Gesture as Predictive Action

Article · April 2016

2 authors:

Wim T. J. L. Pouw
Erasmus University Rotterdam
12 PUBLICATIONS **86** CITATIONS
SEE PROFILE

Autumn B. Hostetter
Kalamazoo College
28 PUBLICATIONS **1,122** CITATIONS
SEE PROFILE

Gesture as Predictive Action

Wim Pouw[1] and Autumn B. Hostetter[2]

1. Department of Psychology, Education and Child Studies, Erasmus University Rotterdam
2. Department of Psychology, Kalamazoo College

Author Note. Correspondence should be addressed to Autumn B. Hostetter, Department of Psychology, 1200 Academy Street, Kalamazoo, MI 49006. Email: Autumn.Hostetter@kzoo.edu.

Acknowledgement. The first author's research is funded by the Netherlands Organisation for Scientific Research (NWO-PROO, project number: 411-10-908).

Abstract

Two broad approaches have dominated the literature on the production of speech-accompanying gestures. On the one hand, there are approaches that aim to explain the origin of gestures by specifying the mental processes that give rise to them. On the other, there are approaches that aim to explain the cognitive function that gestures have for the gesturer or the listener. In the present paper we aim to reconcile both approaches in one single perspective that is informed by a recent sea change in cognitive science, namely, Predictive Processing Perspectives (PPP; Clark, 2013b, 2015). We start with the idea put forth by the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). Under this view, the mental processes that give rise to gesture are re-enactments of sensori-motor experiences (i.e., simulated actions). We show that such anticipatory sensori-motor states and the constraints put forth by the GSA framework can be understood as top-down kinesthetic predictions that function in a broader predictive machinery as proposed by PPP. By establishing this alignment, we aim to show how gestures come to fulfill a genuine cognitive function above and beyond the mental processes that give rise to gesture.

Key words: Gesture & Cognition, Gesture-as-simulated action, Predictive Processing, Problem Solving, Learning

When speakers talk, they often move their hands and arms in a way that mirrors or complements the semantic content of what they are saying. These movements, hereafter referred to simply as gestures, are, in some sense, the epitome of "embodiment" because they are movements of the *body* that are produced in the interest of communication. Yet, to say that gestures are embodied just because they make use of the body is unsatisfactory from the standpoint of Cognitive Science, which uses the term *Embodied Cognition* to mean that cognitive processes make use of perceptual and motor systems, even in situations where such systems would seem to be irrelevant (e.g., Wilson, 2002). To say that gestures are truly embodied then, requires the specification of how, and in virtue of which unique properties (e.g., visual, proprioceptive stimulation), gestures affect cognition (Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). That is, we must move beyond descriptive accounts of what gestures are and towards understanding why they are produced and how they come to have facilitative effects on cognition.

So, why are gestures produced? This question has received increasing attention over the past two decades. While space does not allow a detailed description of the many possibilities, a review of the literature reveals two broad types of answers to this question. First, there are answers that are about *origin*. That is, are the processes or representations that underlie gesture unique to gesture or are they similar to those that are involved in speaking or action generation more generally? Second, there are answers that are about *cognitive function*. Once a gesture is produced, what effect does it have for the speaker and for the listener, and how is this effect brought about?

In this paper, we consider both the origin and function of gesture in a single account. We first describe one theory about the origin of speech-accompanying gestures, namely the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). Under this view, gestures arise from simulated perceptual and action states that are created as a speaker talks about a present or imagined situation. Inspired by contemporaneous ideas about the embodied nature of language comprehension (Barsalou, 2008; Glenberg & Kaschak, 2002), the goal of the GSA framework was to explain how gestures might originate in an embodied cognitive system that is engaged during speech production. However, in the years since the GSA framework was published, a number of theories have taken hold in Cognitive Science that we believe are compatible with, and nicely complement, ideas presented in the GSA framework, namely, Predictive Processing Perspectives.

Predictive Processing Perspectives (hereon PPP; e.g., Clark, 2013b, 2015a, b; Den Ouden, Kok, & De Lange, 2012; Friston, 2010; Glenberg & Gallese, 2012; Hohwy, 2013; Lupyan & Clark, 2015), broadly characterized, postulate that the central work of cognitive systems is to engage in predictions. That is, predictions about sensori-motor consequences that emerge during interaction with the environment, and are updated and acted upon in a way that minimizes surprisal or *prediction error* (i.e., the residual discrepancy between what is predicted and what is encountered). We think PPP are promising for furthering our understanding of how gestures emerge from embodied simulations, while also securing a central role for bodily action in these processes. In this paper, we aim to explore how gestures may be thought about as one instantiation of predictive processing; that is, gestures have functions for speakers and thinkers that can be broadly construed as minimizing prediction error, and these functions can be captured by one broad underlying mechanism giving rise to gestures, namely the activation of simulations (i.e., predictions) in the motor and perceptual system. Thus, by considering gestures as a case of predictive processing, we aim to move beyond considering the origin of gesture and its function as separate explanatory quests, and provide a way to understand how origin and function of gesture are intricately related.

1.1 Outline

Next, (section 2) we provide an overview of the GSA framework's key tenets and the evidence to date. In section 3, we address the cognitive function of gesture. Finally, we introduce PPP, as a means of understanding how gestures not only arise from sensori-motor predictions (i.e., simulations) in the cognitive system (section 4), but also support these predictive processes during on-going cognitive activity (section 5).

2. The Gesture as Simulated Action Framework

The Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008) considers gesture production to be the outgrowth of a cognitive system that is actively engaged in simulating motor and perceptual states. Simulations are neural enactments or re-enactments of interactions with the world; when a speaker engages in simulation, the same motor and perceptual areas of the brain are recruited that would be involved in actually performing the action or viewing the scene. This neural activity in the motor and action systems of the brain has the potential to be expressed alongside speech as gesture. The GSA framework proposes three determinants of whether a simulation is actually expressed alongside speech in any particular instance.

First, the production of a gesture depends on how strongly the simulation evokes thoughts of action. Simulations that are closely tied to action are more likely to engage the motor system strongly enough to result in a movement being produced (e.g., gesture) than simulations with weaker ties to action. For example, a speaker who has experience actually making a pattern he is describing is more likely to gesture about the pattern than he is about a pattern he has only viewed (Hostetter & Alibali, 2010). Action can also be evoked in a simulation of a perceptual scene that was not directly acted on. For example, if speakers can easily imagine interacting with what they are describing, they are particularly likely to gesture about it. Chu and Kita (2015) found that speakers gestured less about a mug that had spikes along its handle than they did about a mug with no spikes that more readily afforded grasping. Masson-Carro, Goudbeek, and Krahmer (2015) further show that the affordances of objects directly predict whether the objects are gestured about. Simulating perceptual scenes may also evoke action if the speaker imagines the scene or its objects in motion. For example, speakers frequently gesture when engaged in mental rotation exercises (e.g., Chu & Kita, 2008, 2011) and gesture more when talking about the process of rotating than when describing the end state of the rotation (e.g., Hostetter, Alibali, & Bartholomew, 2011). Finally, even thinking about a static perceptual experience may engage the motor system because of the tight coupling between perception and action. When we perceive an object, we automatically activate processes about how we would use, grasp, or interact with the object (e.g., Tucker & Ellis, 1998). Under the GSA framework, such activation can be expressed as gestures that depict how to interact with the object, or outline the object's shape.

Second, the production of a gesture depends not only on the absolute strength of action activation involved in the simulation, but also whether this activation is strong enough to pass the speaker's current *gesture threshold*. The gesture threshold is conceptualized as the speaker's current resistance to producing a gesture, but can change from moment to moment during speaking. For example, in situations where speakers think a gesture might benefit their listener, they may lower their threshold and gesture more (e.g., Alibali, Heath, & Myers, 2001). Similarly, if speakers consider the information they are conveying to be particularly important to their listener, they gesture more (e.g., Kelly, Byrne, & Holler, 2011), perhaps as the result of maintaining a lower threshold that even weaker action simulations can surpass. Moreover, speakers may adjust their threshold (either consciously or unconsciously) based on the cognitive demands of the speaking situation. Because gestures are known to have a number of beneficial effects (e.g., Goldin-Meadow & Alibali, 2015), a speaker may find it advantageous to lower her threshold to allow even a weak action activation to be expressed as gesture in certain situations. Conversely, even in situations where there is no clear reason to lower one's threshold, simulations that evoke strong activation of action may result in gesturing regardless because the simulation is strong enough to pass even a heightened threshold (see Hostetter, 2014 for some evidence on this point).

Finally, the GSA framework contends that the occurrence of gesture is particularly likely in situations where the articulatory motor system is already activated in the interest of speaking. Because the speaker must engage his or her motor system for speaking, it is difficult to simultaneously inhibit the manual motor system from also expressing the action activation that occurs during simulation. This is corroborated by findings that show that hand and mouth actions are linked from infancy (Iverson & Thelen, 2000) and heavily constrain one-another throughout further adulthood (Gentilucci, Benuzzi, Gangitano, & Grimaldi, 2001). However, while the GSA framework contends that gestures are more likely to occur with speech than in its absence, the framework by no means precludes the occurrence of gestures without speech. Indeed, since the publication of the GSA framework, a number of reports have been published about gestures that occur in the absence of speech (e.g., Chu & Kita, 2011; Delgado, Gómez, & Sarría, 2011). Such co-thought gestures seem to share many characteristics with co-speech gestures (Chu & Kita, 2015). This evidence supports for the idea that gestures are not dependent on language; while they frequently occur with language, the processes that give rise to gesture are more generally rooted in the sensori-motor, rather than linguistic, system.

3 The cognitive function of gesture

The GSA framework was developed to account for how gestures arise from an embodied cognitive system (Hostetter & Alibali, 2008, p. 495). While not reducing gestures to an epiphenomenon, the issue of how gestures *function* in such a system was left open to further speculation. Consequently, the GSA framework is flexible regarding the possible functions of gesture. Indeed, once a gesture is produced, the GSA framework allows that the movement may have any number of cognitive effects. However, in order to offer a truly embodied account of gesture that considers both their origin and their function, a more detailed specification of how gestures perform their cognitive functions is needed.

What does an embodied account of gesture function entail? Pouw and colleagues (2014; see also Pouw, Van Gog, Zwaan, & Paas, in press) argue that to truly explain the cognitive function of *gesture*, a theory must be able to explicate how this bodily act affects the cognitive system above and beyond neural processes that precede gesturing. That is, it must become clear how the act of gesturing directly affects cognition, which is not accomplished when positing some neural process that generates the gesture as well as its cognitive effect. For example, consider a learner who is attempting to memorize the steps needed to complete a route. The learner may mentally visualize the steps required, and this visualization may lead the learner to gesture about each step and may also lead to improved memory for the steps. However, in order to consider gesture as a causal agent that led to improved memory, it must become clear what additional benefit gesturing brings above and beyond the mental visualization that gives rise to the gesture in the first place.

The idea that the act of gesture might add something to the cognitive toolkit is not new. As the philosopher Andy Clark (2013b) has recently described:

> In gesture, as when we write or talk, we materialize our own thoughts. We bring something concrete into being, and that thing (in this case, the arm motion) can systematically affect our own ongoing thinking and reasoning… … [as such] gesture and overt and covert speech emerge as interacting parts of a distributed cognitive engine, participating in cognitively potent self-stimulating loops whose activity is as much an aspect of our thinking as its result. (Clark, 2013b, p. 263).

In this way of thinking, cognition is not completely brain-bound; rather the physical activity of gesture – in virtue of its co-constitutive role in ongoing cognition - is itself a genuine form of cognition (cf. Clark & Chalmers, 1998). Similarly, McNeill (2005) has argued that gesture and speech exist together in a dialectic, with each influencing and affecting the other. In sum, gestures are not just the result of cognition, they are a critical determinant of cognition (e.g., Goldin-Meadow & Beilock, 2010).

Indeed, there is much research suggesting that gestures affect cognition in a variety of ways (see Goldin-Meadow & Alibali, 2015 for a recent review). For example, speakers who gesture have better memory for what they gesture about than speakers who do not gesture (Cook, Yip, & Goldin-Meadow, 2010). In addition to strengthening the representation being described, gestures also appear to reduce general working memory demands, such that there are more cognitive resources available to devote to a secondary task when speakers gesture than when they do not (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). Gestures appear to affect how speakers solve spatial problems, by influencing the strategy choice (e.g., Alibali, Spencer, Knox, & Kita, 2011) or by focusing attention on perceptual elements of the problem (e.g., Beilock & Goldin-Meadow, 2010). Given these effects on cognitive processing, it is perhaps not surprising that gestures also help speakers communicate, particularly about concepts that are highly spatial or motoric (e.g., Hostetter, 2011). There is some evidence that gestures may actually prime relevant words or ideas in the lexicon (e.g., Krauss, 1998), and that they may help speakers conceptualize what they want to say and package the ideas into the linear stream of speech (e.g., Alibali, Yeo, Hostetter, & Kita, under review). In sum, the cognitive functions of gesture are varied, and have been shown in a variety of domains ranging from problem solving to memory to language.

We believe that these varied functions can be explained under a general mechanism suggested by Predictive Processing Perspectives (PPP). Not only are PPP highly compatible with the GSA framework and the suggestion that gestures arise out of embodied neural simulations, but they also provide further explanation for how gestures' function is not reducible to these neural simulations. Rather, action (and thus gesture) is central to the job description of the cognitive system assigned by PPP, namely, prediction optimization.

4. *Predictive Processing Perspectives*

PPP are a recent sea change in cognitive science (for broad overviews see Clark, 2013a, 2015b; Hohwy, 2013). As PPP are rapidly adapting, they are becoming more divergent from each other (see e.g., Pickering & Clark, 2014). Yet, what unites these models is that they assign a single job description to the cognitive system under which most, if not all, cognitive feats (e.g., perception, action, social cognition, language production and comprehension) can be subsumed. Namely, the cognitive system is engaged in optimizing predictions about the continuous flow of sensory data that perturb the system during the ongoing flux of (potentially hazardous) interactions with the environment. Minimizing prediction error is not some abstract project, but key to the perseverance of life: simply put, "*avoid surprises and you will last longer*" (original emphasis, Friston, Thornton, & Clark, 2012, p. 2).

We will argue that gesture is one special way to optimize predictions. We do this by showing that action-oriented models in PPP (Clark, 2015a, b; Friston, 2009) are compatible with the GSA framework, and are able to clarify the mechanism by which gestures benefit cognition. Two important aspects of PPP will be considered. First, PPP put sensori-motor neural simulations in a broader context of a hierarchical predictive architecture. Second, PPP assign a pivotal role of action within this broader predictive machinery. Thus, by considering gesture as a special case of action, we can use PPP to understand both the cognitive origin and function of gesture.

Before introducing basic tenets of PPP, some preliminary remarks are in place. First, we only provide a broad conceptual overview of some key mechanisms of PPP (e.g., Clark, 2015b) that we think are relevant to thinking about gesture, neglecting statistical formalisms (e.g., Bayes Theorem) that ground PPP (see Friston, 2009, 2010; but see Hohwy, 2013 for an approachable introduction). Second, although it is largely undisputed that there is *a* predictive component in many central cognitive processes, such as attention (e.g., Hohwy, 2013), vision (O'Regan & Noë, 2001), action (e.g., Franklin & Wolpert, 2011), and language comprehension and production (e.g., Lupyan & Clark, 2013; Pickering & Garrod, 2013), models within PPP are still highly debated, and at present there is no evidence to decisively choose among competing models. Thus, our account is an attempt to show the preliminary utility of PPP for understanding gesture's production and function, rather than an endorsement of any one view.

4.1 Introduction to PPP

Predictive Processing Perspectives (PPP), as presented by Clark (2013a, b; 2015a, b; based on Friston, 2009, 2010), entail that the cognitive system has, and continuously adapts, a body of knowledge (called a 'hierarchical generative model') that allows the agent to self-generate data (called 'predictions' or 'prior expectations') about the world that capture the statistical regularities of incoming sensory data. These predictions mostly run on automatic pilot, and need not be subject of awareness to do their work (although they may be constrained by conscious processing). Importantly, a generative model is never perfect, and its predictions never completely match the incoming sensory input. In fact, these discrepancies between incoming sensory input and the predictions are informative and continuously monitored. These discrepancies, called 'prediction-error', are used to update the generative model in order to issue more precise predictions in the future. Thus, prediction errors are used to calibrate future predictions, and over time enable the generative model to make better predictions about the world.

The generative model is "hierarchical" because predictions are issued on multiple higher and lower order levels. Lower order levels issue fast-changing *sensory* predictions, likely to operate on timescales ranging from hundreds of milliseconds to seconds. For example, when reaching for a mug, lower order haptic predictions are produced about the instant consequences of picking up the mug. Higher order levels are more likely to be abstract and multi-modal, and operate on longer time-scales. That is, these levels are not concerned with one particular sensory consequence, but with complex multi-modal regularities that emerge over longer periods of time. Slower predictions might concern keeping track of a trajectory of a moving object, or for slower predictions still, how particular types of situations generally unfold (e.g., restaurant visits, idle conversations etc.). In each level, the model predicts the output from the level below and compares this predicted output to the actual input received from the level below. This results in a complex multi-layered predictive machinery that works in concert to track relevant small- and large-scale changes that have proved to be relevant to the agent in the past.

In many ways, PPP are a reversal of classical models of perception - wherein the cognitive system passively receives input in a bottom-up fashion. Rather, in PPP, the mind has a more active, anticipatory and self-adaptive role[1]. In PPP, action is an essential part of perception, as active sampling can make perceptual patterns that are predicted come true and allow the agent to make better predictions. For example, I may not know for sure that my coffee cup is empty, but when I grasp the handle without enough muscle tone to account for its filled weight, I sample unexpected proprioceptive feedback (i.e., prediction-error is produced) that informs me to adjust my prediction about the cup's fullness, as well as what action is appropriate. To continue with this example, in some versions of PPP (Clark, 2015b; Friston, 2009; 2010) the proprioceptive consequences of a *full* coffee-mug result in prediction error that activates the motor system to adjust to a grasp that optimally deals with a full cup instead of the present grasp (empty cup). In fact, in such versions of PPP, *all actions* are produced by the motor system to resolve prediction errors by making predictions about the consequences of actions true by actually performing those actions. Simply put, when an action is predicted in a particular context, the consequences of those predicted actions are compared to the present state of the system. This results in prediction-errors that are resolved by acting on those predictions.

Minimizing prediction error with action is called active inference. Active inference reduces prediction error using what Pickering and Clark (2014, p. 451) refer to as "twin strategies." Predictions are altered to fit the environment and the environment and body

[1] Interestingly, computer vision research has yielded productive results by implementing just such an active model of vision (e.g., Rao & Ballard, 1999), and this model has unique explanatory power with regards to persistent optical illusions experienced by humans (for an overview see Hohwy, 2013).

are altered through action to fit the predictions. Further, action can simplify what we might call the "predictive load" of the generative model (see Clark, 2015a, b). Namely, actions are informative for updating predictions, as active sampling provides information otherwise not (as reliably) available in a passive mode. As Clark (2015a, p. 15) says, "the course of embodied action to novel patterns of sensory stimulation, may thus acquire forms of knowledge that were genuinely out-of-reach prior to such physical-manipulation-based re-tuning of the generative model." As an illustrative instance, Clark (2015a) points to abacus-training, wherein children are able to learn to perform complex arithmetic by using an abacus, and learn to perform these calculations without an abacus after sufficient training (Stigler, 1984). Learning by acting on an abacus allows the generative model to shape predictions with more reliable inputs (e.g., the results of the actions themselves), and effectively reduces the degrees of complexity of the generative model itself (see Kirsh & Maglio, 1994 for a similar example).

How is an agent able to flexibly employ the different strategies for prediction error-minimization? For example, in some cases active sampling is not an option, and inference on the basis of present input is more appropriate. PPP employ precision-estimation as a mechanism that allows for the flexibility of predictive strategies. Namely, every prediction and sensory input is given a certain second-order 'weighting' (called a 'precision estimate') on the bases of its predicted accuracy. That is, given the context (e.g., say a misty day, or a dark room), the cognitive system may treat incoming sensory signals as less reliable (i.e., lower precision estimate), which results in relatively higher precision estimates of top-down predictions. This allows the agent to behave according to prior knowledge (e.g., anticipating stop signs on the misty road you are driving on; navigating the dark room based on memory) as a more reliable way to reduce prediction error than relying only on sensory bottom-up information. In contrast, in a completely novel situation, it may be difficult to form top-down predictions with any amount of accuracy. In such situations, action becomes increasingly important as a means of learning the environment. Thus, precision estimates allow the system more flexibility, as in some situations the environment can be used as its own best model, whereas in others, a top-down model for the environment may be more effective (Clark, 2015a, b). Precision estimates allow the system to determine which is best.

4.2 Casting the GSA framework in terms of PPP

We believe that there is synergy between the key concepts of PPP and those of the GSA framework. In the sections that follow, we will explore how each of the three determinants of whether a gesture is produced as proposed in the GSA framework can be explained by PPP. Further, by considering gestures as action in a PPP, some predictions about gesture function naturally emerge.

4.2.1 Action simulations are strongly activated when prediction error is high

The GSA framework holds that gestures arise from action-simulations, wherein the strength of motor activations predicts (in part) the likelihood of overt gestures. The strength of activation is determined in large part by the manual motor-affordances that are solicited by the environment or content of speech (e.g., Chu & Kita, 2015; Masson-Carro et al., 2015).

In PPP, action is produced as a means of resolving the prediction error that exists when an action is predicted but one's body state is different (e.g., static). In order to think about an event that involves action, speakers' cognitive system must predict what actions are involved and what the proprioceptive and visual consequences of those actions would be. Creating such predictions in the absence of overt movement results in high prediction error, as the cognitive system predicts that movement should be occurring but does not receive the kinematic feedback of such movement. To resolve this prediction error, the speaker's motor system may be activated to produce congruent movement. Such movement is recognized as gesture when it occurs alongside speech.

Thus the claim that gestures occur when action simulations are strongly activated in the mind of a speaker is compatible with the basic claim of PPP. For example, PPP can accommodate the idea that when the relevant content of speech is actional (e.g., throwing a ball) gestures are more likely, than when the content of speech is about visual-spatial (e.g., seeing a house) or abstract concepts (e.g., democracy), as action is part of the prediction formed by the cognitive system in the former case. To think (and talk) about throwing a ball without actually producing the corresponding action requires the cognitive system to tolerate a higher amount of prediction error in the motor-system. Under PPP, such a state is not desirable; thus, an action is likely to be produced as a means of resolving the prediction error.

Consider the case of mental rotation, in which participants are asked to imagine the visual consequence of rotating some object a specified amount around its axis. In such a task, the visual information associated with rotation of the object must be predicted top-down by a generative model that captures sensory consequences that co-uccur with such rotations based on previous experience. This requires spatiotemporally fine-grained visual predictions of a moving object. In the terminology of PPP, prediction error in such a situation is high, as the top-down predictions of the object's visual appearance as well as motor-associations following rotation do not match the sensory input of the objects' given starting position. To resolve the error, an action may be initiated, even one which does not actually manipulate the object. Indeed, during mental rotation, either co-occuring with or without speech, participants naturally adopt gestures as-if manipulating the object to be rotated (Chu & Kita, 2008; 2011). In terms of the GSA framework, such gestures occur because the movement involved in rotation is being simulated strongly in the speaker's mind, in order to determine its endstate (see Hostetter, Alibali, & Bartholomew, 2011).

Of course, not all gestures are direct pantomimes of action. Many gestures take a form of outlining or tracing a described object. For example, a speaker might say "it was round" while tracing a circle shape in the air. In such an instance, it is difficult to see how the gesture could be reducing prediction error between a predicted action and the kinematic absence of such action. However, in PPP, predictions are multimodal, meaning that they are not limited to action predictions but can also involve visual predictions. Thinking (and talking) about a ball does not only involve predicting what corresponding actions go along with a ball, but also what the ball looks like. Creating an image of the ball with one's hands could be a way to minimize the prediction error that is inherent to talking about how an object looks without getting sensory input about the object's actual appearance. Indeed, speakers gesture less about objects that are visually present than about objects that are not present (e.g., Morsella & Krauss, 2004). This could be because there is less prediction error involved in talking about an object that is visually present in the environment, so action is not as likely to be initiated. In contrast, when there is no visual object present, gestures make the inferences about the object made in speech become true. Under this view, proprioceptive predictions, that are first inherent to action processing, become multi-modally associated with depictive visual-spatial processing (e.g., shape of a house) over development. We speculate that the proprioceptive feedback of an action or gesture comes to activate relevant visual-spatial details as well (see Cooperrider, Wakefield, & Goldin-Meadow, 2015; Pouw, Mavilidi, Van Gog, Zwaan, & Paas, under review, for some evidence on this point).

In sum, the GSA framework proposes that gestures are automatically activated as the result of activation in the sensori-motor system during speaking and thinking. This is congenial to the idea that top-down predictions about sensori-motor events are continuously employed by the cognitive system. Furthermore, that gestures are most likely to be produced when there is a disconnect between the physical and mental environment (e.g., when action is being talked about or when a visual scene is being described that is not visually present) suggests that gestures may emerge precisely when prediction error related to the motor-system is high. In the terms of the GSA framework, action simulations become strongly activated in such situations, and this high activation leads to gesture.

4.2.2 Gesture threshold is adjusted based on precision estimates

Recall that the GSA framework argues that simulations underlying gestures are automatically activated, but that their overt production as a gesture is dependent on a number of contextual factors captured by the 'gesture-threshold'. Speakers can adjust their gesture rate (either consciously or unconsciously) as the result of such things as believing that a gesture will be helpful to either themselves or their listener.

This is similar to the way that predictions and sensory inputs are given precision estimates in PPP, so that the cognitive system can rely more on one or the other in a particular situation. When sensory input is degraded, the cognitive system may favor top-down predictions. When top-down predictions seem insufficient, the cognitive system will seek out sensory input to provide new information through active inference. For example, Tetris players often rotate blocks to decide where to best place them (Kirsh & Maglio, 1994). Producing the rotation movement provides more reliable bottom-up information than top-down predictions (i.e., mental rotations) of where the pieces will fit best.

As mentioned above, a similar process has been observed with the use of gesture during mental rotation tasks (Chu & Kita, 2008, 2011). Most important for our discussion of prediction estimates and the gesture threshold, however, is the finding that participants do not always gesture during such tasks. In cases where the rotational angle is smaller, those participants who generally gesture in more difficult trials may not adopt gestures. At smaller rotation angles, the task is easier, and as such, the precision estimate of top-down visual and motor predictions of the rotation is set to be more reliable (given previous successes in the past) and thus active inference (gesture) is less likely to occur.

This could also explain the findings that those with a lower (as opposed to higher) working memory capacity are more likely to gesture during speech production (e.g., Chu, Meyer, Foulkes, & Kita, 2013; Gillespie, James, Federmeier, & Watson, 2014). For participants with limited working memory systems, top-down predictions are generally more unreliable, leading them to adopt gestures as a means of providing more accurate sensori-motor predictions. In the terms of the GSA framework, such speakers intuit the potential benefit of gesture and thereby set a low gesture threshold so that many of their simulations come to be expressed in gesture. Indeed, Dunn and Risko (2015) found that metacognitive judgments of whether an external rather than internal strategy is more efficient directly predicts how problem solvers approach a task. Thus, precision estimates and reliability judgments may determine whether gestures are produced.

The idea in the GSA framework that speakers can intuit whether a gesture is helpful or not is compatible with the idea in PPP that the cognitive system employs precision estimates as a way to give preference to sensory inputs or top-down predictions. In situations where producing a gesture could help the system visualize the details of the top-down prediction, a gesture is more likely to be produced.

4.2.3 Simultaneous speaking prevents complete inhibition of motor system

In the GSA framework, the final predictor of gesture is whether speech is accompanying the simulation. The GSA framework proposes that because the vocal articulators must be moved during speaking, it is difficult to completely inhibit the motor activity involved in simulation from being expressed as gesture. Although gestures can and do occur without speech (e.g., Chu & Kita, 2015), gestures are typically more prevalent alongside speech than in its absence.

This explanation is in line with the mechanics of PPP. Recall that sensori-motor predictions can be inhibited in situations where top-down predictions are estimated to be more accurate. This is sometimes referred to as "gain control", or the system's ability to gate sensori-motor predictions so that only weak signals are sent to the muscles (e.g., Grush, 2004). However, when the motor system must be involved in the interest of producing speech, it is difficult to completely inhibit all motor signals from being sent to the muscles. In their Action-Based Language theory, Glenberg and Gallese (2012) offer a PPP on language, positing that language learning, comprehension, and production capitalize on systems for motor control. They follow the GSA framework in proposing that gestures are the result of activating relevant actions alongside speech paired with an inability to completely block these movements from being expressed because speaking requires movement of the mouth and vocal articulators. Thus at least one PPP has already offered an account of gesture fully in line with that provided in the GSA framework.

As evidence for this account, consider that the articulatory/oral system and manual system are closely entrained. For example, humans often open and close their mouths during skillful manual manipulation (Darwin, 1998). It has been found that when grasping an object with one's mouth, the size of the mouth opening correspondingly affects the size of index-thumb aperture (Gentiulucci et al., 2001). This is also the case the other way around; the size of the manual grasp of an object affects the size of the aperture of the mouth. Furthermore, when participants had to grasp an object and simultaneously name a syllable printed on it (e.g., "GU", "GA"), the size of the manual grasp aperture affected lip opening as well as voice patterns during syllable expression, showing a clear entrainment between manual action and the articulatory system. In sum, the proposal that gestures are likely to occur alongside speech because motor activity cannot be completely inhibited is compatible with PPP and the existing literature about the mutual entrainment of the oral and manual systems.

4.3 Summary

We have shown that the basic tenets put forth by the GSA framework regarding how gestures emerge in the cognitive system are compatible with the claims made by Perspective Processing Perspectives (PPP). Put simply, gesture is produced by prediction errors that reach the motor plant. Namely, when the system predicts some motor-activity, it will produce prediction-errors - as there is no motor-activity yet that matches the predictions - which will in turn activate gestures. This is akin to what the GSA framework calls strong activation of action simulations. Yet there are constraints, on whether the motor system is activated. For example, when precision estimates of incoming motor-sensory signals are low relative to top-down predictions, than prediction will not be quashed by action, as the prediction error that results will be deemed less reliable and top-down predictions will suffice. This is akin to one way in which the GSA framework conceptualizes the gesture threshold, or the idea that action simulations must be strong enough to pass some resistance to gesture. When gesture does not seem useful to the cognitive or communicative situation, action will not be activated strongly enough to be realized as gesture.

Utilizing PPP to think about gesture offers more than just a shift in terminology. On the contrary, in PPP, active inference is a central catalyst for cognition, suggesting that action in the form of gesture may also benefit the cognitive system. As will be explained in the following section, thinking about gesture not just as simulated action, but also as predictive action offers a general explanatory mechanism for how gestures have their facilitative effects on cognition.

5. Gesture as Predictive Action

In PPP, action can serve as a means of reducing predictive load. That is, by engaging the motor system in action, the cognitive system is able to sample information about the consequences of a particular action that is more precise than the information gleaned from a top-down prediction. We contend that gesture, like action more generally, can have this same effect, by providing the cognitive system with useful sensori-motor information.

We suggest that the act of gesturing provides visual and proprioceptive feedback about the consequences of action that is not available in a static state. Gestures thus provide multimodal information that corresponds to (as they normally co-occur with) the causal consequences of *actually* acting on an object. These consequences thus inform top-down visual and motor predictions with actual kinematic information, which is arguably more reliable than having to predict such consequences completely top-down. What results is a generative model dealing with more reliable externally supplanted (visual, and proprioceptive) information, that allows for less risky (i.e., more accurate) perceptual inferences. This process has a number of potential benefits to the cognitive system.

For example, consider the well-documented finding that gestures reduce working memory demands, as speakers who gesture during a primary task (e.g., explaining their solution to a math problem) are able to perform better on a secondary task (e.g., remembering a string of letters) than speakers who do not gesture (e.g., Goldin-Meadow et al., 2001). In our view, this effect occurs because gestures have reduced the predictive load involved in the primary task. For instance, as speakers describe how to solve a mathematical factoring problem, they use their hands to explore how the numbers move to the relevant positions in the problem space. These gestures provide visual and proprioceptive feedback about where the numbers should be positioned, thereby making it easier for the generative model to operate as the solution is described. As a result, the cognitive system has more resources available to devote to a secondary task (e.g., remembering a list of letters).

The feedback provided by gesture may help problem solving, as well. For example, when speakers are solving a mental rotation problem, moving their hand as they would if they were actually turning the block to be rotated will activate visual information about the end state of that rotation. As they attempt to predict the objects' end state, participants gesture as a form of active inference to determine what the sensorimotor consequences of various amounts of rotation will be, and in doing so, actually provide themselves with information about what those sensorimotor consequences are. The same effect is seen as participants solve the Tower of Hanoi. Producing gestures as-if manipulating the physical apparatus affects problem-solving performance compared to not gesturing (Cooperrider et al., 2015). Such gestures inform top-down predictions with relevant kinematic information. Indeed, when this kinematic information is not relevant to solving the task, performance is hampered (Beilock & Goldin-Meadow, 2010).

A similar effect occurs during mental abacus. Abacus users, after repeated training, learn to do complex arithmetic without the abacus; top-down predictions are doing most of the work in these cases. Interestingly however, abacus users transitioning to do arithmetic without the abacus often use gestures, *as-if* manipulating the beads of an actual abacus (e.g., Hatano & Osawa, 1983). With time, these gestures dissipate, and abacus users learn to do calculations without moving, although they appear to still be using a strategy that involves imagining use of the abacus (see Frank & Barner, 2012 for evidence). Thus, gestures seem to offer some in-between strategy wherein they can supplant the now absent information normally afforded by a physical abacus. Indeed, while the abacus is absent, the affordance of generating proprioceptive and visual consequences that normally occur with acting on an abacus are ready to hand when gesturing. Using this second-hand information afforded by gesture (rather than interaction with the abacus), allows the generative model to deal with a certain amount of uncertainty still present in top-down predictions.

Can this account also explain the effects of gesture on linguistic processing? For example, gesture production has been shown in some circumstances to act as a cross-modal prime that speeds access to corresponding words (Krauss, 1998). In their Action-Based Language Model, Glenberg and Gallese (2012) explain this as occurring because the predictors associated with a physical action and the predictors associated with the articulation of the lexical label for that action are overlapping. We build on this explanation to offer the following account. When speakers are thinking of describing a particular action, they attempt to access the action plan for articulating the correct lexical label. When the precision estimate for accessing this label is low, speakers engage gesture as a means of gathering more information. Because linguistic knowledge is grounded in sensorimotor experience (e.g., Barsalou, 2008; Glenberg & Gallese, 2012), the act of gesture can provide proprioceptive, visual, or kinematic cues that then strengthen activation of the word.

This is especially apparent in the case of gesture-speech mismatches, in which a speaker conveys information in gesture that is not conveyed in the immediately accompanying speech (e.g., Church & Goldin-Meadow, 1986). Such gesture-speech mismatches have been observed in children (and adults) in a wide variety of learning tasks (e.g., solving mathematical equations, balance beam problems, and chemistry problems), and predict children's learning trajectory (for a review see, Goldin-Meadow & Alibali, 2015). When learning a new task, children tend to first produce incorrect solutions in both gesture and speech. With additional learning, it becomes more likely that a correct solution is expressed in either gesture or speech (but not both), before the child finally settles into a stable state where gesture and speech both express a correct strategy (Alibali & Goldin-Meadow, 1993). Thus, it seems that learning does not follow an either-or transition of understanding (i.e., eureka!), but a negotiation of different ways of understanding brought forth through gesture and speech.

From the present perspective, these different "ways" of understanding correspond to the different kinds of predictive processing that govern gesture and speech. Namely, explaining a solution in speech involves predictions that are linear and rule-based (i.e., knowing-that). Speech targets regularities that are present on slower time-scales, which can be applied to several phenomena independent of a single observation (e.g., in a conservation task, knowing that any action could be undone to return to the original state). Yet, these abstract regularities need to be observed and discovered to become articulable. Here, gesture comes into play. As the child thinks about the task without a clear top-down solution in mind (i.e., top-down prediction estimates for speech are low), the child initiates gesture as action to explore the manual affordances of the apparatus. These gestures are governed by the task's predicted manual affordances (e.g., know-how) and not necessarily by representations of an abstract rule (Pouw, Van Gog, et al., in press). The gestures provide proprioceptive and kinematic information about the transformation that goes beyond what the child can see in the stimulus. Through repeated instances of gesturing, invariants can be discovered that become parsed in meaningful sequences that correspond (or not) with segments in speech. Under this view, a stable state is reached when the prediction error between discovered higher order invariants in gesture are resolved with categorical speech predictions that target those invariants.

In sum, we believe that many of the documented effects of gesture on cognition and language can be explained by considering gesture as a case of active inference. By engaging action/gesture, the cognitive system creates new bottom-up input that can inform the top-down predictions necessary for a problem solving or language task.

6. Concluding Remarks

In conclusion, we have offered a preliminary sketch for considering gesture as a case of predictive action. Considering gesture as an example of action in a Predictive Processing Perspective offers a powerful description of how gestures come to have their facilitative effects, as well as how they arise out of anticipatory sensori-motor states (simulations vis-à-vis top-down proprioceptive predictions). Under this view, the distinction between the origin of gesture and their function in the cognitive system is not so clear. Gestures occur because they can have powerful effects on the cognitive system, yet the effects they have are the direct result of their origin as simulated and predictive actions in that cognitive system.

References

Alibali, M., & Goldin-Meadow, S. (1993). Gesture–speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25,* 468–523.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*(2), 169-188.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138-114.

Alibali, M. W., Yeo, A., Hostetter, A. B., & Kita, S. (under review). Representational gestures help speakers package information for speaking. In R. Church, S. Kelly, & M. Alibali (Eds.), *Why Gesture?*

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Beilock, S. L. & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, *21*, 1605-1610.

Clark, A. (2015a). Radical Predictive Processing. *The Southern Journal of Philosophy*, *53*(S1), 3-27.

Clark, A. (2015b). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Clark, A., (2013a). Whatever Next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181-204.

Clark, A. (2013b). "Gesture as thought," in *The Hand, an Organ of the Mind: What the Manual Tells the Mental*, ed. Z. Radman (Cambridge, MA: MIT Press), 255–268.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 7-19.

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*(4), 706.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102-115.

Chu, M., & Kita, S. (2015). Co-thought and co-speech gestures are generated by the same action generation process. Advance online publication. doi:10.1037/xlm0000168.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2013). Individual differences in frequency and salience of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*, 694–709.

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

Cook, S. W., Yip, T., & Goldin-Meadow, S. (2010). Gesture makes memories that last. *Journal of Memory and Language, 63,* 465-475.

Cooperrider, K., Wakefield, E., & Goldin-Meadow, S (2015). More than Meets the Eye: Gesture Changes Thought, even without Visual Feedback. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Darwin C (1998) *The Expression of the Emotions in Man and Animals*. London: Harper Collins.

Delgado, B., Gómez, J. C., & Sarriá, E. (2011). Pointing gestures as a cognitive tool in young children: Experimental evidence. *Journal of experimental child psychology*, *110*(3), 299-312.

Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in psychology*, *3*, 548. doi: 10.3389/fpsyg.2012.00548

Dunn, T. L., & Risko, E. F. (2015). Toward a Metacognitive Account of Cognitive Offloading. *Cognitive Science*. Advance online publication. doi: 10.1111/cogs.12273

Frank, M.C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General, 141*, 134-149.

Franklin, D. W., and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72*, 425–442.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, *13*(7), 293-301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138.

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in psychology*, 1-7.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology*, *86*(4), 1685-1699.

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*(2), 174-180.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905-922.

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

Goldin-Meadow, S., & Alibali, M.W. (2015). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *123*, 448-453.

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

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516-522.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(03), 377-396.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, *15*(1), 95-110.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*, 297-315. doi: 10.1037/a0022128

Hostetter, A. B. (2014). Action attenuates the effect of visibility on gesture rates. *Cognitive Science*, 1-14. doi: 10.1111/cogs.12113

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, *15*(3), 495-514.

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language, 63,* 245-253.

Hostetter, A. B., Alibali, M. W., & Bartholomew, A. E. (2011). Gesture during mental rotation. In *Proceedings of the 33rd Annual Conference of Cognitive Science Society* (pp. 1448-1453). Austin, TX: Cognitive Science Society.

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

Kelly, S., Byrne, K., & Holler, J. (2011). Raising the ante of communication: evidence for enhanced gesture use in high stakes situations. *Information*, *2*(4), 579-593.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*(4), 513-549.

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

Lupyan, G., & Clark, A. (2015). Words and the World Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279-284.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2015). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, 1-11.

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

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, *117*(3), 411-424.

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

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 04, 329-347.

Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends in cognitive sciences*, *18*(9), 451-456.

Pouw, W. T. J. L., De Nooijer, J. A., Van Gog, T., Zwaan, R. A., & Paas, F. (2014a). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*, 359.

Pouw, W. T. J. L., Mavilidi, M., Van Gog, T., Zwaan, R., & Paas, F. (under review). Gesturing during Mental Problem Solving Reduces Eye Movements, Especially for Individuals with Lower Visual Working Memory Capacity.

Pouw, W. T. J. L., Van Gog, T., Zwaan, R. A., & Paas, F (in press). Are gesture and speech mismatches produced by an integrated gesture-speech system? A more dynamically embodied perspective is needed for understanding gesture-related learning. *Behavioral and Brain Sciences*.

Rao, R.P.N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra classical receptive field effects, *Nature Neuroscience 2, 1,* 79–87.

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology*, *16*(2), 145-176.

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

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625-636.