individual study, taken together, the evidence is largely in favor of the predictions. Thus, we believe the framework has largely stood up to empirical test. At the same time, we appreciate the perspectives of the framework's critics (e.g., Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014) and of those who have offered alternative conceptualizations of gestures (e.g., Novack & Goldin-Meadow, 2017). In the next section, we address several specific challenges to the framework that have been raised.

Challenges to the GSA framework

Co-thought gestures also occur

The GSA framework claims that a simulation is more likely to be expressed as gesture when the motor system is simultaneously engaged in the process of speech production than when it is not. However, this claim does not entail that the engagement of the motor system in speech production is *required* in order for gesture to be produced. On the contrary, the GSA framework holds that whenever people simulate actions or perceptual states at a level that exceeds their current gesture threshold, they should produce gestures. Thus, simulations that are very highly activated may be expressed as gestures even in the absence of speech.

From the perspective of the GSA framework, co-thought gestures—that is, gestures that occur when thinking but not speaking—arise from the same processes that give rise to co-speech gestures. As one might expect, then, speakers' rates of co-speech gestures are correlated with their rates of co-thought gestures, and rates of both types of gestures increase when imagined interaction with an object is more likely (Chu & Kita, 2016).

One prediction from the GSA framework is that, other things being equal, people should be more likely to gesture when they are producing speech than when they are not; this is because concurrent speech makes it more difficult to inhibit the motor activation involved in simulation from being expressed as gesture. One study in the literature has provided evidence for this prediction, even though this issue was not the

focus of the analysis. In the study of gear movement prediction problems described above (Alibali et al., 2011), participants in the first experiment talked aloud as they solved the problems, and participants in the second experiment solved the problems silently. In all, 90% (38 of 42) of the participants who talked aloud while solving produced gestures, whereas only 54% (28 of 52) of those who did not talk loud while solving did so, just as the GSA framework would predict.

In brief, within the GSA framework, co-thought and co-speech gestures can be explained via the same basic mechanisms. However, further research is needed that directly tests how co-thought gestures and co-speech gestures differ in terms of their likelihood in various kinds of tasks, their modulation according to the gesture threshold, and the producers' awareness of them.

Gestures are not exactly like actions

Gestures differ from actions in many ways. For example, gestures are more closely synchronized with speech than actions are (Church, Kelly, & Holcomb, 2014), and listeners have a more difficult time ignoring speech-accompanying gestures than speech-accompanying actions (Kelly, Healey, Özyürek, & Holler, 2015).

Furthermore, the form of a gesture is not usually identical to the form of the corresponding action. Gestures often fail to replicate some of the specific details of the actions they depict (e.g., Annett, 1990; Kita et al., 2017). At the same time, gestures can also elaborate on actions, sometimes adding information that helps speakers transform their thinking (Nemirovsky, Kelton, & Rhodhamel, 2012). This discrepancy between gesture and action has led to the criticism that the GSA framework puts too much emphasis on actions. Some researchers contend that the emphasis should instead be on gestures' status as *representations* of action (Novack & Goldin-Meadow, 2017). Gestures seem to strengthen speakers' mental representations (Goldin-Meadow & Beilock, 2010; Trofatter, Kontra, Beilock, & Goldin-Meadow, 2015), and they can encourage generalization of information to new contexts (Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014)—qualities that may not be shared by actions.

We agree that gestures are a special form of action (Alibali, Boncoddio & Hostetter, 2014). Indeed, by emphasizing their status as reflecting *simulated* actions, we contend that gestures are a natural outgrowth of the motor activation and planning processes that are central to cognition. Importantly, however, gestures are not subject to the same constraints as actions, so the form of a gesture need not directly reproduce or mimic the form of the corresponding action. On the contrary, speakers may simulate specific aspects of an action as they are speaking, and only the aspects that are highly activated within the speaker's simulation will

be expressed as gestures. For example, a speaker who is describing throwing a Frisbee might move her arm in the way that she would if she were actually throwing a Frisbee, but she might not configure her hand in the way that she would if holding a real Frisbee. However, this might vary if the focus of her utterance were on the nature of her grip, rather than the direction of her throw. More generally, people can zero in on or *schematize* particular elements of visual, spatial, or motoric information within their simulations, and their gestures may express these specific elements in a more or less veridical way (Kita et al., 2017), depending on the nature of the simulation, the speaker's communicative or cognitive goals at that moment, and cultural expectations about appropriate gesture form.

In some cases, people may even use body parts to gesture that are different from those they would use in performing the actual actions that are being simulated. For example, a speaker talking about kicking a branch out of his path might depict this action using his right hand instead of his right foot. The speaker schematizes certain elements of the action (forceful forward movement to the right side of the body) and expresses those elements with his hand rather than his foot. There are several reasons why this might occur. It may be that movements are expressed with the hands because the gesturer is simultaneously speaking, and neural activation spreads readily from mouth to hands. Such an explanation suggests that gestures about nonmanual actions might be more likely to be expressed with the hands when the gesturer is simultaneously speaking than when he is not—a prediction that could be put to empirical test. Alternatively, movements may be expressed with the hands because of cultural norms that dictate that gestures should be produced manually. Young children, who have not yet learned this cultural norm, often use body parts other than the hands to gesture (for examples, see McNeill, 1992). This explanation suggests that cognitive effort might be associated with producing a manual gesture about a nonmanual action, because the motor code involved in the nonmanual simulation must be transferred to the hands.

Under both of these proposed explanations, it seems likely that speakers may not completely suppress nonmanual movements, even when they also gesture about the movement manually. For example, when a speaker gestures manually about kicking, he may also move his foot, but it may go unnoticed by addressees (and gesture researchers!) who are primarily focused on the speaker's hands. Furthermore, the movements may be very small; Popescu and Wexler (2012) found that head movements produced during a spatial imagery task that required realignment of the visual field were less than 1% of the amplitude of the actual movements necessary to perform the realignment. Future studies could investigate whether action simulations that involve body parts other than the hands also give rise to movements with the corresponding body parts, even when speakers also engage their hands.

Gestures are not always about actions

Some have interpreted the GSA framework as applying only to action pantomime gestures (e.g., P. Wagner, Malisz, & Kopp, 2014) or to gestures that directly mimic actions. Although the framework does account for such gestures, it also applies to other sorts of gestures, including gestures that depict visuospatial properties. For example, a speaker who says “it was a tall tree” might move her hands vertically, so as to depict the height of the tree she is describing. Although such a movement does not directly mimic action—she is not showing how she would climb the tree, for example—we argue that the origin of such a movement is still an action simulation.

Recall that the GSA framework follows other theories about the embodiment of cognition and perception (e.g., Gibson, 1979; Prinz, 1997) in holding that there are intrinsic links between action and perception, as well as among motor imagery, spatial imagery, and visual imagery. Visual imagery of a tree, in our view, activates spatial imagery of properties of the tree (such as its height or its orientation), as well as motor imagery of how one might interact with that tree (such as by climbing it). Thus, forming a visual image of a tree might evoke simulations of actions such as climbing, but it can also evoke simulations of more subtle actions, such as those involved in perceiving the tree or in thinking about the spatial properties of the tree.

In perceiving a tall tree, a speaker might move her head and eyes vertically to take in the height of the tree. This vertical movement could be reactivated when the speaker forms a mental image of the tree, just as the eye movements people make when perceiving a scene are reactivated when they visualize the scene in memory (e.g., Bone et al., 2018; Brandt & Stark, 1997; Laeng & Teodorescu, 2002). Furthermore, in thinking about the spatial properties of the tree, a speaker could activate actions that would highlight such properties in her mental image, for example, scanning from the base to the top of the tree. Scanning a shape in this way presumably involves creating a motor plan. Indeed, research suggests that overt or covert eye movements are involved in shifting attention within a spatial image (e.g., de Vito, Buonocore, Bonnefon, & Della Sala, 2014). Thus, when people form spatial images of objects or scenes, they may also activate motor information about moving between positions in the images.

Regardless of whether its origin is in perception or in imagery, we propose that the simulation of motion between spatial locations is overlaid on the hands during speech when it is strongly activated or when the threshold for producing a gesture is low. But how does the motor plan become realized specifically by the *hands*? Although there is ample evidence that oculomotor plans are involved in the formation and maintenance of spatial imagery (e.g., Pearson et al., 2014), there is much less evidence that *manual* movements are involved.

However, some evidence suggests that eye movements and hand movements are indeed related. First, manual movements facilitate the planning and execution of eye movements in the same direction, and they interfere with eye movements in a different direction (e.g., Niehorster, Siu, & Li, 2015). Second, manual movements can trigger (or disrupt) shifts in attention, just as eye movements can (Eimer, Forster, van Velzen, & Prabhu, 2005; Gherri & Eimer, 2010). Third, manual movements can facilitate (or interfere with) maintenance of spatial information in working memory (e.g., Lawrence et al., 2001). Thus, it seems possible that movements between spatial locations are represented in a way that can be realized by either the hands or the eyes, a proposal that makes sense, given the evolution of hand–eye coordination. Below we offer some speculative thoughts on how this could occur.

One possibility is that shared representations are formed for the shape of an object as revealed through vision and for the shape of the object as revealed through manual exploration. Several studies have shown that overlapping areas in the inferior temporal gyrus are activated when objects are discriminated on the basis of vision or touch (e.g., Pietrini et al., 2004), and some have suggested that spatial representations are multisensory, with ties to both vision and haptic perception (Lacey, Campbell, & Sathian, 2007). As such, it is possible that thinking about how an object looks necessarily activates information about how the object would feel if it were touched, and this tactile activation may be the catalyst for manual gestures that trace the object’s spatial properties.

A second possibility is that spatial representations activate the dorsal pathway, or the vision-for-action stream, of the visual system, and this activation may underlie manual movements. It is generally accepted that there are separate neural pathways for processing the “what” of vision and the “where” of vision, with the latter being especially involved in the coordination of reaching and grasping (e.g., Goodale, 2011). Furthermore, thinking about the size or shape of an object activates posterior parietal areas in the brain that are part of this dorsal (vision-for-action) stream (Oliver & Thompson-Schill, 2003). As applied to the present account, it is possible that thinking about spatial properties of objects or spatial locations within a scene activates the dorsal stream that is involved in the premotor planning of reaching and grasping. The activation of this action system could make manual movement more likely.

We contend that, regardless of the neural mechanisms involved, thinking about the spatial properties of a visual image activates motor representations of scanning, tracing, or moving between the locations of the parts of the image. From this perspective, gestures that depict spatial properties have their origins in the action simulations that are activated when the speaker thinks about those properties. In support of this view, there is some evidence that spatial properties do indeed evoke action plans that can be realized with the hands. Bach,

Griffiths, Weigelt, and Tipper (2010) asked participants to categorize objects on the basis of whether they are typically found outside or inside the home. Participants responded by gesturing either a circle shape or a square shape, depending on their categorization of the object. Critically, the words named objects that either had a circle shape (e.g., moon; dart board) or that had a square shape (e.g., billboard; book). Bach et al. found that participants were faster to produce gestures that matched the shapes of the objects they were categorizing than gestures that did not match. For example, they were faster to produce circle gestures when they were categorizing the location of circle-shaped objects than square-shaped objects. Furthermore, when participants responded to square objects with circle gestures, their circle gestures became more “square-like” in their shape. This evidence suggests that thinking about an object with a particular shape automatically activated a motor plan that traced the contour of that shape. Executing a gesture that deviated from that activated motor plan was difficult.

In sum, the GSA framework accounts for more than just pantomimes of manual actions. By *simulated action*, we refer to any action that is activated in thinking about images, including motor images, spatial images, and visual images. Such simulations can be very action-like, as in the case of motor imagery, but they can also be actions related to perceiving or highlighting the spatial elements of a visual image (see Alibali & Kita, 2010; Hostetter & Boncoddio, 2017, for more on this point). Furthermore, as we noted earlier, other types of imagery (such as auditory, tactile, or even olfactory imagery) have the potential to activate action simulations and be expressed in gestures, as well. Thinking about where a sound came from could activate a simulation of turning toward the sound; thinking about the smell of a rose could activate a simulation of lifting the rose and sniffing it.

Gestures are affected by language

Gestures differ across languages, and some of these differences coincide with differences in how languages “package” information in speech. For example, speakers of different languages display different patterns in gestures about path and manner in motion events, and these differences coincide with differences in how these languages package information in the verb and in satellites (e.g., Özyürek et al., 2005). Within a particular language, gestures are also affected by the specific syntactic framing chosen for a given utterance (e.g., Kita et al., 2007). Such evidence has been taken as supporting the view that the representations that give rise to gestures and those that give rise to speech interact during production (e.g., Kita & Özyürek, 2003). Furthermore, this perspective is sometimes portrayed as an alternative to the stance taken by the GSA framework that gestures arise from simulated action, which

could be construed as “not requiring explicit interactions between speech and gesture” (e.g., Özyürek, 2017, p. 41).

We do not see a conflict between the GSA framework and accounts that hold that gestures align with linguistic structure. In our view, both gestures and speech arise from thoughts that are sometimes grounded in embodied simulations of the world. The possibilities and constraints for linguistic expression in the language being spoken affect how speakers simulate events and which aspects of the events are most salient in their simulations; these differences in simulation affect both how speakers express information in gesture and how they express it in speech. Because speakers formulate their simulations in the interest of communicating about them, the details of simulations—at least simulations that underlie *co-speech* gestures—are shaped by linguistic factors. Co-thought gestures may be less affected by linguistic factors, as some studies have suggested (Özçalışkan, Lucero, & Goldin-Meadow, 2016a).

Gestures are designed for the listener

Recently, H. H. Clark (2016) proposed that iconic gestures belong to a class of behaviors he terms “depictions,” or intentional actions that are used to stage a scene in order to help others imagine that same scene. Nemirovsky and colleagues (Nemirovsky, Kelton, & Rhodehamel, 2012) offer a related proposal, arguing that gestures allow speakers and listeners to collectively imagine. Such accounts hold that gestures are designed deliberately, and perhaps consciously, by speakers for listeners.

The GSA framework does not deny that gestures can be produced quite deliberately by speakers in some situations. We consider these cases as situations in which the gesture threshold is lowered because a gesture is expected to be particularly beneficial for the listener or speaker, and this lowering of the threshold may occur quite consciously. Moreover, we see no reason why once such gestures are produced, they cannot become part of the speaker’s and listener’s common ground. However, according to the GSA framework, not all gestures are produced in this way, because speakers often gesture without awareness, and in some cases, they do not seem to intend their gestures to communicate (see Cooperrider, 2018, for further discussion of this issue).

At the same time, there is also evidence that speakers make subtle (and likely unconscious) changes to their gestures, depending on aspects of the discourse context, such as the knowledge (e.g., Jacobs & Garnham, 2007) or location (e.g., Özyürek, 2002) of their addressees. For example, speakers produce larger and more precise gestures when communicating information to someone who has not heard the information before than when speaking to someone who has (Galati & Brennan, 2013). Speakers also produce gestures higher in space when speaking to someone who has no knowledge of

the particular actions being described than when speaking to someone who has observed the actions (Hilliard & Cook, 2016). Speakers also produce smaller gestures when they do not want their listeners to understand well because they will soon be competing in a game (Hostetter, Alibali, & Schrager, 2011).

In the original formulation of the GSA framework, we considered three ways in which variations in the discourse context could lead to variations in gesture: by encouraging speakers to adjust the heights of their gesture thresholds, by affecting how strongly speakers activate particular elements of a simulation, and by affecting the coordination of the oral and manual systems at the moment of gesturing (Hostetter & Alibali, 2008, p. 506). At the same time, however, we believe that it is fair to say that, in its original formulation, the GSA framework did not provide a detailed or mechanistic account of how variations in the discourse context might give rise to variations in gesture form. We return to this point below.

Does the GSA framework account for gesture throughout development?

To date, most of the research that has been explicitly motivated by the GSA framework has focused on adults. Thus, one might reasonably ask whether the theory applies only to adults, or whether it applies throughout development. Indeed, the framework is intended as a general characterization of the processes that give rise to gestures, and as such, it should apply to infants and children, as well as to adults.

One tenet of the GSA framework is that gestures emerge along with speech because of sensorimotor linkages between the manual system and the oral–vocal system. These linkages are evident already in infancy. For example, there are close temporal relationships between infants' production of rhythmic arm movements and their reduplicated babble (e.g., Ejiri, 1998; Iverson, Hall, Nickel, & Wozniak, 2007). Iverson and Thelen (1999) argued that vocal and manual actions are mutually entrained through development, and that eventually, due to the sustained, close coordination of vocal and manual activity, motor activation of the mouth for producing speech automatically gives rise to activation of the hands. According to the GSA framework, this spreading activation—from the mouth to the hands—is one reason why many gestures are produced along with speech; motor activation from speech increases activation of the hands, making it more likely that the level of activation will exceed the gesture threshold. Importantly, because these oral–motor linkages are present from infancy onward, this process should operate in the same way in both children and adults.

A second tenet of the GSA framework is that gesture rates are dependent on a gesture threshold, which is sensitive to the presence of an audience or the cognitive difficulty of the task, among other factors. There is some evidence for the operation

of a gesture threshold in young children. For example, kindergarteners decrease their gesture rates when they cannot see their listeners (Alibali & Don, 2001). Children also adjust their gesture rates depending on the cognitive difficulty of the task; children gesture more when the conceptual demands of speaking are greater (Alibali, Kita, & Young, 2000).

A third tenet of the GSA framework is that gestures emerge when speakers activate visuospatial or motor simulations rather than solely verbal codes. Infants' gestural behavior provides support for this claim. Infants rely extensively on gestures to communicate during the one and two-word stages of language development (Iverson, Capirci, & Caselli, 1994; Iverson & Goldin-Meadow, 2005), and their early gestures often presage their speech, in the sense that infants express certain meanings initially in gestures and only later in speech (Acredolo & Goodwyn, 1988). The early emergence of gestural expressions makes sense from the perspective of the GSA framework; because gestures derive from simulated actions or perceptual states, the meanings of gestures are based in actions and perceptions, and early in development, these meanings may not be linked directly to words.

Over time, children become more proficient in speech, and as they rely more on verbal codes rather than action or perceptual simulations, their use of gesture decreases. However, some data suggest that young children's gestures remain closely tied to the corresponding actions. McNeill (1992) presents several compelling examples from children's narratives to make the case that children's gestures are essentially *enactments*, which use body parts and space in the same ways that the corresponding actions do. For example, when children gesture about actions that are performed with body parts other than the hands (such as kicking with the feet), they may use those same body parts to produce the gestures. In McNeill's view, children's gestures "lack full separation from action" (McNeill, 1992, p. 303)—as they might if they arise from highly activated action simulations, as the GSA framework contends.

Furthermore, much evidence suggests that children continue to express information in their gestures that is not also present in their speech throughout the elementary school years (e.g., Alibali, Evans, Hostetter, Ryan, & Mainela-Arnold, 2009; Church & Goldin-Meadow, 1986; Perry, Church, & Goldin-Meadow, 1988). Even though children's reliance on imagery and simulation decreases as they become more proficient with verbal representations, still, in some instances their verbal knowledge and their imagistic knowledge diverge. Gestures may be particularly likely in such situations (e.g., Mainela-Arnold, Alibali, Hostetter, & Evans, 2014), as the GSA framework predicts.

Whereas many studies have examined gesture in young children, much less is known about how children's gestures develop and change as they approach adulthood. Relatively few studies have directly compared gesture production in

children and adults. The available data are based mainly on cartoon-retell tasks, and they suggest that children gesture at lower rates than adults on such tasks (Alibali et al., 2009; Colletta, Pellenq, & Guidetti, 2010; Mayberry, Jacques, & DeDe, 1998; Reig Alamillo, Colletta, & Guidetti, 2012). At first glance, this seems potentially inconsistent with the idea that gestures should be more prevalent with rich imagistic simulations, because there is no reason to assume that children would rely *less* on visuospatial representations than adults. However, there is evidence that motor imagery and the ability to map action representations onto one's own body do not fully develop until adolescence (Choudhury, Charman, Bird, & Blakemore, 2007; Skoura, Vinter, & Papaxanthis, 2009; Spruijt, van der Kamp, & Steenbergen, 2015). Thus, it is possible that when narrating a cartoon (i.e., describing visuospatial information that was observed rather than performed), children gesture less than adults because the visuospatial imagery of the cartoon does not automatically activate their motor systems to the same degree as it does in adults. In contrast, when the task involves thinking about actions that one has performed oneself, there may be no differences in the gesture rates of children and adults, as research with problem-solving tasks has shown (Pouw, van Gog, Zwaan, Agostinho, & Paas, 2018).

Finally, it should be noted that interpreting differences in gesture rates between adults and children is complicated, because many developmental changes take place in children's cognitive and social skills—skills that the GSA framework proposes should matter for determining gesture rates. For example, inhibitory control is not fully developed in childhood (e.g., Williams, Ponesse, Schachar, Logan, & Tannock, 1999). If inhibitory control is needed to prevent highly activated simulations from being expressed as gestures, as the GSA framework contends, then children should gesture more than adults, all else being equal. The problem, of course, is that all else is not equal, as there are also developmental differences in narrative length and complexity (e.g., Colletta et al., 2010; Reig Alamillo et al., 2012), in imagery abilities (e.g., Spruijt et al., 2015), in sensitivity to the audience's knowledge (e.g., Fukumura, 2016), and in other verbal skills across development. Thus, it is difficult to use the GSA framework to interpret differences in gesture rates between children and adults, because so many factors that are proposed to affect gesture rates also change with development.

Even though comparing gestures across children and adults may not be fruitful as a means to test the claims of the GSA framework, the basic tenets of the GSA framework should operate in children, and as such, children's gestures should be sensitive to the same factors as adults' gestures. More research will be needed to systematically address key claims of the framework in children at various ages. For example, it would be valuable to test experimentally whether children gesture at higher rates when motor simulations are more

highly activated, as is the case for adults (e.g., Hostetter & Alibali, 2010).

Is the GSA framework falsifiable?

The GSA framework contends that gestures arise as a result of a dynamic relationship between (1) the level of activation of simulated actions and perceptual states and (2) the current height of the gesture threshold. However, neither of these factors is directly observable, and because both are proposed to matter, it is possible to explain the occurrence (or lack thereof) of any observed gesture. For example, if a speaker describes an idea closely tied to action but does not gesture, perhaps the gesture threshold was very high at that moment. If a speaker gestures while describing something very abstract, perhaps the speaker is thinking about a spatial metaphor for the idea being described, and that spatial image evokes an action plan that is expressed in gesture. Thus, it is indeed the case that observations of individual gestures cannot falsify the GSA framework.

However, we contend that the GSA framework *could* be falsified—or supported—in experiments that carefully manipulate the factors that, according to the framework, should influence gesture. Although it is not possible to know exactly how high an individual speaker's gesture threshold is at any one moment, it is possible to put speakers in communicative situations in which their thresholds should generally be lower and to observe their resulting gesture rates (e.g., Kelly et al., 2011). Similarly, it is possible to compare patterns of gesture use across different speaking topics and observe that rates are generally higher when the topics are more closely related to action (e.g., Pine et al., 2010).

The framework makes more nuanced predictions, as well, such as the notion that the gesture threshold (which varies with the communicative situation) and the level of activation on action simulations (which varies with speaking topic) should interact. Specifically, when action simulations are highly activated, the communicative situation should matter less than when simulations are less highly activated, because highly activated simulations should surpass even a very high threshold (see Hostetter, 2014). Furthermore, although gestures may have cognitive benefits in many speaking situations, the cognitive cost associated with *not* producing a gesture should be particularly apparent in situations in which a highly activated action simulation is present. Without such a simulation, no motor plan for gesture needs to be inhibited, so there should be no associated cognitive cost of not gesturing (although there could still be a cognitive benefit of producing a gesture in such situations).

In short, although it is not possible to directly observe how strongly action is being simulated or how high an individual's gesture threshold is at a given moment, these factors can be manipulated, and the GSA framework does make specific

predictions about the resulting effects on gesture. To the extent that these predictions are supported, the framework is supported. To the extent that they are not supported, the framework is falsified. We encourage experimenters who wish to test the claims of the GSA framework to preregister their predictions, so as to assure that post hoc explanations are not being applied in order to explain the results.

Conclusion

The GSA framework has not been without its critics, and we have attempted to summarize and respond to the main criticisms here. In our view, these criticisms do not undermine the key claims of the framework. However, this is not to imply that the framework is without limitations. In the following section, we turn our attention to several important limitations of the framework as it was originally proposed.

Limitations of the GSA framework

The GSA framework does not account for nonrepresentational gestures

The gestures that emerge from simulations of actions and perceptual states are most obviously iconic gestures, or gestures that enact or depict the motor and spatial aspects of what is being described. Indeed, in the original formulation of the GSA framework, only representational gestures were considered—including iconic gestures, metaphoric gestures, and deictic gestures. Beat gestures, the rhythmic gestures that align with speech prosody or that highlight discourse structure, were not considered (though see Hostetter, 2008). Furthermore, although deictic gestures that indicate particular locations in space were included under the broad category of representational gestures, the original framework did not specify how they might relate to underlying simulations.

Deictic gestures Alibali and Nathan (2012) described more specifically how the basic tenets of the GSA framework can be applied to deictic gestures. They suggested that pointing gestures reflect the speaker's indexing of ideas to objects or locations in space. Building on the indexical hypothesis of language comprehension (Glenberg & Robertson, 1999), which claims that language is indexed to specific objects in the environment, Alibali and Nathan (2012) proposed that this indexing occurs during language production, as well. When a person thinks or talks about a specific object that is present (or imagined to be present) in the environment, that person's cognitive system orients toward the object, and may simulate touching, reaching toward, or orienting the eyes and body

toward the object. This cognitive indexing may be expressed as a pointing gesture.

This idea aligns with theoretical arguments about the ontogenetic origins of pointing in reaching (e.g., Vygotsky, 1978) or in object exploration (e.g., Carpendale & Carpendale, 2010; Lock, Young, Service, & Chandler, 1990). The basic idea is that pointing derives from the orienting of attention, which may (at least in some cases) evoke simulations of reaching or touching. Of course, pointing is not only attentional—it is also social in nature, and it seems clear that social cognitive abilities, exposure to others' pointing behaviors, and cultural factors also influence pointing behavior (see, e.g., Cooperrider, Slotka, & Núñez, 2018; Matthews, Behne, Lieven, & Tomasello, 2012; Tomasello, Carpenter, & Liszkowski, 2007). Our focus here is on the attentional underpinnings of pointing, and the claim is that pointing can emerge from simulated orienting actions, even as it is also influenced by social and cultural factors.

One implication of this idea is that, although pointing gestures are produced in specific, culturally determined ways, they may vary subtly in form and execution depending on the underlying simulations. For example, Gonseth, Vilain, and Vilain (2013) have shown that speakers produce larger and more prolonged points to objects that are further away in space than to objects that are nearer, just as reaching for something that is farther away requires a larger, more prolonged movement than reaching for something that is closer. Thus, deictic gestures may be a special case of representational gestures, in which motoric aspects of the simulations are channeled to a more highly conventionalized, culturally specified form. At present, this idea is clearly speculative. However, some empirical predictions can be derived from this view. If deictic gestures are based on simulations of reaching or manual exploration, people may be more likely to point to objects they would like to touch (e.g., a clean, soft toy) than to objects they would not like to touch (e.g., a soiled, slimy toy). As a second example, people might point to objects using slightly different handshapes or hand orientations when encouraged to think about reaching for the objects versus touching the surfaces of the objects.

Beat gestures Beat gestures are small, rhythmic movements that often coincide with stress in speech (e.g., Leonard & Cummins, 2011; McClave, 1994). Within the GSA framework, beat gestures have been explained as occurring when there is close coordination between the oral and manual systems, without a strongly activated action simulation to enact on top of that coordination (Hostetter, 2008; see also Tuite, 1993). If this is the case, there should be an inverse relation between the number of representational gestures that occur with a description and the number of beats, because when there are fewer simulations to express as representational gestures, beats should be more prevalent. To test this idea,

Hostetter (2008) examined gesture rates as speakers described words that varied in how strongly they were associated with actions (e.g., *hammer* vs. *impossibility*). The number of representational gestures increased as the amount of action evoked by a word increased, but the number of beat gestures decreased. In a related study, however, Masson-Carro et al. (2016a) did not replicate this pattern. They conclude that representational and nonrepresentational (i.e., beat) gestures have their origins in different cognitive processes.

Some recent evidence supports the notion that beats may be simplified expressions of representational gestures. Yap, Brookshire, and Casasanto (2018) examined the beat gestures produced by speakers as they retold stories with spatial schemas implying a particular direction (e.g., a rocket taking off; grades improving). The researchers found that beat gestures moved in the direction implied in the story more often than would be expected by chance, and they concluded that beats can have iconic properties. This finding supports the notion that beat and representational gestures may emerge from the same cognitive processes.

To our knowledge, no current theoretical models of gesture simultaneously account for all types of gesture (both representational and nonrepresentational). Although the mechanisms specified in the GSA framework can be applied to explain nonrepresentational gestures, such as beats, more empirical data are needed to establish whether the predictions hold. For example, studies directly comparing the timing of deictic, beat, and iconic gestures relative to speech could be informative, as could more studies that examine whether deictic and beat gestures have iconic properties (e.g., Yap et al., 2018).

The GSA framework does not specify the function of gestures for speakers

Many researchers have argued that gestures serve a functional role in cognition (see Goldin-Meadow & Alibali, 2013, for a review). Gestures appear to help speakers resolve lexical difficulties (e.g., Krauss, 1998), package spatial information into the linear stream of speech (Alibali, Yeo, Hostetter, & Kita, 2017; Kita, 2000), and bypass the need to encode the information verbally (e.g., Hostetter & Alibali, 2011). Beyond their benefits for speech production, gestures also appear to influence cognitive processes more broadly. Gestures focus speakers' attention on perceptual information (Alibali & Kita, 2010; Hostetter & Boncoddio, 2017), they help speakers to remember information (Cook, Yip, & Goldin-Meadow, 2012b; So, Ching, Lim, Cheng, & Ip, 2014), they help speakers form new representations that are helpful in problem solving (Boncoddio, Dixon, & Kelly, 2010; Bucciarelli, Mackiewicz, Khemlani, & Johnson-Laird, 2016), and they help learners generalize what they have learned (Goldin-Meadow, 2016). Producing gestures appears to be more powerful for learning than producing actions on objects (Novack

et al., 2014; Stieff, Lira, & Scopelitis, 2016), using spatial language (So, Shum, & Wong, 2015), or seeing someone else produce gestures (Goldin-Meadow et al., 2012). Recent theoretical descriptions of how gestures might come to have their facilitative effects have focused on the role of gesture in schematizing information for speakers (Kita et al., 2017) or on their status as representational (rather than actual) actions (Novack & Goldin-Meadow, 2017).

The GSA framework does not specify how gestures come to have facilitative effects for speakers. However, it also does not deny that gestures are functional. In fact, the possibility that externalizing a simulation by expressing it as gesture might be helpful is one factor that has been proposed to affect the gesture threshold. In addition, we suggest that thinking about gestures as arising from an embodied cognitive system can be helpful in thinking about the functions they may serve. Many studies from an embodied cognition perspective indicate that movements of the body affect cognitive processing; for example, producing fluid rather than abrupt movements while drawing promotes more creative thinking (Slepian & Ambady, 2012), and tilting one's head appears to aid in mental rotation (Popescu & Wexler, 2012; Risko, Medimorec, Chisholm, & Kingstone, 2014). Thinking about gesture as arising from the embodied mind situates gesture within the embodied cognition paradigm and suggests parallels with these other demonstrated effects.

Furthermore, recent accounts of the embodiment of cognitive processes have focused, not just on how the mind simulates real-world interactions, but on how these simulations are used to *predict* the sensorimotor consequences of interactions with the environment (e.g., A. Clark, 2013; Glenberg & Gallese, 2012; Lupyan & Clark, 2015; Pezzulo & Cisek, 2016). According to these perspectives, the general task of the cognitive system is to predict incoming sensory information as accurately as possible. When actual sensory information from the environment conflicts with the prediction generated by the cognitive system, this yields "prediction error," which the system is motivated to reduce. In the case of a simulated action that is not actually produced (e.g., imagined rotation of an object), the cognitive system predicts the sensorimotor consequences of actually producing the action (e.g., how the object will look after it is rotated). However, because the action has not actually been produced, the incoming sensorimotor information from the environment (e.g., the shape of the unrotated object) does not match the predicted sensorimotor consequences that have been generated by the system (e.g., how the object should look after rotation), resulting in high prediction error. One way to minimize the prediction error is to actually produce the action that has been simulated (e.g., to physically rotate the object), thereby bringing the incoming sensorimotor information more into line with the predictions.

Pouw and Hostetter (2016) have proposed that the same system is in operation, even when no object is present to actually act on, and such attempts to reduce prediction error

through action are realized as gesture. For example, when speakers imagine rotating a picture of an object, they simulate what the object would look like after rotation, but that mental image does not match the incoming sensorimotor information of the unrotated picture. To resolve the discrepancy, they attempt to produce an action that will bring the world more in line with their prediction. No actual object is present to manipulate, so the action is realized as a gesture. According to this view, gestures derive from an attempt to predict the sensorimotor consequences of an action in the world—indeed, the attempt to predict is, in essence, a simulation. At the same time, producing the gesture has the effect of relieving demands on the cognitive system by reducing the prediction error that is inherent to thinking about an action without actually producing it, and producing the gesture can provide multimodal information about the sensorimotor consequences of actions. In this way, there is a direct correspondence between the origin of gesture (e.g., efforts to resolve prediction error that are inherent to simulating an action without producing it) and the function those gestures come to have for the cognitive system (i.e., making the consequences of those actions more apparent and thereby reducing prediction error).

Nathan and Martinez (2015) offered a related account of how gesture can both reflect and be causally involved in the formation of a situation model when reading a text. They proposed an elaboration of the GSA framework, which they call the Gesture as Model Enactment (GAME) framework. In the GAME framework, gestures occur when speakers describe knowledge that activates a situation model, or a mental model that combines new information with long-term knowledge to produce relational and predictive understanding. Situation models may involve visual, spatial, or motor imagery that activates gesture, as articulated in the GSA framework. However, in the GAME framework, sensorimotor information from gesture feeds back to inform and update the current situation model by activating a set of predictor–controller modules that simultaneously predict the various potential model-based consequences of each intended action. As gestures are enacted, they help prune these choices to those situation models that are most plausible, given the activated knowledge. When gestures suggest new information about the situation model, this information is incorporated into the imagery that supports revision of the situation model, thereby facilitating more accurate prediction and inference making. The result is that gestures not only reflect, but also influence, the construction of situation models. Nathan and Martinez supported this claim with data showing that speakers who gestured while describing a text about the human circulatory system were more successful at making inferences that went beyond the text than were speakers who engaged in a spatial tapping task (and who therefore were unable to gesture).

Although some have claimed that thinking about the source of gesture and thinking about the function of gesture are separate endeavors (e.g., Novack & Goldin-Meadow, 2017), we

argue that considering the source of gestures can inform theories about their function. Specifically, if gestures arise from a mind that is engaged in active simulation, this may explain how gestures accomplish the functions they do. For example, if the verbal codes of language are associated with and activated by actions (see Glenberg & Gallese, 2012) and gestures are simulations of these actions, then it becomes clear how gesture could function to influence lexical access. Furthermore, if the motor plan for a gesture is largely formed from a simulation that already underlies speaking and thinking, it becomes clear how planning and producing a gesture could actually require fewer cognitive resources than inhibiting that movement. If gestures are simulations and simulations are predictive, then it becomes clear how gestures can generate information that helps thinking and reasoning.

The GSA framework does not specify how (or whether) gestures function for listeners

Just as the GSA framework does not directly address how gestures might facilitate cognition for speakers who produce them, it also does not address whether gestures have facilitative effects for listeners. Although a speaker's belief about whether listeners might benefit from gesture is one factor that is hypothesized to affect the gesture threshold (and consequently, the speaker's gesture rate), the framework does not make strong claims regarding whether listeners actually do benefit from gestures. Nonetheless, some researchers have used the basic tenets of the GSA framework to speculate about how gestures may have communicative effects for listeners (e.g., Austin & Sweller, 2014), and there is substantial evidence that listeners do benefit from seeing speakers' gestures in many situations (e.g., Hostetter, 2011).

One proposal for how gestures might come to have their facilitative effects for listeners is that seeing a speaker gesture activates a simulation of the gesture in the listeners' motor systems, or in some cases, encourages the listener to actually produce a similar gesture (e.g., Alibali & Hostetter, 2010; Iani & Bucciarelli, 2017; Ping, Goldin-Meadow, & Beilock, 2014; Quandt, Marshall, Shipley, Beilock, & Goldin-Meadow, 2012). Indeed, seeing others' gestures prompts listeners to use similar gestures when they talk about the same topics (Child, Theakston, & Pika, 2014; Kimbara, 2006; Mol, Krahmer, Maes, & Swerts, 2012; Morett, 2018).

Moreover, there is evidence that the processes involved in producing gesture and those involved in comprehending gesture might both be based on simulations of actions and their relation to mental imagery. Listeners benefit less from gesture when they engage in a simultaneous task that involves the hands and arms than when they are not engaged in such a task (Iani & Bucciarelli, 2017; Ping et al., 2014), suggesting that when it is difficult to simulate the speaker's gestures with one's own effectors, the ability to understand and benefit from

the gestures is also decreased. This explanation aligns well with the GSA framework, and suggests that both gesture production and gesture comprehension are based on simulations of actions. In the case of gesture production, activation of a mental image leads to simulation and a motor plan that is expressed as gesture, as we have argued. In the case of gesture comprehension, seeing a gesture evokes a motor plan, which then activates a corresponding mental image that can promote comprehension of the message. One outstanding question is whether seeing gestures promotes comprehension more than seeing actions, or whether seeing gesture and seeing actions yield similar simulations and comparable increases in understanding.

The GSA framework does not account fully for adaptations in gesture form

The primary focus of the GSA framework is on predicting the likelihood of a gesture occurring at a particular moment. However, gesture frequency is not the only dimension of gestural behavior that a successful theory should explain. Many dimensions of gesture form, including viewpoint (Parrill, 2010), orientation (Özyürek, 2002), size (Hostetter, Alibali, & Schrager, 2011), complexity (Parrill et al., 2013), precision (Galati & Brennan, 2013), location (Hilliard & Cook, 2016), and the representational technique used (e.g., imitating, drawing, or molding; Masson-Carro et al., 2016a; Müller, 1998), can be affected by the cognitive or communicative situation. A complete account of gesture should offer an explanation, not only for whether or not a gesture occurs, but also for the details of its execution.

Within the GSA framework, variations in gesture form are attributable to differences in how people simulate the information that they are speaking and/or gesturing about. In our view, these simulations can be affected by the cognitive or communicative situation, and such differences in simulations are reflected in qualitative differences in gestures. For example, Hostetter and Alibali (2008) argued that variations in gesture viewpoint are the result of variations in the type of simulation underlying speech; simulations of motor imagery (e.g., first-person actions) were hypothesized to give rise to character viewpoint gestures, whereas simulations of visuospatial imagery were hypothesized to give rise to observer viewpoint gestures. Along similar lines, variations in the representational techniques used in iconic gestures are thought to depend on variations in simulated or actual experience with the objects being iconically represented. When people think about manipulable objects, they tend to simulate *using* those objects, and therefore produce gestures that enact those actions, whereas when people think about nonmanipulable objects, they tend to simulate *touching* or *viewing* those objects, and therefore produce gestures that rely on molding, sculpting, or tracing (Masson-Carro et al., 2016a).

Furthermore, we contend that the amount of detail contained in a simulation may also affect aspects of gesture form. If speakers simulate the precise details rather than a broad overview of an event, this could lead to gestures that are relatively complex, as those additional details will be reflected in gesture (e.g., Parrill et al., 2013). For example, if a speaker imagines the way that a squirrel's feet moved as it ran, this could be reflected in a gesture that shows the squirrel's wiggly feet, in addition to its trajectory across space.

This explanation fits well with evidence that people think about events at different levels of abstraction. According to construal-level theory (Trope & Liberman, 2010), the amount of detail included in an event or object representation depends on the perceived psychological distance of the event or object. If an event is perceived to be further away in time or space or is perceived as unlikely to happen, the event is thought about more abstractly and with less detail than if the event is perceived as near or as very likely to occur. Similarly, the amount of detail that is included in a simulation could depend on communicative factors. When speakers talk with someone to whom they feel psychologically close, they may form more detailed, concrete simulations (see Amit, Waksak, & Trope, 2013) that result in more precise and complex gestures. This could be one reason why people high in empathy produce more salient (i.e., larger, more visible) gestures than people low in empathy (Chu et al., 2014). Furthermore, if a speaker knows that the listener has no knowledge of what he or she is describing and is motivated to help the listener understand, the speaker may imagine more precisely the information that the listener needs to know, which could lead to more precise and detailed gestures (Galati & Brennan, 2013).

Thus, we suggest that speakers may alter their simulations—and as a consequence, their gestures—depending on the communicative situation. In line with this idea, speakers also modify their gestures when there is a “trouble spot” in communication that stems from a lack of common ground. For example, when students in math classrooms indicate lack of understanding of lesson content (by saying “I don’t get it” or by asking questions), teachers adjust their gestures, often increasing their gesture rates, and sometimes breaking information into smaller units, depicting additional details, or adjusting the placement of the gestures (for examples, see Alibali, Nathan, et al., 2013; Nathan & Alibali, 2011). These gestural adaptations suggest that, when common ground is breached, speakers may “zoom in” on relevant parts of the simulation, or resimulate the content from the listener’s point of view, in an effort to make their communication more successful.

Although many variations in gesture form (e.g., viewpoint, complexity) may be traced to variations in the underlying simulation, such an explanation cannot account for every type of change in gesture form that has been documented in the literature. For instance, there is some evidence that speakers

produce their gestures higher in space when they think the listener might not understand (Hilliard & Cook, 2016) and that speakers produce larger gestures when they really want the listener to understand (Hostetter, Alibali, & Schrager, 2011). Although it is possible, we doubt that such differences arise from the features of size or location being simulated differently depending on the communicative context; instead, it seems more likely that these differences occur during motor execution, rather than during the simulation and initiation of the movement. Just as actions generally begin as only a crude motor plan that is then shaped online during the movement by sensory feedback and predictions about the consequences of the current action (e.g., Cisek, 2005), it seems likely that the precise kinematics of a gesture are shaped as it is being executed. How this process may unfold in the case of gesture remains to be specified, but it seems likely that it is affected by complex top-down information regarding audience knowledge and what is situationally or culturally appropriate.

Cultural variations in gesture are ubiquitous, and these variations may also be manifest as gestures are being executed. The size of the gesture space differs across cultures (Kita, 2009), as does the preferred effector used in pointing. In Ghana, speakers point only with their right hand; there is a taboo on pointing with the left (Kita & Essegbey, 2001). In Papua New Guinea, speakers sometimes point with their nose or head (Cooperrider et al., 2018), whereas in Laos, people point with their lips (Enfield, 2001). Such cultural differences in gesture form are not accounted for in the GSA framework. However, we believe that these differences may emerge after the initial motor plan for a gesture is formed. That is, once the motor activation involved in simulation exceeds the gesture threshold and a gesture is initiated, the form of the unfolding gesture based on this motor activation can be further shaped by cultural or situational constraints on gesture form, size, and so forth.

Presumably, children learn such cultural conventions over development. Indeed, the tendency of adults in most cultures to gesture primarily with the hands, rather than with other articulators (such as the head), also seems to be learned. Although systematic data on this issue are lacking, McNeill (1992) claims that young children often use body parts other than the hands to gesture, as discussed above. Adults' tendency to gesture with the hands, even to represent actions produced by the legs and feet, may be one example of a cultural constraint on gesture form that is applied once an action simulation surpasses threshold for gesture production. However, as discussed earlier, applying the constraint may require cognitive effort and may not result in the complete inhibition of movement in the feet.

In sum, the GSA framework explains how a gesture is “born”—that is, when it is most likely to occur. The original framework did not specify fully how the form of the gesture is selected or executed. Here, we suggest that the specifics of

gesture form may be the combined result of differences in simulation as well as the application of constraints to the movement as it is executed. However, it is clear that additional theoretical and empirical work on this issue will be needed.

The precise neural mechanisms involved in simulation and gesture are underspecified

The GSA framework has its roots in evidence that cognitive processes involve the activation of neural areas that are used in actually interacting with the world (e.g., Pezzulo & Cisek, 2016). At the same time, however, the precise neural mechanisms involved in gesture remain underspecified. For example, what exactly is a “simulation” in the brain? There is disagreement over this issue, even among cognitive neuroscientists (e.g., Pezzulo et al., 2013), and the GSA framework does not take a clear stance. How does simulation differ, both phenomenologically and neurally, from motor imagery, emulation, or mental models (see Iani & Bucciarelli, 2017; O'Shea & Moran, 2017)? Again, the GSA framework does not imply a clear stance on these issues.

Furthermore, when an action is simulated, how specific is the activation in the motor cortex? That is, when speakers think about “kicking,” is it the case that specifically the foot area of motor cortex is activated (e.g., Hauk, Johnsrude, & Pulvermüller, 2004) and that this motor activation somehow gets transferred to the hand during speaking and gesturing? Or is the simulation of “kicking” a more general activation of the motor system to engage in a forward, forceful trajectory that could just as easily be produced with the hand as with the foot? These questions are beyond the level of specificity provided in the GSA framework.

Interest in understanding the brain bases of gesture has increased in the years since the GSA framework was formulated (e.g., I. Helmich & Lausberg, 2014; Hilverman, Cook, & Duff, 2016). Because the precise neural mechanisms involved in “simulation” are unspecified, it is difficult to address how the results of such studies relate to the GSA framework. For example, Hilverman et al. compared the gestures of patients with hippocampal damage (but whose neural systems supporting action understanding are presumably intact) to those of age-matched controls, and found that the hippocampal patients described fewer episodic details and also gestured at lower rates when asked to describe memories (e.g., how to make a sandwich). The researchers concluded that gestures originate from representations in the hippocampus, rather than from action representations, and contend that these results are incompatible with the claim in the GSA framework that gestures arise from action representations. We suggest that the evidence that gesture rates decrease with a decrease in episodic detail is compatible with the GSA framework, in that we would expect less vivid imagery to result in both fewer details being mentioned in speech and fewer gestures. Thus, it could

be that impaired hippocampal functioning leads to less rich imagery about events from one's past (e.g., Bird, Bisby, & Burgess, 2012), and these impoverished images activate action simulations less strongly and lead to fewer gestures.

Indeed, the GSA framework makes predictions about gesture based on the details of underlying simulations; however, the exact nature of these simulations is unknowable at the present time. This means that, to some extent, it is possible to explain the occurrence of any gesture by appealing to the nature of the underlying simulation. To get around this issue, independent measures of simulation are needed, and such measures are not yet available. Nonetheless, even as the precise details of a simulation are unknowable, we do believe that we can alter the likelihood that speakers simulate something in a particular way by altering their experience in particular ways. For example, giving people experience making the pattern they describe increases the amount of gesture they use, because—we contend—their simulation is more likely to include strong activation of action simulations about making the pattern (Hostetter & Alibali, 2010). As a second example, giving people a concurrent task that taxes their action planning system might decrease the amount of gestures they use, because—we contend—their simulations might be less rich or less highly activated.

Importantly, the GSA framework is not a neural or computational model; it was proposed as a general framework for relating gesture to the embodiment of the mind. To the extent that it is helpful to think about gesture in this way, it is a useful framework, and it does yield concrete behavioral predictions that can be tested empirically. However, without further specification of the neural mechanisms involved in simulation, this framework is unlikely to make useful predictions about the specific brain regions or neural mechanisms involved in gesture production and comprehension.

Conclusions

Ten years after we initially proposed the GSA framework, we have sought to evaluate the support for the framework and to consider its limitations. Our review of the literature indicates that the framework's main predictions have been largely supported. At the same time, challenges to the framework have highlighted its limitations and have underscored important questions for future research. Three sets of issues are central. First, more research will be needed to fully understand the mechanisms that give rise to systematic variations in gesture form, including variations in size, placement, representational techniques, and so on. Second, research will be needed to elucidate how different types of gestures (co-thought vs. co-speech; representational vs. beat) relate to one another and to the cognitive processes that underlie them. Third, reliable measures of simulation that are independent from gesture

would be helpful in further testing the claims of the framework.

In sum, it is clear that gestures are an important aspect of human behavior that have the potential to both reflect and affect cognitive processes. As the study of gesture continues to increase in psychology, education, and related fields, it is important that gesture researchers consider how theories of gesture can inform and be informed by ongoing discussions in neuroscience and cognitive psychology. The GSA framework has been an influential account of how gestures relate to embodied cognition, and its predictions have thus far been supported. Thus, we continue to argue that gestures do—at least in many cases—reflect a mind that is actively engaged in embodied simulation.

Author note We thank Rebecca Boncoddio, Mingyuan Chu, Breckie Church, Susan Wagner Cook, Susan Goldin-Meadow, Michael Kaschak, Spencer Kelly, Sotaro Kita, Mitchell Nathan, Wim Pouw, and Amelia Yeo for helpful discussions that have informed our thinking over the years. We also thank Colleen Bruckner for help in gathering source material for this review.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Acredolo, L. P., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, 59, 450–466. <https://doi.org/10.2307/1130324>
- Akhavan, N., Nozari, N., & Göksun, T. (2017). Expression of motion events in Farsi. *Language, Cognition, and Neuroscience*, 32, 792–804. <https://doi.org/10.1080/23273798.2016.1276607>
- Alibali, M. W., Boncoddio, R., & Hostetter, A. B. (2014). The role of gesture in reasoning and problem solving. In L. Shapiro (Ed.), *Routledge handbook of embodied cognition* (pp. 150–159). New York, NY: Routledge.
- Alibali, M. W., & Don, L. S. (2001). Children's gestures are meant to be seen. *Gesture*, 1, 113–127. <https://doi.org/10.1075/gest.1.2.02ali>
- Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., & Mainela-Arnold, E. (2009). Gesture–speech integration in narrative discourse: Are children less redundant than adults? *Gesture*, 9, 290–311. <https://doi.org/10.1075/gest.9.3.02ali>
- Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188. <https://doi.org/10.1006/jmla.2000.2752>
- Alibali, M. W., & Hostetter, A. B. (2010). Mimicry and simulation in gesture comprehension (Commentary on P. Niedenthal, M. Maringer, M. Mermillod, & U. Hess, The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression). *Behavioral and Brain Sciences*, 33, 433–434.
- Alibali, M. W., & Hostetter, A. B. (2018). *When hand movements are prohibited, do people use other body parts to gesture?* Manuscript in preparation.
- Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture*, 10, 3–28. <https://doi.org/10.1075/gest.10.1.02ali>
- Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and*

- Cognitive Processes*, 15, 593–613. <https://doi.org/10.1080/016909600750040571>
- 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.
- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, 21, 247–286. <https://doi.org/10.1080/10508406.2011.611446>
- Alibali, M. W., Nathan, M. J., Church, R. B., Wolfgram, M. S., Kim, S., & Knuth, E. J. (2013). Gesture and speech in mathematics lessons: Forging common ground by resolving trouble spots. *ZDM—International Journal on Mathematics Education*, 45, 425–440. <https://doi.org/10.1007/s11858-012-0476-0>
- Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Jacobs, S. A., Martinez, C. J., & Knuth, E. J. (2014). How teachers link ideas in mathematics instruction using speech and gesture: A corpus analysis. *Cognition and Instruction*, 32, 65–100. <https://doi.org/10.1080/07370008.2013.858161>
- Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gesture influences strategy choices in problem solving. *Psychological Science*, 22, 1138–1144. <https://doi.org/10.1177/0956797611417722>
- Alibali, M. W., Yeo, A., Hostetter, A. B., & Kita, S. (2017). Representational gestures help speakers package information for speaking. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking, and communicating* (pp. 15–37). Philadelphia, PA: Benjamins.
- Alibali, M. W., Young, A. G., Crooks, N. M., Yeo, A., Wolfgram, M. S., Ledesma, I. M., ... Knuth, E. J. (2013). Students learn more when their teacher has learned to gesture effectively. *Gesture*, 13, 210–233. <https://doi.org/10.1075/gest.13.2.05ali>
- Amit, E., Waksalak, C., & Trope, Y. (2013). The use of visual and verbal means of communication across psychological distance. *Personality and Social Psychology Bulletin*, 39, 43–56. <https://doi.org/10.1177/0146167212460282>
- Annett, J. (1990). Relations between verbal and gestural explanations. In C. Hammond (Ed.), *Cerebral control of speech and limb movements* (pp. 327–346). Amsterdam, The Netherlands: Elsevier Science.
- Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, 122, 92–103. <https://doi.org/10.1016/j.jecp.2013.12.008>
- Bach, P., Griffiths, D., Weigelt, M., & Tipper, S. P. (2010). Gesturing meaning: Non-action words activate the motor system. *Frontiers in Human Neuroscience*, 4, 1–12. <https://doi.org/10.3389/fnhum.2010.00214>
- Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58, 495–520. <https://doi.org/10.1016/j.jml.2007.02.004>
- Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, 13, 63–92. <https://doi.org/10.1075/gest.13.1.03bav>
- Beilock, S., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, 21, 1605–1610. <https://doi.org/10.1177/0956797610385353>
- Bird, C. M., Bisby, J. A., & Burgess, N. (2012). The hippocampus and spatial constraints on mental imagery. *Frontiers in Human Neuroscience*, 6, 142:1–12. <https://doi.org/10.3389/fnhum.2012.00142>
- Boncoddo, R., Dixon, J. A., & Kelly, E. (2010). The emergence of a novel representation from action: Evidence from preschoolers. *Developmental Science*, 13, 370–377. <https://doi.org/10.1111/j.1467-7687.2009.00905.x>
- Bone, M. B., St-Laurent, M., Dang, C., McQuiggan, D. A., Ryan, J. D., & Buchsbaum, B. R. (2018). Eye movement reinstatement and neural reactivation during mental imagery. *Cerebral Cortex*. Advance online publication. <https://doi.org/10.1093/cercor/bhy014>
- Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, 9, 27–38. <https://doi.org/10.1162/jocn.1997.9.1.27>
- Bub, D. N., Masson, M. E. J., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition*, 106, 27–58. <https://doi.org/10.1016/j.cognition.2006.12.010>
- Bucciarelli, M., Mackiewicz, R., Khemlani, S. S., & Johnson-Laird, P. N. (2016). Children's creation of algorithms: Simulations and gestures. *Journal of Cognitive Psychology*, 28, 297–318. <https://doi.org/10.1080/20445911.2015.1134541>
- Campisi, E., & Mazzone, M. (2016). Do people intend to gesture? A review of the role of intentionality in gesture production and comprehension. *Reti, Saperi, Linguaggi: Italian Journal of Cognitive Sciences*, 2, 285–300. <https://doi.org/10.12832/85417>
- Carpendale, J. I. M., & Carpendale, A. B. (2010). The development of pointing: From personal directedness to interpersonal direction. *Human Development*, 53, 110–126. <https://doi.org/10.1159/000315168>
- Casey, S., Emmorey, K., & Larrabee, H. (2012). The effects of learning American Sign Language on co-speech gesture. *Bilingualism: Language and Cognition*, 15, 677–686. <https://doi.org/10.1017/S1366728911000575>
- Cevasco, J., & Ramos, F. M. (2013). The importance of studying prosody in the comprehension of spontaneous spoken discourse. *Revista Latinoamericana de Psicología*, 45, 21–33.
- Child, S., Theakston, A., & Pika, S. (2014). How do modelled gestures influence preschool children's spontaneous gesture production? Social versus semantic influence. *Gesture*, 14, 1–25. <https://doi.org/10.1075/gest.14.1.01chi>
- Chisholm, J. D., Risko, E. F., & Kingstone, A. (2014). From gestures to gaming: Visible embodiment of remote actions. *Quarterly Journal of Experimental Psychology*, 67, 609–624. <https://doi.org/10.1080/17470218.2013.823454>
- Choudhury, S., Charman, T., Bird, V., & Blakemore, S.-J. (2007). Development of action representation during adolescence. *Neuropsychologia*, 45, 255–262. <https://doi.org/10.1016/j.neuropsychologia.2006.07.010>
- Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, 137, 706–723. <https://doi.org/10.1037/a0013157>
- Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140, 102–116. <https://doi.org/10.1037/a0021790>
- Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 257–270. <https://doi.org/10.1037/a0021790>
- Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, 143, 694–709. <https://doi.org/10.1037/a0033861>
- Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23, 43–71. [https://doi.org/10.1016/0010-0277\(86\)90053-3](https://doi.org/10.1016/0010-0277(86)90053-3)
- Church, R. B., Kelly, S., & Holcomb, D. (2014). Temporal synchrony between speech, action, and gesture during language production.

- Language, Cognition, and Neuroscience*, 29, 345–354. <https://doi.org/10.1080/01690965.2013.857783>
- Cisek, P. (2005). Neural representations of motor plans, desired trajectories, and controlled objects. *Cognitive Processing*, 6, 15–24. <https://doi.org/10.1007/s10339-004-0046-7>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, 123, 324–347. <https://doi.org/10.1037/rev0000026>
- Colletta, J.-M., Pellenq, C., & Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Communication*, 52, 565–576. <https://doi.org/10.1016/j.specom.2010.02.009>
- Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, 113, 98–104. <https://doi.org/10.1016/j.cognition.2009.06.006>
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012a). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, 27, 594–610. <https://doi.org/10.1080/01690965.2011.567074>
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012b). Gesturing makes memories that last. *Journal of Memory and Language*, 63, 465–475. <https://doi.org/10.1016/j.jml.2010.07.002>
- Cooperrider, K. (2018). Foreground gesture, background gesture. *Gesture*, 16, 176–202. <https://doi.org/10.1075/gest.16.2.02coo>
- Cooperrider, K., Slotta, J., & Núñez, R. (2018). The preference for pointing with the hand is not universal. *Cognitive Science*, 42, 1375–1390. <https://doi.org/10.1111/cogs.12585>
- Cornoldi, C., & Vecchi, T. (2003). *Visuo-spatial working memory and individual differences*. New York, NY: Psychology Press.
- de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, 4, 232–248. <https://doi.org/10.1111/j.1756-8765.2012.01183.x>
- de Vito, S., Buonocore, A., Bonnefon, J. F., & Della Sala, S. (2014). Eye movements disrupt spatial but not visual mental imagery. *Cognitive Processes*, 15, 543–549. <https://doi.org/10.1007/s10339-014-0617-1>
- Demir, Ö. E., Levine, S. C., & Goldin-Meadow, S. (2015). A tale of two hands: Children's early gesture use in narrative predicts later narrative structure in speech. *Journal of Child Language*, 42, 662–681. <https://doi.org/10.1017/S0305000914000415>
- Eielts, C., Pouw, W., Ouwehand, K., van Gog, T., Zwaan, R. A., & Paas, F. (2018). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychological Research*. Advance online publication. <https://doi.org/10.1007/s00426-018-1065-9>
- Eigsti, I.-M. (2013). A review of embodiment in autism spectrum disorder. *Frontiers in Psychology*, 4, 224:1–10. <https://doi.org/10.3389/fpsyg.2013.00224>
- Eimer, M., Forster, B., van Velzen, J., & Prabhu, G. (2005). Covert manual response preparation triggers attentional shifts: ERP evidence for the premotor theory of attention. *Neuropsychologia*, 43, 957–966. <https://doi.org/10.1016/j.neuropsychologia.2004.08.011>
- 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.
- Enfield, N. J. (2001). "Lip-pointing": A discussion of form and function with reference to data from Laos. *Gesture*, 1, 185–211. <https://doi.org/10.1075/gest.1.2.06enf>
- Fukumura, K. (2016). Development of audience design in children with and without ASD. *Developmental Psychology*, 52, 71–87. <https://doi.org/10.1037/dev0000064>
- Galati, A., & Brennan, S. E. (2013). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition, and Neuroscience*, 29, 435–451. <https://doi.org/10.1080/01690965.2013.796397>
- Gerofsky, S. (2010). Mathematical learning and gesture: Character viewpoint and observer viewpoint in students' gestured graphs of functions. *Gesture*, 10, 321–343. <https://doi.org/10.1075/gest.10.2-3.10ger>
- Gherri, E., & Eimer, M. (2010). Manual response preparation disrupts spatial attention: An electrophysiological investigation of links between action and attention. *Neuropsychologia*, 48, 961–969. <https://doi.org/10.1016/j.neuropsychologia.2009.11.017>
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.
- Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition*, 132, 174–180. <https://doi.org/10.1016/j.cognition.2014.03.012>
- Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48, 905–922. <https://doi.org/10.1016/j.cortex.2011.04.010>
- Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, 28, 1–26. <https://doi.org/10.1080/01638539909545067>
- Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science*, 8, 573–585. <https://doi.org/10.1177/1745691613498098>
- Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (2015). Spontaneous gesture and spatial language: Evidence from focal brain injury. *Brain and Language*, 150, 1–13. <https://doi.org/10.1016/j.bandl.2015.07.012>
- Goldinger, S. D., Papesch, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, 4, 959–978. <https://doi.org/10.3758/s13423-015-0860-1>
- Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Belknap Press.
- Goldin-Meadow, S. (2016). Using our hands to change our minds. *WIREs Cognitive Science*, 8, e1368:1–6. <https://doi.org/10.1002/wcs.1368>
- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. <https://doi.org/10.1146/annurev-psych-113011-143802>
- Goldin-Meadow, S., & Beilock, S. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5, 664–674. <https://doi.org/10.1177/1745691610388764>
- Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, 15, 876–884. <https://doi.org/10.1111/j.1467-7687.2012.01185.x>
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12, 516–522. <https://doi.org/10.1111/1467-9280.00395>
- Gonseth, C., Vilain, A., & Vilain, C. (2013). An experimental study of speech/gesture interactions and distance encoding. *Speech Communication*, 55, 553–571. <https://doi.org/10.1016/j.specom.2012.11.003>
- Goodale, M. A. (2011). Transforming vision into action. *Vision Research*, 51, 1567–1587. <https://doi.org/10.1016/j.visres.2010.07.027>
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377–396. <https://doi.org/10.1017/S0140525X04000093>

- Gu, Y., Mol, L., Hoetjes, M., & Swerts, M. (2017). Conceptual and lexical effects on gestures: The case of vertical spatial metaphors for time in Chinese. *Language, Cognition, and Neuroscience*, 32, 1048–1063. <https://doi.org/10.1080/23273798.2017.1283425>
- Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, 14, 53–82. <https://doi.org/10.1075/pc.14.1.05gul>
- Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, 33, 251–277. <https://doi.org/10.1007/s10919-009-0073-2>
- Hall, R., & Nemirovsky, R. (2012). Introduction to the special issue: Modalities of body engagement in mathematical activity and learning. *Journal of the Learning Sciences*, 21, 207–215. <https://doi.org/10.1080/10508406.2011.611447>
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–307. [https://doi.org/10.1016/S0896-6273\(03\)00838-9](https://doi.org/10.1016/S0896-6273(03)00838-9)
- Helmich, I., & Lausberg, H. (2014). Hand movements with a phase structure and gestures that depict action stem from a left hemispheric system of conceptualization. *Experimental Brain Research*, 232, 3159–3173. <https://doi.org/10.1007/s00221-014-4006-x>
- Helmich, R. C., de Lange, F. P., Bloem, B. R., & Toni, I. (2007). Cerebral compensation during motor imagery in Parkinson's disease. *Neuropsychologia*, 45, 2201–2215. <https://doi.org/10.1016/j.neuropsychologia.2007.02.024>
- Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Research*, 1428, 71–79. <https://doi.org/10.1016/j.brainres.2011.06.026>
- Hilliard, C., & Cook, S. W. (2016). Bridging gaps in common ground: Speakers design their gestures for their listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 91–103. <https://doi.org/10.1037/xlm0000154>
- Hilliard, C., & Cook, S. W. (2017). A technique for continuous measurement of body movement from video. *Behavior Research Methods*, 49, 1–12. <https://doi.org/10.3758/s13428-015-0685-x>
- Hilliard, C., O'Neal, E., Plumert, J., & Cook, S. W. (2015). Mothers modulate their gesture independently of their speech. *Cognition*, 140, 89–94. <https://doi.org/10.1016/j.cognition.2015.04.003>
- Hilveman, C., Cook, S. W., & Duff, M. C. (2016). Hippocampal declarative memory supports gesture production: Evidence from amnesia. *Cortex*, 85, 25–36. <https://doi.org/10.1016/j.cortex.2016.09.015>
- Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, 79–80, 1–17. <https://doi.org/10.1016/j.jml.2014.10.004>
- Hoetjes, M., Krahmer, E., & Swerts, M. (2015). On what happens in gesture when communication is unsuccessful. *Speech Communication*, 72, 160–175. <https://doi.org/10.1016/j.specom.2015.06.004>
- Hoetjes, M., & Masson-Carro, I. (2017). Under load: The effect of verbal and motoric cognitive load on gesture production. *Journal of Multimodal Communication Studies*, 4, 29–35. <https://doi.org/10.14746/jmcs>
- Hostetter, A., Boneff, K., & Alibali, M. (2018). Does extraneous perception of motion affect gesture production? In C. Kalish, M. Rau, J. Zhu, & T. T. Rogers (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 1821–1826). Austin, TX: Cognitive Science Society.
- Hostetter, A. B. (2008). *Mind in motion: The Gesture as Simulated Action framework* (Doctoral dissertation), University of Wisconsin, Madison, WI.
- Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137, 297–315. <https://doi.org/10.1037/a0022128>
- Hostetter, A. B. (2014). Action attenuates the effect of visibility on gesture rates. *Cognitive Science*, 38, 1468–1481. <https://doi.org/10.1111/cogs.12113>
- 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. <https://doi.org/10.1075/gest.7.1.05hos>
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15, 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language*, 63, 245–257. <https://doi.org/10.1016/j.jml.2010.04.003>
- Hostetter, A. B., & Alibali, M. W. (2011). Cognitive skills and gesture–speech redundancy: Formulation difficulty or communicative strategy? *Gesture*, 11, 40–60. <https://doi.org/10.1075/gest.11.1.03hos>
- Hostetter, A. B., Alibali, M. W., & Bartholomew, A. E. (2011). Gesture during mental rotation. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1448–1454). Austin, TX: Cognitive Science Society.
- Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my minds' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22, 313–336. <https://doi.org/10.1080/01690960600632812>
- Hostetter, A. B., Alibali, M. W., & Schrager, S. (2011). If you don't already know, I'm certainly not going to show you! Motivation to communicate affects gesture production. In G. Stam & M. Ishino (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 61–74). Philadelphia, PA: Benjamins.
- Hostetter, A. B., & Boncoddio, R. (2017). Gestures highlight perceptual–motor representations in thinking. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking, and communicating* (pp. 155–174). Philadelphia, PA: Benjamins.
- Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language*, 82, 22–29. [https://doi.org/10.1016/S0093-934X\(02\)00009-3](https://doi.org/10.1016/S0093-934X(02)00009-3)
- Hostetter, A. B., & Potthoff, A. L. (2012). Effects of personality and social situation on representational gesture production. *Gesture*, 12, 62–83. <https://doi.org/10.1075/gest.12.1.04hos>
- Hostetter, A. B., & Skirving, C. (2011). The effect of visual vs. verbal stimuli on gesture production. *Journal of Nonverbal Behavior*, 35, 205–223. <https://doi.org/10.1007/s10919-011-0109-2>
- Hostetter, A. B., & Sullivan, E. (2011). Gesture production during spatial tasks: Its not all about difficulty. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1965–1970). Austin, TX: Cognitive Science Society.
- Humphries, S., Holler, J., Crawford, T. J., Herrera, E., & Poliakoff, E. (2016). A third-person perspective on co-speech action gestures in Parkinson's disease. *Cortex*, 78, 44–54. <https://doi.org/10.1016/j.cortex.2016.02.009>
- Iani, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, 96, 110–121. <https://doi.org/10.1016/j.jml.2017.05.004>
- Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, 9, 23–43. [https://doi.org/10.1016/0885-2014\(94\)90018-3](https://doi.org/10.1016/0885-2014(94)90018-3)
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16, 367–371. <https://doi.org/10.1111/j.0956-7976.2005.01542.x>

- Iverson, J. M., Hall, A. J., Nickel, L., & Wozniak, R. H. (2007). The relationship between onset of reduplicated babble and laterality biases in infant rhythmic arm movements. *Brain and Language*, 101, 198–207. <https://doi.org/10.1016/j.bandl.2006.11.004>
- Iverson, J. M., & Thelen, E. (1999). Hand, mouth, and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6, 19–40. [Reprinted in R. Núñez & W. J. Freeman (Eds.). (2000). *Reclaiming cognition: The primacy of action, intention and emotion* (pp. 19–40). Thorverton, UK: Imprint Academic.]
- Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, 56, 291–303. <https://doi.org/10.1016/j.jml.2006.07.011>
- Kamermans, K. L., Pouw, W., Fassi, L., Aslanidou, A., Paas, F., & Hostetter, A. B. (2018). *The role of gesture as simulated action in reinterpretation of mental imagery*. Manuscript under review.
- Kelly, S., Byrne, K., & Holler, J. (2011). Raising the ante of communication: Evidence for enhanced gesture in high stakes situations. *Information*, 2, 579–593. <https://doi.org/10.3390/info2040579>
- Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, 22, 517–523. <https://doi.org/10.3758/s13423-014-0681-7>
- Kim, J. K., Stierwalt, J. A. G., LaPointe, L. L., & Bourgeois, M. S. (2015). The use of gesture following traumatic brain injury: A preliminary analysis. *Aphasiology*, 29, 665–684. <https://doi.org/10.1080/02687038.2014.976536>
- Kimbara, I. (2006). On gestural mimicry. *Gesture*, 6, 39–61. <https://doi.org/10.1075/gest.6.1.03kim>
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge, England: Cambridge University Press. <https://doi.org/10.1017/CBO9780511620850.011>
- Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, 24, 145–167. <https://doi.org/10.1080/01690960802586188>
- Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The Gesture-for-Conceptualization hypothesis. *Psychological Review*, 124, 245–266. <https://doi.org/10.1037/rev0000059>
- Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes*, 24, 761–775. <https://doi.org/10.1080/01690960802327971>
- Kita, S., & Essegbey, J. (2001). Pointing left in Ghana: How a taboo on the use of the left hand influences gestural practice. *Gesture*, 1, 73–94. <https://doi.org/10.1075/gest.1.1.06kit>
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48, 16–32. [https://doi.org/10.1016/S0749-596X\(02\)00505-3](https://doi.org/10.1016/S0749-596X(02)00505-3)
- Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, 22, 1212–1236. <https://doi.org/10.1080/01690960701461426>
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York, NY: Oxford University Press.
- Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7, 54–60. <https://doi.org/10.1111/1467-8721.ep13175642>
- Lacey, S., Campbell, C., & Sathian, K. (2007). Vision and touch: Multiple or multisensory representations of objects? *Perception*, 36, 1513–1521. <https://doi.org/10.1068/p5850>
- Laeng, B., & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, 26, 207–231. [https://doi.org/10.1016/S0364-0213\(01\)00065-9](https://doi.org/10.1016/S0364-0213(01)00065-9)
- Lakoff, G., & Núñez R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.
- Laurent, A., & Nicoladis, E. (2015). Gesture restriction affects French–English bilinguals’ speech only in French. *Bilingualism: Language and Cognition*, 18, 340–349. <https://doi.org/10.1017/S1366728914000042>
- Lavelli, M., & Majorano, M. (2016). Spontaneous gesture production and lexical abilities in children with specific language impairment in a naming task. *Journal of Speech, Language, and Hearing Research*, 59, 784–796. https://doi.org/10.1044/2016_JSLHR-L-14-0356
- Lawrence, B. M., Myerson, J., Oonk, H. M., & Abrams, R. A. (2001). The effects of eye and limb movements on working memory. *Memory*, 9, 433–444. <https://doi.org/10.1080/09658210143000047>
- Leman, M., & Maes, P.-J. (2014). The role of embodiment in the perception of music. *Empirical Musicology Review*, 9, 236–246. <https://doi.org/10.18061/emr.v9i3-4.4498>
- Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26, 1457–1471. <https://doi.org/10.1080/01690965.2010.500218>
- Lock, A., Young, A., Service, V., & Chandler, P. (1990). Some observations on the origin of the pointing gesture. In V. Volterra & C.J. Erting (Eds.), *From gesture to language in hearing and deaf children* (pp. 42–55). New York, NY: Springer.
- Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language–perception–cognition interface. *Current Directions in Psychological Science*, 24, 279–284. <https://doi.org/10.1177/0963721415570732>
- Mainela-Arnold, E., Alibali, M. W., Hostetter, A. B., & Evans, J. L. (2014). Gesture–speech integration in children with specific language impairment. *International Journal of Communicative Disorders*, 49, 761–770. <https://doi.org/10.1111/1460-6984.12115>
- Marstaller, L., & Burianova, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin & Review*, 20, 496–500. <https://doi.org/10.3758/s13423-012-0365-0>
- Masson, M. E. J., Bub, D. N., & Warren, C. M. (2008). Kicking calculators: Contribution of embodied representations to sentence comprehension. *Journal of Memory and Language*, 59, 256–265. <https://doi.org/10.1016/j.jml.2008.05.003>
- Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016a). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition, and Neuroscience*, 31, 430–440. <https://doi.org/10.1080/23273798.2015.1108448>
- Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016b). Imposing cognitive constraints on reference production: The interplay between speech and gesture during grounding. *Topics in Cognitive Science*, 8, 819–836. <https://doi.org/10.1111/tops.12217>
- Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2017). How what we see and what we know influence iconic gesture production. *Journal of Nonverbal Behavior*, 41, 367–394. <https://doi.org/10.1007/s10919-017-0261-4>
- Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture: A training study. *Developmental Science*, 15, 817–829. <https://doi.org/10.1111/j.1467-7687.2012.01181.x>
- Mayberry, R. I., Jacques, J., & DeDe, G. (1998). What stuttering reveals about the development of the gesture–speech relationship. In J. M. Iverson & S. Goldin-Meadow (Eds.), *The nature and function of gesture in children’s communication* (pp. 77–87). San Francisco, CA: Jossey-Bass.
- McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, 23, 45–66. <https://doi.org/10.1007/BF02143175>

- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, 4, 119–141. <https://doi.org/10.1075/gest.4.2.02mel>
- Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). Communicative gestures and memory load. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1569–1574). Austin, TX: Cognitive Science Society.
- Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language*, 66, 249–264. <https://doi.org/10.1016/j.jml.2011.07.004>
- Morett, L. M. (2018). In hand and in mind: Effects of gesture production and viewing on second language word learning. *Applied Psycholinguistics*, 39, 355–381. <https://doi.org/10.1017/S0142716417000388>
- Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology*, 117, 411–424.
- Müller, C. (1998). Iconicity and gesture. In S. Santi, I. Guatiella, C. Cave, & G. Konopczynski (Eds.), *Oralité et gestualité: Interactions et comportements multimodaux dans la communication. Actes du colloque [Orality and gestuality: Multimodal interaction and behavior in communication. Proceedings of the meeting of ORAGE 2001]* (pp. 407–410). Paris, France: L'Harmattan.
- Nathan, M. J., & Alibali, M. W. (2011). How gesture use enables inter-subjectivity in the classroom. In G. Stam & M. Ishino (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 257–266). Amsterdam, The Netherlands: Benjamins.
- Nathan, M. J., & Martinez, C. V. J. (2015). Gesture as model enactment: The role of gesture in mental model construction and inference making when learning from text. *Learning: Research and Practice*, 1, 4–37. <https://doi.org/10.1080/23735082.2015.1006758>
- Nemirovsky, R., Kelton, M. L., & Rhodamel, B. (2012). Gesture and imagination: On the constitution and uses of phantasms. *Gesture*, 12, 130–165. <https://doi.org/10.1075/gest.12.2.02nem>
- Nicoladis, E., Marentette, P., & Navarro, S. (2016). Gesture frequency linked primarily to story length in 4–10-year old children's stories. *Journal of Psycholinguistic Research*, 45, 189–210. <https://doi.org/10.1007/s10936-014-9342-2>
- Nicoladis, E., Pika, S., & Marentette, P. (2009). Do French–English bilingual children gesture more than monolingual children? *Journal of Psycholinguistic Research*, 38, 573–585. <https://doi.org/10.1007/s10936-009-9121-7>
- Niehorster, D. C., Siu, W. W. F., & Li, L. (2015). Manual tracking enhances smooth pursuit eye movements. *Journal of Vision*, 15, 11:1–14. <https://doi.org/10.1167/15.15.11>
- Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, 25, 903–910. <https://doi.org/10.1177/0956797613518351>
- Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin and Review*, 24, 652–665. <https://doi.org/10.3758/s13423-016-1145-z>
- O'Carroll, S., Nicoladis, E., & Smithson, L. (2015). The effect of extroversion on communication: Evidence from an interlocutor visibility manipulation. *Speech Communication*, 69, 1–8. <https://doi.org/10.1016/j.specom.2015.01.005>
- O'Shea, H., & Moran, A. (2017). Does motor simulation theory explain the cognitive mechanisms underlying motor imagery? A critical review. *Frontiers in Human Neuroscience*, 11, 72:1–13. <https://doi.org/10.3389/fnhum.2017.00072>
- Oliver, R. T., & Thompson-Schill, S. L. (2003). Dorsal stream activation during retrieval of object size and shape. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 309–322. <https://doi.org/10.3758/CABN.3.4.309>
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016a). Does language shape silent gesture? *Cognition*, 148, 10–18. <https://doi.org/10.1016/j.cognition.2015.12.001>
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016b). Is seeing gesture necessary to gesture like a native speaker? *Psychological Science*, 27, 737–747. <https://doi.org/10.1177/0956797616629931>
- Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effect of addressee location on representational gestures. *Journal of Memory and Language*, 46, 688–704. <https://doi.org/10.1006/jmla.2001.2826>
- Özyürek, A. (2017). Function and processing of gesture in the context of language. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 39–58). Philadelphia, PA: Benjamins.
- Özyürek, A., Kita, S., Allen, S. E., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures? Insights from crosslinguistic variations and similarities. *Gesture*, 5, 219–240. <https://doi.org/10.1075/gest.5.1.15ozy>
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45, 255–287. <https://doi.org/10.1037/h0084295>
- Parrill, F. (2010). Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25, 650–668. <https://doi.org/10.1080/01690960903424248>
- Parrill, F., Bergen, B. K., & Lichtenstein, P. V. (2013). Grammatical aspect in multimodal language production: Using gesture to reveal event representations. *Cognitive Linguistics*, 24, 135–158.
- Parrill, F., Bullen, J., & Hoburg, H. (2010). Effects of input modality on speech–gesture integration. *Journal of Pragmatics*, 42, 3130–3137. <https://doi.org/10.1016/j.pragma.2010.04.023>
- Parrill, F., Cabot, J., Kent, H., Chen, K., & Payneau, A. (2016). Do people gesture more when instructed to? *Gesture*, 15, 357–371. <https://doi.org/10.1075/gest.15.3.05par>
- Parrill, F., & Stec, K. (2018). Seeing first person changes gesture but saying first person does not. *Gesture*, 17, 159–176.
- Pearson, D. B., Ball, K., & Smith, D. T. (2014). Oculomotor preparation as a rehearsal mechanism in spatial working memory. *Cognition*, 132, 416–428. <https://doi.org/10.1016/j.cognition.2014.05.006>
- Pecher, D., Boot, I., & Van Dantzig, S. (2011). Abstract concepts: Sensory–motor grounding, metaphors, and beyond. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 217–248). San Diego, CA: Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-385527-5.00007-3>
- Perlman, M., Clark, N., & Johansson Falck, M. (2015). Iconic prosody in story reading. *Cognitive Science*, 39, 1348–1368. <https://doi.org/10.1111/cogs.12190>
- Perlman, M., & Gibbs, R. W., Jr. (2013). Pantomimic gestures reveal the sensorimotor imagery of a human-fostered gorilla. *Journal of Mental Imagery*, 37, 73–96.
- Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, 3, 359–400. [https://doi.org/10.1016/0885-2014\(88\)90021-4](https://doi.org/10.1016/0885-2014(88)90021-4)
- Pezzulo, G., Candidi, M., Dindo, H., & Barca, L. (2013). Action simulation in the human brain: Twelve questions. *New Ideas in Psychology*, 31, 270–290. <https://doi.org/10.1016/j.newideapsych.2013.01.004>
- Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences*, 20, 414–425. <https://doi.org/10.1016/j.tics.2016.03.013>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329–392. <https://doi.org/10.1017/S0140525X12001495>

- Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, C. W.-H., Cohen, L., ... Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences*, 101, 5658–5663. <https://doi.org/10.1073/pnas.0400707101>
- Pine, K. J., Gurney, D. J., & Fletcher, B. (2010). The semantic specificity hypothesis: When gestures do not depend upon the presence of a listener. *Journal of Nonverbal Behavior*, 34, 169–178. <https://doi.org/10.1007/s10919-010-0089-7>
- Ping, R. M., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, 34, 602–619. <https://doi.org/10.1111/j.1551-6709.2010.0112.x>
- Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, 143, 195–204. <https://doi.org/10.1037/a0032246>
- Popescu, S. T., & Wexler, M. (2012). Spontaneous body movements in spatial cognition. *Frontiers in Psychology*, 3, 136:1–13. <https://doi.org/10.3389/fpsyg.2012.00136>
- Pouw, W. T. J. L., & Hostetter, A. B. (2016). Gestures as predictive action. *Reti, Saperi, Linguaggi: Italian Journal of Cognitive Sciences*, 5, 57–80. <https://doi.org/10.12832/83918>
- Pouw, W. T. J. L., Mavilidi, M.-F., van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing*, 17, 269–277. <https://doi.org/10.1007/s10339-016-0757-6>
- Pouw, W. T. J. L., de Nooijer, J. A., van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, 5, 359:1–14. <https://doi.org/10.3389/fpsyg.2014.00359>
- Pouw, W. T. J. L., van Gog, T., Zwaan, R. A., Agostinho, S., & Paas, F. (2018). Co-thought gestures in children's mental problem solving: Prevalence and effects on subsequent performance. *Applied Cognitive Psychology*, 32, 66–80. <https://doi.org/10.1002/acp.3380>
- Pouw, W. T. J. L., Wassenburg, S., Hostetter, A. B., de Koning, B., & Paas, F. (2018). Does gesture strengthen sensorimotor knowledge of objects? The case of the size-weight illusion. Manuscript under review.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129–154. <https://doi.org/10.1080/713752551>
- Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*, 1, 110–122. <https://doi.org/10.1111/j.1745-6916.2006.00008.x>
- Provost, A., Johnson, B., Karayanidis, F., Brown, S., & Heathcote, A. (2013). Two routes to expertise in mental rotation. *Cognitive Science*, 37, 1321–1342. <https://doi.org/10.1111/cogs.12042>
- Quandt, L. C., Marshall, P. J., Shipley, T. F., Beilock, S. L., & Goldin-Meadow, S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: An EEG study of action production and gesture observation. *Neuropsychologia*, 50, 2745–2751. <https://doi.org/10.1016/j.neuropsychologia.2012.08.005>
- Reig Alamillo, A., Colletta, J. M., & Guidetti, M. (2012). Gesture and language in narratives and explanations: The effects of age and communicative activity on late multimodal discourse development. *Journal of Child Language*, 40, 511–538. <https://doi.org/10.1017/S0305000912000062>
- Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotated text: A natural behavior approach to investigating cognitive offloading. *Cognitive Science*, 38, 537–564. <https://doi.org/10.1111/cogs.12087>
- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research*, 71, 491–507. <https://doi.org/10.1007/BF00248742>
- Sassenberg, U., Foth, M., Wartenburger, I., & van der Meer, E. (2011). Show your hands—Are you really clever? Reasoning, gesture production, and intelligence. *Linguistics*, 49, 105–134. <https://doi.org/10.1515/LING.2011.003>
- Sassenberg, U., & van der Meer, E. (2010). Do we really gesture more when it is more difficult? *Cognitive Science*, 34, 643–664. <https://doi.org/10.1111/j.1551-6709.2010.01101.x>
- Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, 20, 457–497. https://doi.org/10.1207/s15516709cog2004_1
- Skoura, X., Vinter, A., & Papaxanthis, C. (2009). Mentally simulated motor actions in children. *Developmental Neuropsychology*, 34, 356–367. <https://doi.org/10.1080/87565640902801874>
- Slepian, M. L., & Ambady, N. (2012). Fluid movement and creativity. *Journal of Experimental Psychology: General*, 141, 625–629. <https://doi.org/10.1037/a0027395>
- Smithson, L., & Nicoladis, E. (2014). Lending a hand to imagery? The impact of visuospatial working memory interference upon iconic gesture production in a narrative task. *Journal of Nonverbal Behavior*, 38, 247–258. <https://doi.org/10.1007/s10919-014-0176-2>
- So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese–English bilinguals. *Language and Cognitive Processes*, 25, 1335–1353. <https://doi.org/10.1080/01690961003694268>
- So, W. C., Ching, T. H.-W., Lim, P. E., Cheng, X., & Ip, K. Y. (2014). Producing gestures facilitates route learning. *PLoS ONE*, 9, e112543:1–21. <https://doi.org/10.1371/journal.pone.0112543>
- So, W. C., Shum, P. L.-C., & Wong, M. K.-Y. (2015). Gesture is more effective than spatial language in encoding spatial information. *Quarterly Journal of Experimental Psychology*, 68, 2384–2401. <https://doi.org/10.1080/17470218.2015.1015431>
- Spruijt, S., van der Kamp, J., & Steenbergen, B. (2015). Current insights in the development of children's motor imagery ability. *Frontiers in Psychology*, 6, 787:1–12. <https://doi.org/10.3389/fpsyg.2015.00787>
- Stieff, M., Lira, M. E., & Scopelitis, S. A. (2016). Gesture supports spatial thinking in STEM. *Cognition and Instruction*, 34, 80–99. <https://doi.org/10.1080/07370008.2016.1145122>
- Talmy, L. (1983). How language structures space. In H. Pick & L. Acredolo (Eds.), *Spatial orientations: Theory, research, and application* (pp. 225–282). New York, NY: Plenum Press.
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78, 705–722. <https://doi.org/10.1111/j.1467-8624.2007.01025.x>
- Topolinski, S., & Strack, F. (2009). Motormouth: Mere exposure depends on stimulus-specific motor simulations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 423–433. <https://doi.org/10.1037/a0014504>
- Trofatter, C., Kontra, C., Beilock, S., & Goldin-Meadow, S. (2015). Gesturing has a larger impact on problem-solving than action, even when action is accompanied by words. *Language and Cognitive Neuroscience*, 30, 251–260. <https://doi.org/10.1080/23273798.2014.905692>
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychology Review*, 117, 440–463. <https://doi.org/10.1037/a0018963>
- Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, 116, 185–203. <https://doi.org/10.1016/j.actpsy.2004.01.004>
- Tuite, K. (1993). The production of gestures. *Semiotica*, 93, 83–106. <https://doi.org/10.1515/semi.1993.93.1-2.83>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209–232. <https://doi.org/10.1016/j.specom.2013.09.008>

- Wagner, S., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is hand-waving spatial? *Journal of Memory and Language*, 50, 395–407. <https://doi.org/10.1016/j.jml.2004.01.002>
- Wartenburger, I., Kuhn, E., Sassenburg, U., Foth, M., Franz, E. A., & van der Meer, E. (2010). On the relationship between fluid intelligence, gesture production, and brain structure. *Intelligence*, 38, 193–201. <https://doi.org/10.1016/j.intell.2009.11.001>
- Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2009). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience*, 22, 2387–2400. <https://doi.org/10.1162/jocn.2009.21386>
- Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life-span. *Developmental Psychology*, 35, 205–213.
- Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science*, 20, 201–206. <https://doi.org/10.1177/0963721411408770>
- Yannier, N., Hudson, S. E., Wiese, E. S., & Koedinger, K. R. (2016). Adding physical objects to an interactive game improves learning and enjoyment: Evidence from EarthShake. *ACM Transactions on Computer–Human Interaction*, 23(4), 26:2–31. <https://doi.org/10.1145/2934668>
- Yap, D. F., Brookshire, G., & Casasanto, D. (2018). Beat gestures encode spatial semantics. In C. Kalish, M. Rau, J. Zhu, & T. T. Rogers (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (p. 1211). Austin, TX: Cognitive Science Society.
- Yeo, A., & Alibali, M. W. (2018). Does visual salience of action affect gesture production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 826–832. <https://doi.org/10.1037/xlm0000458>
- Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, 18, 229–234. <https://doi.org/10.1016/j.tics.2014.02.008>