cortex xxx (2011) 1–18

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

# Action-based language: A theory of language acquisition, comprehension, and production

*Arthur M. Glenberg* [a,b,*] *and Vittorio Gallese* [c,d]

[a] *Department of Psychology, Arizona State University, Tempe, AZ, USA*
[b] *Department of Psychology, University of Wisconsin, Madison, WI, USA*
[c] *Department of Neuroscience, University of Parma, Parma, Italy*
[d] *IIT (Italian Institute of Technology), Brain Center for Social and Motor Cognition, Parma, Italy*

ARTICLE INFO

*Article history:*
Received 3 February 2010
Reviewed 27 May 2010
Revised 12 October 2010
Accepted 13 April 2011
Published online xxx

*Keywords:*
Language
Embodiment
Mirror neurons
HMOSAIC model of action control

ABSTRACT

Evolution and the brain have done a marvelous job solving many tricky problems in action control, including problems of learning, hierarchical control over serial behavior, continuous recalibration, and fluency in the face of slow feedback. Given that evolution tends to be conservative, it should not be surprising that these solutions are exploited to solve other tricky problems, such as the design of a communication system. We propose that a mechanism of motor control, paired controller/predictor models, has been exploited for language learning, comprehension, and production. Our account addresses the development of grammatical regularities and perspective, as well as how linguistic symbols become meaningful through grounding in perception, action, and emotional systems.

## 1. Introduction

The nature of language and the evolutionary process producing it are still matters of debate. This is partly due to the complexity and multidimensional nature of language. What do we refer to when we speak about the language faculty? Viewing cognition as an embodied, situated, and social enterprise offers the possibility of a new approach. This view of language and cognition has important philosophical antecedents, especially in the phenomenological tradition (see Gallese, 2007, 2008), which argues that meaning does not inhabit a pre-given Platonic world of ideal and eternal truths to which mental representations connect and conform. Instead, phenomenology entertains a perspective compatible with many empirical results of contemporary cognitive neuroscience: meaning is the outcome of our situated interactions with the world.

With the advent of language, the notion of meaning changes, as it is amplified by freeing itself from being dependent upon specific instantiations of actual experience. Language affords the opportunity to connect all possible actions within a network, thereby expanding the meaning of individual situated experiences. Language does this by evoking the near totality of possibilities for action the world presents us, and by structuring those actions within a web of related meanings. By endorsing this perspective, it follows

---

* *Corresponding author*. Department of Psychology, Arizona State University, Mail Code 1104, Tempe, AZ 85287, USA.
E-mail address: glenberg@asu.edu (A.M. Glenberg).

0010-9452/$ – see front matter © 2011 Elsevier Srl. All rights reserved.
doi:10.1016/j.cortex.2011.04.010

that if we confine language solely to its predicative use, we inappropriately reify just one part of language's nature. Instead, our understanding of linguistic expressions is not solely an epistemic attitude; it is first and foremost a pragmatic attitude directed toward action.

Data from psychology, psycholinguistics, and neuroscience have demonstrated the importance of action systems to perception (Wilson and Knoblich, 2005), social processes such as mentalizing (Gallese and Goldman, 1998; Gallese, 2003a; Sommerville and Decety, 2006), and to language comprehension (Glenberg and Robertson, 1999; Pulvermüller, 2002, 2005, 2008; Gallese, 2007, 2008). The action-related account of language suggests that the neuroscientific investigation of what language is and how it works should begin from the domain of action. However, no formal theories of this interaction have been proposed. Here we adapt well-tested theories of motor control, the modular selection and identification for control (MOSAIC) and hierarchical modular selection and identification for control (HMOSAIC) theories (Haruno et al., 2003) to produce our theory of action-based language (ABL). We apply the theory to language acquisition, comprehension, and some aspects of production, namely gesture.

We begin with a brief review of recent work on the relation between language and action (for more complete reviews see Gallese, 2007, 2008; Glenberg, 2007; Pulvermüller, 2005; Rizzolatti and Craighero, 2007) and the neurophysiology of the connection between language and action. This review is followed by a description of the MOSAIC and HMOSAIC models and how we modify them to apply to language phenomena. In brief, we propose that (a) the neural networks underlying the HMOSAIC model include pre-motor mirror neurons and canonical neurons, and (b) Hebbian learning underlies the association of neural networks used in speech production (e.g., uttering the word "give") and action control (e.g., the act of giving) so that the meaning of the utterance is grounded in the action and the expected outcome of the action. We then discuss how the model applies to the acquisition of nouns, verbs, and syntactic structures, how it explains simple language comprehension, and we apply the model to gesture as one component of language production.

One caveat is important. Whereas we focus on the relation between language and action, we do not claim that *all* language phenomena can be accommodated by action systems. Even within an embodied approach to language, there is strong evidence for contributions to language comprehension by perceptual systems (e.g., Pulvermüller, 2002; Kaschak et al., 2005; Pecher et al., 2003; Saygin et al., 2010) and emotional systems (Havas et al., 2010, 2007), and we address some of this work in the discussion. Our primary goal, however, is to make progress in understanding what appear to be major contributions of action to language.

## 2. Language and action

At first blush, action would seem to have little in common with language (conceived as a cognitive, modular system). Nonetheless, strong links between language and action have been found in analyses based on evolution (e.g., Gentilucci and Corballis, 2006; Rizzolatti and Arbib, 1998), neurophysiology (see Rizzolatti and Craighero, 2007, for a review), and behavior. Here we focus on behavioral data and briefly touch on other approaches.

The Indexical Hypothesis (Glenberg and Robertson, 1999) asserts that sentences are understood by creating a simulation of the actions that underlie them. Glenberg and Kaschak (2002) tested this proposal in a task in which participants judged the sensibility of sentences describing the transfer of concrete objects such as "Andy delivered the pizza to you/You delivered the pizza to Andy" and abstract information, such as "Liz told you the story/You told Liz the story". As in these examples, half of the sensible sentences described transfer toward the reader and half described transfer away. Participants responded using a three-button box held in the lap so that the buttons were aligned on the front/back axis. Participants viewed a sentence by holding down the middle button with the preferred hand. In one condition, the "sensible" response was made by moving the preferred hand to the far button, thus requiring a movement consistent with a simulation of transfer to another person. In the other condition, the "sensible" response was made by pressing the near button, thus requiring a movement consistent with transfer from another person to the reader.

As predicted by the Indexical Hypothesis, there was an interaction in the time needed to judge sentence sensibility: judgments were faster when the action implied by the sentence matched the action required to make the response, and this was true for both the concrete and the abstract transfer sentences. Glenberg and Kaschak refer to this sort of interaction as an Action-sentence Compatibility Effect (ACE). De Vega (2008) has reported an ACE-type of interaction in understanding counterfactual sentences such as, "If the jeweler had been a good friend of mine he would have shown me the imperial diamond". Note that the sentence is abstract in that the precondition does not exist (i.e., the jeweler is not a good friend) nor did the event occur (i.e., the jeweler did not show the diamond).

Zwaan and Taylor (2006) obtained similar results using a radically different ACE-type of procedure. Participants in their experiments turned a dial clockwise or counterclockwise to advance through a text. If the meaning of a phrase (e.g., "he turned the volume down") conflicted with the required hand movement, reading of that phrase was slowed. Furthermore, Taylor and Zwaan (in preparation), have demonstrated an effect of forces described in sentences on the force used in responding to the sentence.

These results have been mirrored in the imaging and neuropsychology literatures. For example, using brain-imaging techniques, it has been shown that when processing language with content related to different effectors, effector-specific sectors of the pre-motor and motor cortical areas become active (Hauk et al., 2004; Tettamanti et al., 2005). Bak and Hodges (2003) discuss how degeneration of the motor system associated with motor neuron disorder (amyotrophic lateral sclerosis – ALS) affects comprehension of action verbs more than nouns. Glenberg et al. (2008a) demonstrate how use-induced plasticity in the motor system affects the processing of both concrete and abstract language.

Similarly, behavioral and kinematic studies have shown a modulation of motor responses related to the content of the

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

language material (Buccino et al., 2005; Scorolli et al., 2009). Furthermore, motor activation occurs very soon after a stimulus is presented, and only 22 msec after peak activation in auditory temporal areas (Pulvermüller, 2008). This early differential activation is difficult to reconcile with the hypothesis that motor effects reflect motor imagery after understanding is completed. Instead, the early activation is more consistent with the embodied simulation account of language understanding (Gallese, 2007, 2008).

## 3. Neurophysiology and the language-action connection

The neurophysiological basis for the language-based modulation of the motor system is most likely related to the properties of a set of neurons, the so-called mirror neurons, first discovered in area F5 of the monkey pre-motor cortex (Gallese et al., 1996; Rizzolatti et al., 1996). Here we review properties of mirror neurons [(and mirror neuron systems or mirror mechanisms – MMs)] that are relevant to the development of the ABL theory described later.

Mirror neurons discharge when the animal performs an object-related action with the hand or the mouth and when it observes the same or a similar action done by another individual. What matters most in driving a mirror neuron's discharge is the specific, goal-related motor act irrespective of the movements required to accomplish it (Rochat et al., 2010). A major step forward in the research on the MM consisted of the discovery that pre-motor and parietal mirror neurons not only code the goal of an executed/observed motor act, like grasping an object, but that they also code the overall action intention (e.g., bringing the grasped object to the mouth or into a container, Fogassi et al., 2005; Bonini et al., 2010). More recently it has been shown that a sub-group of pre-motor mirror neurons, part of the corticospinal tract, display complete suppression of discharge during action observation and when the observed action is intransitive, Kraskov et al., 2009. The authors of this study proposed that this suppression of mirror neuron discharge might be involved in the inhibition of self-movement during action observation. A similar finding for humans was also reported by Mukamel et al. (2010): some cells showed reverse mirror properties, that is, suppression of activity during action observation. Perhaps more importantly, by using single-cell recording in humans, Mukamel et al. firmly established the existence of mirror neurons in humans. Previous work with humans, often using imaging techniques and transcranial magnetic stimulation (TMS), had not been completely convincing (cf. Dinstein et al., 2008).

The MM has a hierarchical organization. At a high level of the hierarchy (and anterior in pre-motor cortex Rizzolatti and Craighero, 2004) are action intentions, such as grasping-to-eat. At a lower level, these intentions are realized by integrating sequences of goal-related motor acts, such as grasping, holding, bringing, placing in mouth – the different "words" of a "motor vocabulary" (see Rizzolatti et al., 1988) – to obtain different and parallel intentional "action sentences". These action sentences consist of temporarily chained sequences of motor acts properly assembled to accomplish a more distal goal-state. For example, the "motor vocabulary" of grasping-related neurons, by sequential chaining, reorganizes itself to map the fulfillment of the action intention (e.g., grasping-to-eat). The overall action intention is the goal-state of the ultimate goal-related motor act of the chain (e.g., placing food in the mouth). In our account of language, we will specifically adopt this hierarchical organization and propose that it is the backbone of syntax.

As already noted, the Mukamel et al. (2010) data firmly establish the existence of mirror neurons in humans. However, because of the location (determined by surgical considerations) and the sparseness of the single-cell recording, the data are not ideal for developing a theory of language. More relevant data have been produced by using imaging and TMS procedures. For example, it has been shown in humans that the observation of actions done with different effectors (hand, foot, and mouth) recruits the same motor representations active during the actual execution of those same actions (Buccino et al., 2001). These findings strongly support the existence of mirror neurons in the human motor system and have lead to the notion of a human MM involving areas in the frontal lobes (notably, Broca's area) and parietal lobes. The MM can also be activated by the typical sound of an action and even when actions are described verbally (for reviews see Rizzolatti and Craighero, 2004; Buccino et al., 2004, 2006; Gallese, 2007, 2008; Rizzolatti and Sinigaglia, 2007). Aziz-Zadeh et al. (2006) observed somatotopic organization and overlap between motor areas activated during observation of actions and motor areas activated during the comprehension of sentences describing those actions.

Although a thorough discussion of the homology between macaque monkey pre-motor area F5 and Broca's area (BA 44 and 45) in the human brain is beyond the scope of this study, a few considerations are in order. Neuroanatomical studies show that BA 44, the posterior part of Broca's area (also designated as *pars opercularis*), shares cytoarchitectonic properties with pre-motor cortex (BA 6). More specifically, comparative studies indicate that – in regard to cytoarchitecture – the *pars opercularis* of the inferior frontal gyrus is the human homolog of macaque area F5 (Von Bonin and Bailey, 1947; Petrides and Pandya, 1994). Furthermore, as discussed by Rizzolatti and Craighero (2004) and by Fadiga et al. (2006), functional evidence shows that human BA 44, in addition to contributing to speech, contributes (as does monkey area F5) to hand movements. Altogether, these data strongly suggest that human BA 44 is the homolog of monkey area F5.

Another aspect of the MM relevant to the ABL theory is the finding of an overlap between mechanisms of speech perception and speech production. Consider first the findings of Fadiga et al. (2002) who used TMS to probe motor activity during speech perception. When their Italian-speaking participants listened to words having a trilled double-r sound, they observed more activation of the tongue muscles than when listening to a double-f sound. These findings have been complemented by studies using functional magnetic resonance imaging (fMRI) (e.g., Wilson et al., 2004), as well as additional studies using TMS (Meister et al., 2007; Watkins et al., 2003; Watkins and Paus, 2004). One of the most convincing TMS studies was reported by D'Ausilio et al. (2009). They directed a double-pulse TMS stimulation at the motor

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

area that controls a particular articulator (e.g., tongue or lips), and they observed selectively enhanced identification of phonemes whose production depends upon the stimulated articulator. Taken together, these results suggest that there are speech mirror neurons (cf. Galantucci et al., 2006; Guenther et al., 2006), that is, neural structures that respond both to heard and observed speech and speech production.

Finally, in developing the ABL theory, we will also make use of neurophysiological findings regarding canonical neurons, also found in area F5 of the macaque pre-motor cortex. Canonical neurons are grasping-related neurons that fire not only when a grasping action is carried out, but also when the animal merely observes the object in absence of any detectable action on that object (Rizzolatti et al., 2000; Rizzolatti and Luppino, 2001; Gallese, 2003b). Unlike mirror neurons, however, canonical neurons do not respond to observed action. Thus, canonical neurons help to explain how the mere appearance of a graspable object in visual space activates the appropriate motor program of the intended type of hand–object interaction.

Given that both monkeys and humans have MMs, why do only humans have language? There are at least two reasons. First, compared to monkeys and other great apes, humans are ultra-social, and in particular, have a more highly developed sense of intersubjectivity and the ability for intention-sharing that may have encouraged the evolution of language. Second, compared to other great apes, control of the human hand (using Broca's area) may require more complex hierarchical control structures which could serve as templates for grammatical structures as explained below.

In conclusion, there is strong behavioral and neurophysiological evidence pointing to a close connection between the motor system and language: there is overlap between speech perception and production in Broca's area, and Broca's area contributes to both speech articulation and hand articulation. Furthermore, as briefly noted above (e.g., Aziz-Zadeh et al., 2006; Glenberg et al., 2008a, 2008b), a human MM responds to both action production and descriptions of the actions. Nonetheless, there are few formal accounts of the connection. We turn next to developing such an account.

## 4. The MOSAIC and HMOSAIC theories of action control

In this section we describe a theory of motor control developed by Wolpert and colleagues (e.g., Wolpert et al., 2003). After describing the theory in some detail, we take two novel steps. The first is to relate some operations of the theory to mirror and canonical neurons. We then discuss how a simple modification of this model can link it to language processes.

Two types of models are often invoked in theories of motor control. A controller (also referred to as a backward or inverse model) computes the context-sensitive motor commands to accomplish goals. Thus a controller might produce the commands to control effector trajectory and forces in reaching for and lifting a cup at a particular location. As discussed by Wolpert et al. (2003), these computations are far from trivial because the same motor command to the muscles will have different effects depending on muscle fatigue, changes in body configuration such as joint angles and hand position, and characteristics of the objects of interaction. To complicate the problem, the musculoskeletal system is not just high-dimensional, it is also a nonlinear system so that forces in different sequences may result in very different effects. Finally, learning of the controller is difficult because feedback in the form of perceptual information must be used to adjust motor processes.

The second type of model is a predictor (also referred to as a forward model). The function of the predictor is to predict effects (both motor and sensory consequences) of literal actions. The predictor makes use of an efference copy of the commands generated by controllers. That is, the same motor commands that are sent to the body to generate movement are also sent to the predictor and are used to generate predictions. These predictions are useful for (a) fast correction of movement before sensory feedback can be obtained, (b) determining if the movement was successful by comparing the prediction to actual sensory feedback, (c) enhancing perceptual accuracy (Grush, 2004; Wilson and Knoblich, 2005), and importantly (d) comparison of the predicted sensory feedback to actual feedback produces an error signal used in learning.

The Wolpert et al. MOSAIC model consists of multiple pairs of predictors and controllers even for relatively simple actions such as lifting a bottle. Each of the predictors and controllers is implemented as a recurrent neural network (see Wolpert and Kawato, 1998, for computational details). We will refer to a linked pair of a predictor and controller as a *module*. For example, the control of action for lifting a particular container may consist of one module for when the container is full (large force required) and one module for when the container is empty (less force required). Fig. 1 illustrates three such modules. Note that the predictor is responsible for predicting both sensory feedback and changes in the environment due to action. These predictions are expectations regarding how the body and the world will change as a function of the actions. In the terminology of classic cognitive science, the predictor corresponds to a mental model (Johnson-Laird, 1989), and in the terminology of embodied cognition, the predictor corresponds to a simulator (Barsalou, 1999).

In any particular context (e.g., lifting a container when there is uncertainty about the extent to which it is filled), several modules might be activated. The actual motor command is a weighted function of the outputs from the selected controllers (the weighting process is represented by the circle in Fig. 1, see Blakemore et al., 1998, for supporting data). The weighting is determined by the probability that a particular module is relevant in the particular context (the "responsibilities" described shortly). This weighted motor command also becomes the efference copy used by the predictors, that is, each predictor model generates predictions on the basis of the weighted motor commands. The predictions are compared to sensory feedback, and those predictions with small errors lead to an increased weight for the associated controller, as well as changes in the responsibilities.

Fig. 1 shows some of the components of Wolpert and Kawato's (1998) MOSAIC, but for simplicity, we have not shown several other components. Thus, in this illustration, (a) the predictors should be considered to also compare sensory-motor feedback to the predictions to generate an error signal,

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

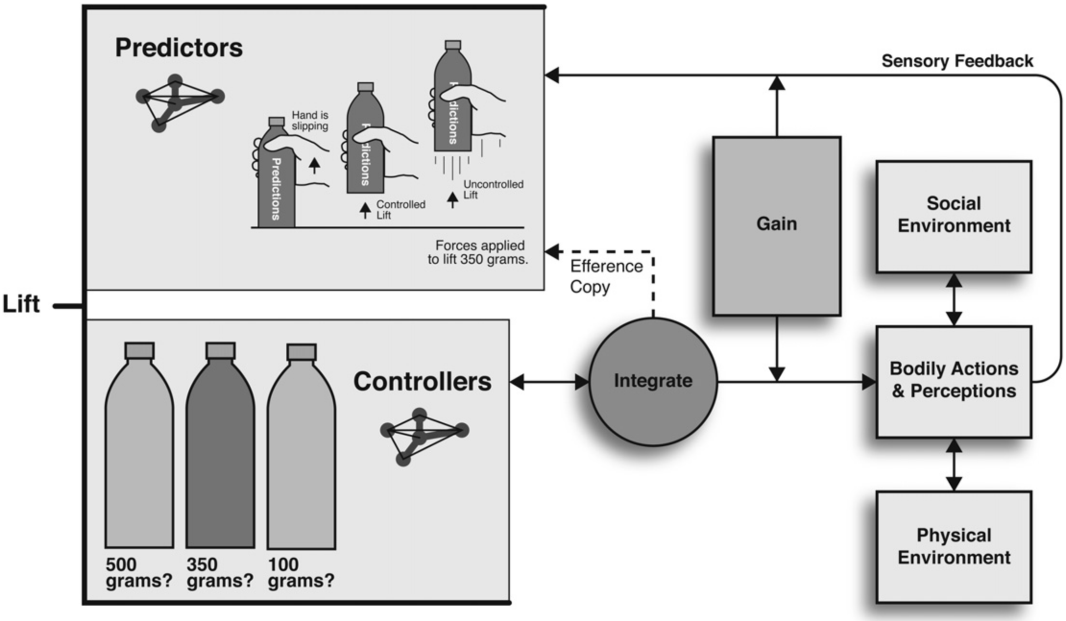

**Fig. 1 – The modified MOSAIC model of movement control for lifting a bottle. Because the container is opaque, it is not clear how much liquid it contains, and hence modules that generate multiple grasping and lifting forces (e.g., forces for 100, 350, and 500 g) are activated. The motor commands are integrated as weighted by the responsibilities, here illustrated by shading. The integrated commands (the efference copy) are fed back to the predictor modules), and each predictor module generates predictions.**

(b) the controller should be considered to compare motor feedback to its motor output to generate an error signal, and (c) we have completely suppressed illustration of the Responsibility Predictors that function in selecting modules for the context. Wolpert and Kawato (1998) demonstrate that the error signals are sufficient to learn (a) predictors, (b) controllers, and (c) the responsibilities.

Although the responsibilities (i.e., probabilities that a module is relevant in a particular context) are central to the operation of the MOSAIC model, here we only sketch how they might be updated. First, perception of the world provides information such as the location of objects and barriers, or the amount of liquid in a bottle, which will directly prime (increase the responsibility of) some modules more than others. Second, the error signal derived from a comparison of predicted and actual sensory feedback, is used to update responsibilities. For example, when lifting an opaque bottle, it may be uncertain as to whether the bottle is full (and requires a large grip force) or empty. When the efference copy of the integrated (i.e., weighted average) motor command is used by the low-force (100 g in Fig. 1) predictor, it predicts a lift off that is too fast for accurate control, whereas the same efference copy used by the high-force (500 g) predictor predicts a slippage of the fingers holding the bottle. If sensory feedback then indicates slippage, the responsibility of the high-force predictor is increased.

The predictors are generally learned faster than controllers because the output of the predictor and the sensory consequences can be coded using the same parameters (e.g., predicted proprioception and obtained proprioception). The relation between the output of the controller (e.g., a force to a particular muscle system) and the feedback (e.g., proprioception) is less direct. Furthermore, a well-trained predictor assists the learning of the controller in two ways. First, the predictor must make accurate predictions for the error signal to be useful in updating the controller. For example, if the controller generates an accurate movement signal, but the predictor makes a poor prediction, then the resulting error signal will force a change in the controller that is working properly. Second, when the predictor makes accurate predictions, the controller can be learned off-line, that is, by generating motor commands (but suppressing their execution) and observing whether the predictions correspond to the desired trajectory.

We add to the Wolpert and Kawato scheme two features addressed by Grush (2004) and Hurley (2004). Grush notes that gain control can be used to gate sensory feedback and thus serves several purposes (note in Fig. 1 the arrow from gain control to feedback, which is not part of the Wolpert et al. scheme). In noisy environments, feedback should have a reduced effect on the updating of modules. Also, as discussed in Hurley (2004), inhibiting the motor output allows the system to be "taken off-line" for simulation, imagery, deliberation, planning, off-line learning of the controller, and as we describe latter, for some language tasks. For example, in an imagery task, after selection of appropriate modules, the gain can be set low to ensure that only weak signals are sent to the muscles and that sensory feedback (of an unchanging body and environment) has little effect on adjusting the modules. Nonetheless, the predictors are free to generate predictions of the sensory feedback that would be obtained if literal action was taken. These predictions of how the environment would change are tantamount to imagery (cf. Grush, 2004).

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

Haruno et al. (2003) introduce a hierarchical version of MOSAIC, HMOSAIC, as illustrated in Fig. 2. In HMOSAIC, a module for a goal-directed action, such as drinking, selects basic motor act elements, such as grasping and lifting for the particular context. Although Fig. 2 illustrates two levels of hierarchy, in fact, more can be added without significant changes in the mathematics underlying HMOSAIC. Haruno et al. (2003), demonstrate how higher-level modules can learn to select the basic motor acts and learn the appropriate temporal orderings of those acts.

Whereas the architecture of the lower and upper levels of the hierarchy is almost identical, there are important differences. At the lowest level, motor commands are generated by the controller, and the predictor generates predictions of sensory-motor consequences based on the efference copy. At higher levels, the controllers generate vectors of prior probabilities that lower-level modules are relevant (thereby controlling the selection and ordering of the lower-level modules), and the higher-level predictors predict the posterior probabilities of the lower-level modules controlling behavior. Thus, the higher-level modules are more "abstract" compared to the lowest level. (Later, we will treat these probabilities as partially-executed simulations, or perceptual symbols as in Barsalou, 1999, and that is what is illustrated in Fig. 2.) Wolpert et al. (2003) suggest that the top-down plans and bottom-up constraints of HMOSAIC are one solution to the symbol grounding problem. Ultimately, control of behavior arises from the interplay of top-down control and prediction of lower-level modules combined with bottom-up feedback from sensation (at the lowest level) and posterior probabilities (higher levels).

## 5. Linking HMOSAIC to language

In this section we develop the HMOSAIC model so that it becomes a model of hierarchical control in language as well as action production, that is, the ABL model. This development is followed by a discussion of the application of ABL to selective aspects of language acquisition (learning nouns, verbs, and multi-word constructions), comprehension, and production (gesture, in particular). Each of these sections includes a description of how the ABL model applies and a brief review of supporting data.

It is often noted that language is a productive system in that a finite number of words and syntactic rules can be used to generate an infinite number of sentences. In communication, those sentences must properly encode a variety of

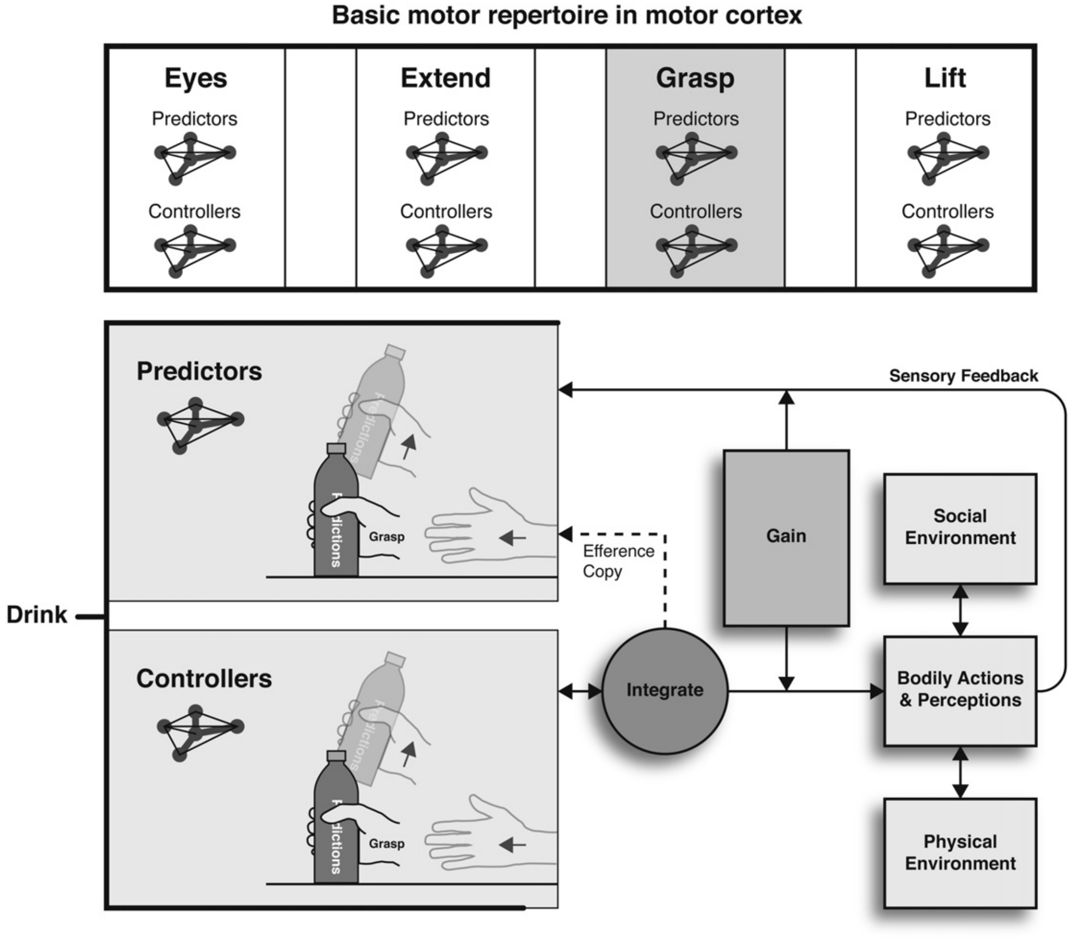

**Fig. 2 – The modified HMOSAIC model of action control for drinking. The figure illustrates the situation mid-way through the act of grasping-to-drink: the eyes have located an appropriate object, the arm has been extended, and the higher-level controller has selected the grasp modules to control the hand.**

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

constraints such as who is doing what to whom, number, gender, aspect, tense, and so on. But, getting combinations that make contextual sense is a difficult problem. For example, although hammers and tractors are both tools, both found on farms, both can be stepped on, and neither is strongly associated with the concept of a ladder, only one makes sense in the following context, “Because the ladder was broken, the farmer stepped on his hammer/tractor to paint the top of the barn wall”, (Glenberg and Robertson, 2000). Thus, an important goal for an embodied account of language is to produce contextually-appropriate and sensible (i.e., communicatively effective) combinations of words, not just syntactically correct combinations.

The problem of creating contextually-appropriate and effective combinations is also endemic to motor control. In fact, our general hypothesis is that the motor system has solved the problem of producing contextually-appropriate and effective behavior by being functionally organized in terms of goal-directed motor acts, and not in terms of movements (Rizzolatti et al., 1988, 2000; Umiltà et al., 2008; Rochat et al., 2010). Furthermore, the brain takes advantage of the solution of one difficult problem, namely contextually-appropriate action, to solve another difficult problem, namely contextually-appropriate language. Gallese and Lakoff (2005) have called this neural exploitation (see also Gallese, 2007, 2008), whereas Anderson (2010) refers to this idea as the “neural re-use” hypothesis.

Thus, we set for ourselves two major challenges. One is to develop a theory of language in which constraints on meaning are primary. The basic idea is that constraints on meaning reflect constraints on effective action. The other challenge is to demonstrate how the hierarchical organization of language can arise from the hierarchical organization of motor control.

Fig. 3 illustrates our version of neural exploitation, namely how the HMOSAIC theory can be linked to language. Along with others (e.g., Fadiga and Gallese, 1997; Rizzolatti and Arbib, 1998; Guenther et al., 2006), we propose that this opportunistic sharing of action control and language was made possible by the development of mirror neurons. Recall that the frontal MM overlaps with Broca’s area, which controls both the speech articulators and the hand (see Fadiga et al., 2006). This overlap is noted in Fig. 3 by adding a speech articulation component for those higher-order modules that correspond to actions associated with words. For these modules, both the predictors and controllers include models of speech articulation.

The overlap in Fig. 3 between the speech articulation and action control is meant to imply that the act of articulation primes the associated motor actions and that performing the actions primes the articulation. That is, we tend to do what we say, and we tend to say (or at least covertly verbalize) what we do. Furthermore, when listening to speech, bottom-up processing activates the speech controller (Fadiga et al., 2002; Galantucci et al., 2006; Guenther et al., 2006), which in turn activates the action controller, thereby grounding the meaning of the speech signal in action. Finally, note how a mechanism developed to produce hierarchical control over serial motor acts (e.g., the motor acts composing the action of drinking) is also used to produce hierarchical control over serial motor acts in speech production.

The theory illustrated in Fig. 3 provides a principled and novel account of what it means to understand a linguistic term such as “drink”, that is, how the word is grounded. First, the term is grounded in action (cf. Glenberg and Kaschak, 2002), that is, the controller for articulation of the word “drink” is associated with or overlaps with the controller for the action of drinking. In addition, the predictor for the articulation of “drink” is associated with the predictor for the action of drinking. That is, part of the knowledge of what “drink” means consists of expected consequences of drinking. Thus, in its essence, linguistic knowledge is predictive and grounded in perception and action.

In the ABL theory, predictions are driven by activity in the motor system (cf. Fiebach and Schubotz, 2006), however, the predictions themselves reside in activity across the brain. For example, predictions of how the body will change on the basis of action result from activity in somatosensory cortices, predictions of changes in spatial layout result from activity in visual and parietal cortices, and predictions of what will be heard result from activity in temporal areas. In this way, although ABL is based on motor processes, it addresses contributions of other systems to language.

The ABL theory makes a novel and strong prediction: adapting an action controller will produce an effect on language comprehension. There is some neuropsychological evidence in this regard. For example, Bak and colleagues (e.g., Bak and Hodges, 2003) demonstrate that motor disorders produced by ALS have a selective effect on language. Glenberg et al. (2008a) tested this prediction behaviorally using a use-induced plasticity paradigm (e.g., Classen et al., 1998). Participants first practiced a toward-the-body action or an away-from-the-body action. This practice then affected time to comprehend sentences: comprehension of sentences describing action toward-the-body was slowed by practice of the toward action, whereas comprehension of sentences describing action away-from-the body was slowed by practice of the away action.

## 6. Learning linguistic constructions

In this section we consider how the ABL theory can provide a unique account of several components of language acquisition. We begin with a consideration of the learning of nouns and verbs. We then demonstrate how the theory accounts for learning syntactic-like constructions such as the double-object construction that describes transfer events, such as “Give the horse [object 1] the apple [object 2]”.

Consider how an infant can associate a verbal label for an object (e.g., “bottle”) with the appropriate action module for the object. In this analysis, we assume that the infant has already developed some skill in how to interact with objects (e.g., a baby bottle).

In many Western cultures, parents often call attention to objects and actions when naming them for babies (Masur, 1997). For example, while displaying the baby’s bottle, a father may say, “Here’s your bottle”. Even when the infant’s attention is not directly elicited, the infant’s capacity for “intention-reading” (e.g., Tomasello, 2003) helps to ensure

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

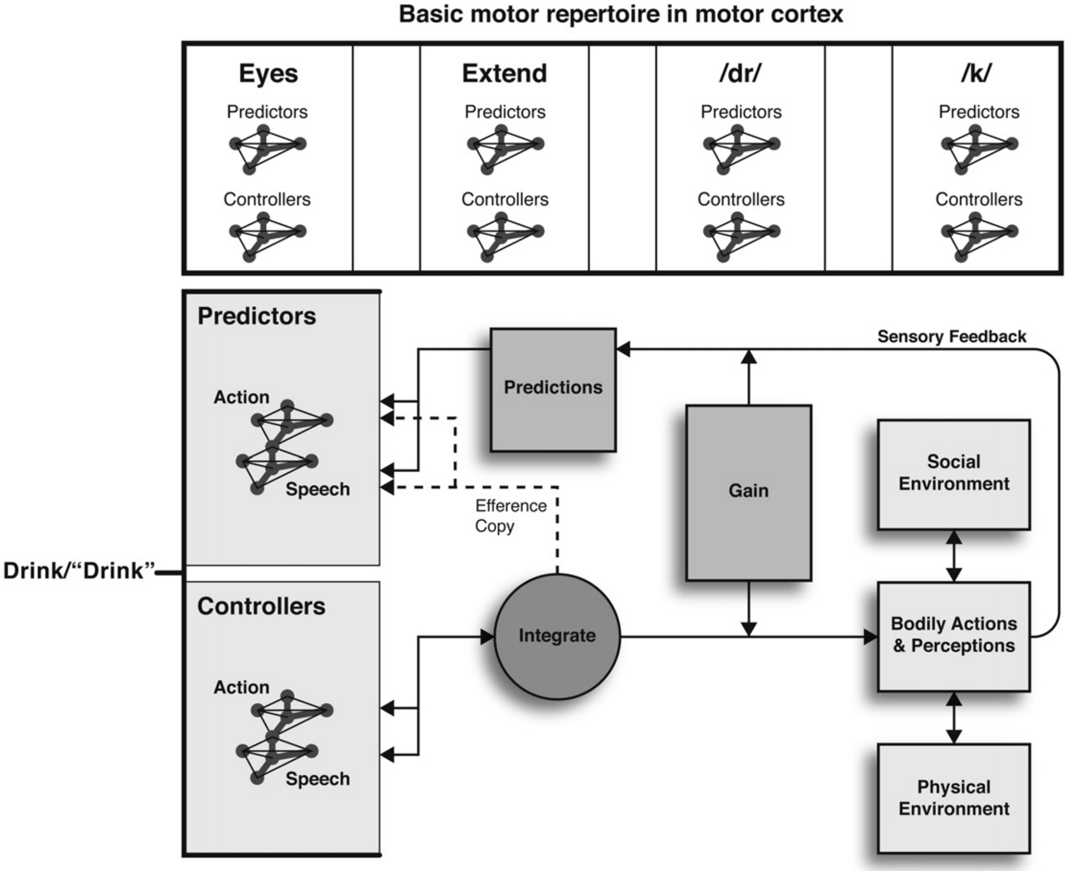

**Fig. 3 – The ABL model for understanding the verb "to drink".**

that parent and child are attending to the same components of the scene that a parent may be talking about.

Upon locating the object, the infant's canonical neuron system will be activated, thereby encoding the actions available to the infant for interacting with the object. That is, the visual information activates the appropriate module and its controller for activity with the bottle. At the same time, for the infant who has learned at least some of the articulations needed to pronounce "bottle", (in the infant's native language, of course) the infant's speech-MM is activated by the parent's spoken words. Note that both sets of activations are likely to be in Broca's area. Thus, the stage is set for Hebbian learning of the meaning of the spoken words by connecting the activated action controller and the activated speech controller. In effect, the module becomes the representation of a construction (Goldberg, 1995) that relates phonology (articulation) to meaning (action).

Based on this scheme, we propose that the meaning of a noun is grounded in two basic motor functions. The first is to call attention to an object named by the noun. According to the pre-motor theory of attention (e.g., Awh et al., 2006; Craighero et al., 1998; Rizzolatti et al., 1987), attention is the preparation of motor plans to act in regard to an object. Often this preparation is realized as the plan for a saccade of the eyes to the location of the object. Note that we are not suggesting that the eye movements are themselves an important component of meaning. Instead, what is critical is the HMOSAIC control of the eye movements and consequent predictions. For example, the controller for the bottle module generates commands to move the eyes. Using the efference copy, the predictor of the module generates the sensory feedback predicted upon moving the eyes and locating the object; in this case, those predictions likely include the size, shape, and color of the bottle. The second way in which a noun is grounded is to make available the affordances of the object that is attended. The motor realization of this second function is the activation of mirror neurons and the activation of canonical neurons.

Consider how the framework of Fig. 3 would control an infant's behavior upon hearing a noun such as "bottle". Hearing the noun activates speech mirror neurons for the word. These in turn activate the controller for interacting with bottles. The controller generates a large prior probability to move the eyes until the sensory feedback corresponding to a bottle is obtained, and a smaller prior probability to move the arm. The lower-level modules for controlling the eyes generate motor commands that are weighted by responsibilities sensitive to the location of the bottle in the context. The efference copy of the weighted commands is used by the predictors to generate a prediction of the sensory feedback (that the eyes will be moved to fixate a particular location and that a bottle will be seen). The predicted feedback is compared to actual feedback so that system can determine that the appropriate object has been attended. Once the bottle has been attended, the higher-level controller updates the prior probabilities so that the probability of moving the eyes is decreased and the probability of moving the arm and grasping (controlled by now-activated canonical neurons) is increased.

The attentional and eye movement function of nouns is attested by eye tracking data. For example, when experimental participants are listening to speech about the environment in which they are situated, the eyes saccade to mentioned objects immediately after the presentation of enough information to identify the noun (e.g., Altmann and Kamide, 2004; Chambers et al., 2004; see Thothathiri and Snedeker, 2008, for evidence with 3-year-old children). The second function of nouns, making affordances available, is supported by behavioral work (e.g., Borghi et al., 2004) and work in brain imaging (e.g., Hauk et al., 2004).

When language is being used to refer to objects and events that are not in the physical environment (e.g., when thinking, reading, or discussing distal events), similar processes ensue, except for differential operation of gain control. Hearing a noun activates the controller, and the first action is to attend to the object by planning for moving the eyes. Gain control inhibits some literal eye movements (but see Richardson and Spivey, 2000, for a demonstration that the eye movements are not completely inhibited). Nonetheless, an efference copy is sent to the predictor. Thus, the predictor generates an expectation, or image, of the object. The controller also activates movement elements such as moving the arm, but the literal movement is inhibited by gain control. The inhibited movement elements can be detected using TMS, as demonstrated by Buccino et al. (2005) and are a likely source of gesture as discussed later.

A similar account can be given for verb learning. For example, consider how an infant who knows how to drink (i.e., the HMOSAIC module for controlling drinking is in place), might learn the verb “to drink”. While an infant is drinking from a bottle, the parent might say, “good drinking!” The infant’s speech mirror neurons are activated by the parent’s speech, and a Hebbian learning process begins to establish connections between the action control for drinking and the motor representation of the speech signal. Later, a parent might say, “Drink your bottle”. If the infant has already learned the noun “bottle”, she might attend to the bottle, grasp it, and begin drinking. Suppose, however, the child focuses instead on the unknown word “drink” and does not undertake corresponding action. At this point, the parent might say, “Watch; this is what drink means” and the parent might mimic drinking from the bottle. Because the child knows how to drink, her MM will activate the controller for drinking, once again setting the stage for Hebbian learning between the modules for speech and the modules for action.

This account is broadly consistent with several types of data. First, in the ABL theory, there is not a principled difference between the learning of nouns (corresponding to how to interact with specific objects) and the learning of verbs. Although it has traditionally been believed that different cortical areas correspond to verb and noun processing (e.g., Friederici et al., 2006), recent data suggest a different story. Vigliocco et al. (2006) noted that much of the previous work indicating a cortical dissociation between nouns and verbs was confounded with word meaning. To address this confound, they studied words naming motor and sensory events either as nouns or verbs (the Italian verbal stimuli were declined so that grammatical class was obvious to the native Italian speakers). Participants listened to the words while positron emission tomography images were obtained. Differences emerged along the motor/sensory dimension, but not along the noun/verb dimension. That is, motor words, whether verbs or nouns, tended to activate motor areas of cortex as well as posterior left inferior frontal gyrus (Broca’s area). Sensory words, whether verbs or nouns, tended to activate left inferior temporal areas, and left anterior and posterior inferior frontal gyrus.

The ABL account of verb learning predicts that infants and children will learn verbs more proficiently if they have first learned the corresponding actions. An analysis of the MacArthur Child Development Inventory (CDI, Angrave and Glenberg, 2007) uncovered data consistent with this prediction. Using the CDI data, Angrave and Glenberg estimated average age of acquisition (in months) for actions such as drinking, sweeping, (pretend) reading, and the average age of production of the corresponding verbs. The correlation between the two was very strong ($p < .001$), and the linear relation was described by the function speech age $= .61$ (gesture age) $+ 17$ months. Thus, development of speech occurred in lockstep with the development of the action, however, the speech was delayed by about a year (see Buresh et al., 2006, for discussion of the gap between action and speech production).

Why might there be a gap between gesture production and speech production? Why is there a gap when the evidence is strong that infants can understand the speech well before they can produce it (e.g., Childers and Tomasello, 2002; Goldin-Meadow et al., 1976)? Part of the answer is that infants must accomplish the difficult task of controlling the speech articulators. The HMOSAIC model also suggests a computational answer for this gap. As noted by Wolpert and Kawato (1998) and described above, the learning accomplished by the predictor models is computationally simpler than the learning accomplished by the controllers. In addition, learning the controller model is necessarily delayed until accurate predictions can be made. Consequently, in the HMOSAIC model there is a delay between accurate prediction and performance of actions. Similarly, in ABL there is a delay between ability to comprehend speech (e.g., use it to direct eye movements, grasping, etc.) and being able to produce the same speech.

In natural situations, nouns are learned faster than verbs (e.g., Gentner, 2006). One reason seems to be that verb meanings are more variable than noun meanings. The ABL theory provides another reason based on the difficulty of learning the underlying actions. For both nouns and verbs, the theory predicts that learning will occur only after the child has learned appropriate actions, either interactions with objects or goal-directed actions such as giving. At the very least, appropriate action predictors must be learned to serve as the ground for speech learning. Typically, predictors for goal-directed actions will be more difficult to learn than predictors for object interactions. First, an infant’s goal-directed

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

behavior often (but certainly not always) involves a complex sequence of actions, whereas interactions with objects are less complex. For example, drinking from a cup involves locating the cup, extending the arm, grasping the cup, bringing the cup to the lips, sipping, and swallowing. In contrast, learning the meaning of a cup initially involves only locating and grasping. Second, the more complex goal-directed actions require learning a module at a higher level to control the order of the actions, as well as learning at lower levels.

Thus, the ABL theory makes the following predictions. Infants will find it easier to learn names of actions and objects for which they have already learned appropriate modes of interaction (cf. Huttenlocher et al., 1983). Also, given equivalent learning of modes of interaction, there should be little difference in the speed with which the infants can learn the associated nouns and verbs.

Before turning to the ABL analysis of larger syntactic units, several caveats are in order. First, we have, at best, provided a sketch of how nouns and verbs are learned (for more details about verb learning see Hirsh-Pasek and Golinkoff, 2006). Second, there are multiple syntactic classes (prepositions, adjective, adverbs, etc.) for which we have provided no account of learning. Development of such an account seems possible, but for now remains a promissory note.

Our hypothesis regarding larger syntactic units has three parts. First, the basic function of motor control is to use a hierarchical mechanism to combine movements in a way that produces effective action, that is, action that succeeds in accomplishing a goal. Second, the basic function of syntax is to combine linguistic components in a way that responds to or produces a communicative goal. Third, syntax emerges from modifying hierarchical control of action to produce hierarchical control of speech and responses to speech.

We propose that the mechanisms illustrated in Fig. 3, with one additional assumption, are sufficient to learn contextually-appropriate responses to simple sentences. The assumption is that when modules generate high-probability predictions that are disconfirmed, then control defaults to other modules (e.g., modules for attending to unpredicted events). However, when modules generate low-probability predictions that are disconfirmed, the modules maintain control over behavior, but the error signal is used to update the modules. This assumption will come into play later when we describe how rule-like behavior can result from the learning of specific actions.

As an example of learning a syntactic unit, consider how a child might begin to learn the meaning of the word “give” and the double-object construction. As before, assume that the child already knows (a) how to interact with various objects (e.g., a bottle, a cup, a cookie), (b) the names of those objects, and (c) how to engage in some complex activities such as giving.

Imagine that the child hears his father say, “Give Dada the cup”. Upon hearing “Dada” the child will attend to Dada and prepare to interact with him. Upon hearing “the cup”, the child will attend to the cup and perhaps grasp it. But suppose that the child has not associated the word “give” with any corresponding action. Consequently, the child will not give Dada the cup (at least not reliably in response to the verbal command). After several repetitions of “Give Dada the cup” without the infant doing much but holding the cup, Dada might extend his arm toward the infant while saying “Give Dada the cup”. Extending the arm will activate the child’s MM and the controllers for extending the infant’s arm. Some of these controllers are part of the infant’s knowledge of the actions of giving, and hence controllers for giving are activated. The act of giving is controlled by a high-level module that orchestrates, at a lower level, (a) attending to a speaker, (b) locating an object, (c) extending the arm, (d) grasping the object, (e) attending again to the recipient and extending the arm, and (f) releasing the object.

With repetition, the child learns the association between a specific string of words (e.g., “give-dada-the-cup”) and a specific high-level module for giving a cup to Dada. This specific high-level module produces a sequence of prior probabilities to (a) look at Dada, (b) move the eyes to the cup, (c) extend arm to the cup, (d) grasp the cup, (e) extend the arm to Dada, and (f) releasing the cup, namely, *give1* in Fig. 4. To say this differently, through repetition the child has learned an association between a specific sequence of words and a specific instantiation of the give action module, namely giving a cup to Dada.

Later, the child hears his mother say, “Give Momma the bottle”. Upon hearing “give”, the module for “give-dada-the-cup” is activated. The speech predictor generates the expectation that the next word heard will be “Dada”, that is, the only thing that the module has learned will happen in this context. Because the module has little else to predict, the prediction is made with confidence. However, the next word is, “Momma”. The disconfirmation of the high-probability prediction disrupts control of behavior. Now Momma must engage in repetition to teach appropriate responding to “give Momma the bottle”, resulting in *give2* in Fig. 4. Similar disconfirmations of high-probability predictions result from hearing “Give Lavinnia the spoon” (resulting in *give3*), “Give Dada the bowl” (resulting in *give4*), and “Give Dada the spoon” (resulting in *give5*).

What has the child learned? First, the child has five modules corresponding to five situations, not a general grammatical form. That is, the child knows how to respond to a stimulus such as “give-momma-the-bottle” rather than knowing how to parse and understand a sentence. This learning corresponds to what Tomasello (2003) calls holophrases, that is, relatively unparsed adult expressions that control behavior in particular situations. Second, each of the higher-level modules follows the same structure: in each case, the module controls a sequence of attending to the named speaker, moving attention to the object and grasping it, and then looking back at the named speaker, extending the arm and releasing the object. As described next, these specific modules can be the source of behavior that appears to arise from having learned a syntactic rule.

Consider now the child hearing the novel sentence, “Give Dada the bottle”. Upon hearing “give”, the speech mirror neurons associated with all of the “give” controllers are activated to an extent determined by their similarity to the context (i.e., the responsibilities). For example, if it is Dada talking, then the modules corresponding to “give-dada-...” will be most strongly activated. The responsibility-weighted efference copy that results from combining the outputs of

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

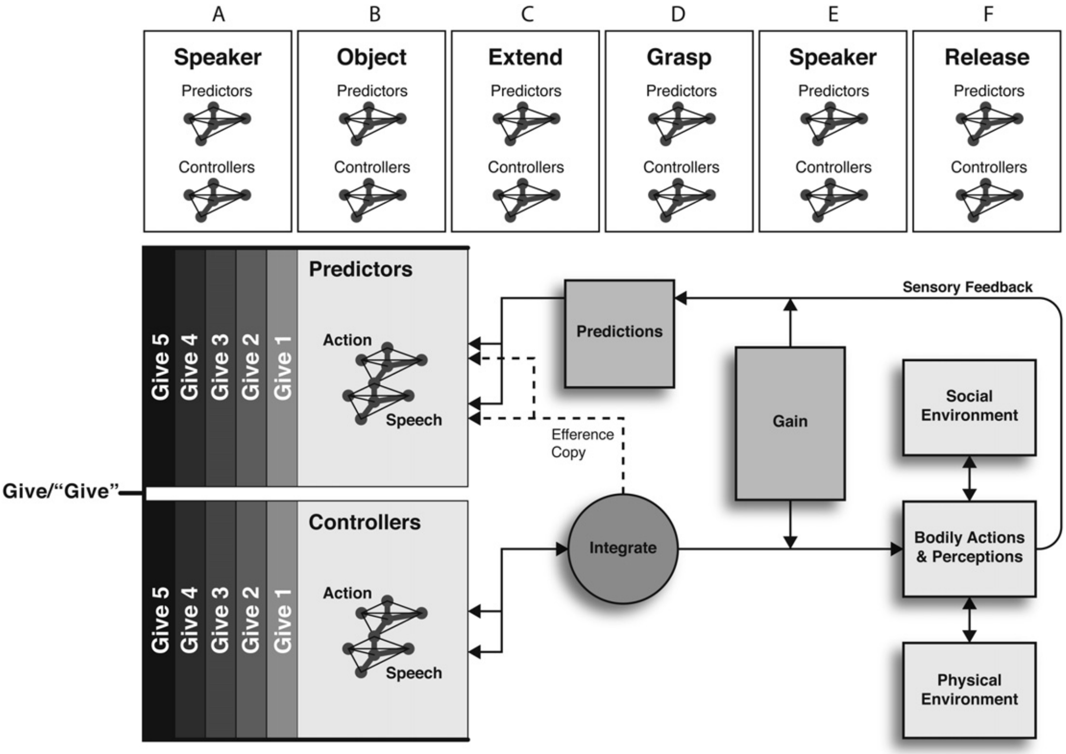

**Fig. 4 – The ABL model for understanding five instances of double-object sentences using the verb "to give". The figure illustrates five modules each of which corresponds to a particular instance of a double-object sentence in context.**

the *give1*–*give5* speech controllers are fed back to the predictors which strongly predict that the next word will be "Dada", and less strongly predict that the next word will be "Momma", or "Lavinnia". Upon hearing "Dada", predictions from *give1*, *give4*, and *give5* are confirmed, and those modules are given added responsibility in this context, whereas the responsibilities of the *give2* and *give3* modules are reduced.

The child is now looking at dada, and the highly-weighted predictors generate expectations for "cup", "bowl", and "spoon". Note that although the three modules have high responsibilities, particular predictions cannot be made with confidence, that is, each is equally likely. Thus, when the child hears "the bottle", there is a disconfirmation of low-probability predictions. In this case of disconfirmation of low-probability predictions, control of behavior by the *give1*, *give4*, and *give5* modules is not disrupted. Instead, upon hearing "the bottle", the infant executes the actions already associated with "the bottle", namely, locate the bottle, extend the arm, and grasp the bottle. Note that these three actions on the bottle correspond (in kind) to actions (b) – (d) predicted by the *give1*, *give4*, and *give5* predictors, thus updating the *give* controllers and predictors. At this point, the *give1*, *give4*, and *give5* controllers orient the child back to the last located person, the child extends the arm (e), and the child hands off the bottle (f). Thus, the child behaves appropriately to the novel sentence, "Give Dada the bottle". Our claim, however, is that the appropriate behavior has emerged from the operation of the HMOSAIC mechanism for integrating particular actions, not from using an abstract syntactic rule.

This generalization mechanism is not a simple blending. For example, it is not that the meaning of "bottle" is a blend of meanings of "bottle", "cup", "spoon" and so on. Instead, the ability to respond correctly to the sentence depends on having a well-articulated understanding of "bottle", namely how to locate and act on that sort of object. The generalization comes about for the following reason: How one literally gives bottles uses the same higher-order control structure for giving cups, spoons, and so on. Consequently, when those higher-order control structures are verbally invoked, they will accept the lower-order structure for acting on bottles.

To summarize, although the infant has learned nothing like a grammatical rule, we propose that having learned a sufficient number of *give* modules, the child can generalize and respond to many more sentences than those already learned. That is, *behavior* consistent with having learned an abstract structure or rule emerges from the operation of the HMOSIAC theory (cf. Hintzman, 1986).

This account of syntactic behavior makes several predictions. First, the theory predicts that modules encoding particular sequences of words (e.g., "give-dada-the-bottle") are maintained in memory and support generalization. This prediction was verified by Bannard and Mathews (2008). In their research, children were asked to repeat four-word sequences. These sequences were frequent in child-directed speech (e.g., "a drink of milk", "when we go out") or less frequent (e.g., "a drink of tea", "when we go in"). Nonetheless, the frequencies of the fourth words (e.g., "milk" and "tea") were matched. Bannard and Mathews demonstrated that children were (a) more accurate in repeating the more frequent sequences, and (b) faster to say the first three words in the more frequent sequences than the same three words in the less frequent sequences. These results imply that the word sequences were available in memory, not just the individual words (cf. Fig. 4).

Second, as described for the learning of "give", parents need to spend an inordinate amount of time teaching children verb use. Cameron-Faulkner, Lieven, and Tomasello (as reported in Tomasello, 2003) report that children hear about 5000–7000 utterances/day, but that more than half of these utterances begin with one of 52 high-frequency frames, and 45% of the utterances in their corpus began with just 17 words!

Third, children's learning of verb use is highly conservative. Tomasello (2003) writes that "within any given verb's development, there was great continuity such that new uses of a given verb almost always replicated previous uses and then made one small addition or modification…"(page 117).

Fourth, and most importantly for the account developed here, the theory describes a type of mesh of affordances. According to Glenberg and Robertson (2000; see also Kaschak and Glenberg, 2000), sentences are only sensible when the affordances of objects named in the sentence can be integrated, as guided by syntactic constructions, to result in goal-directed action. In the example, the child's understanding of how to interact with the bottle (i.e., how to locate it, how to grasp it) can be successfully integrated with the control structure for the *give* modules. If the sentence was instead "Give Fido the cup", the child and the model would reject the sentence as nonsense because, although Fido (a dog) can be located, Fido cannot take the cup. That is, at this point in the child's learning about "give", affordances of Fido do not mesh with the control sequence specified by the *give* modules. Similarly, at this point in learning, the child and the model will reject "Give Dada a kiss", because kisses cannot be literally grasped and handed off. Thus, the generalization mechanism will not always produce an integrated control structure. Instead, generalization is limited (at this stage of acquisition) by constraints on action. This prediction is confirmed by data reported in Tomasello (2003). He notes that children do not have adult-like semantic or syntactic categories. Instead, a particular verb might be used with verb-specific instruments such as "'thing to open with' and 'thing to draw with'" (Tomasello, 2003, page 118). Or, as in our example, the verb "to give" is only used with things that can be literally handed off.

## 7. Comprehension: using the motor system to guide simulation

A number of researchers have proposed that language comprehension is a process of simulation (e.g., Barsalou, 1999), and that the simulation makes use of the motor system (Gallese, 2007, 2008; Gallese and Lakoff, 2005; Glenberg and Robertson, 2000; Zwaan and Taylor, 2006). Here we provide an example of how the ABL theory produces such a simulation, and exactly what it is about that simulation that counts as language comprehension. Consider the understanding of a sentence read to a child as part of a story, "The girl gives the horse an apple". As part of learning about stories, the child has learned to set gain control low, so that he does not take literal action. Upon hearing "the girl", the child's speech mirror neurons activate the speech controller corresponding to the pronunciation of "girl" which in turn activates the associated action controller. The controller generates the motor commands for interacting with a girl, one of which is to move the eyes until the girl is located. An efference copy is used by the predictor of the *girl* module to generate a prediction of the sensory consequences of locating a girl. Note that this prediction corresponds to a mental image of a girl and can be considered the initial step in forming a mental model (Glenberg et al., 1987; Johnson-Laird, 1989).

Upon hearing "gives", the speech mirror neurons are activated in the many modules that encode various give actions. Some of these modules will be associated with double-object forms of give (as in the example sentence) and others will be associated with prepositional forms, such as "The girl gives an apple to the horse". The many double-object modules that might be activated by "give", combined with other components of context, will begin predicting the occurrence of many different recipients, such as "dada", "momma", "teacher", "Fido", and so on. Let's presume that the child has never before heard about giving apples to a horse, but he does know about horses, for example that they can eat things given to them. On the assumption that none of the previous-learned recipients is strongly related to this context, none of the individual predictions is made with a high probability. Nonetheless, virtually all of the double-object predictors predict that the next-named object will (a) require a new fixation (to dada, momma, etc.) and (b) that the object fixated will have the morphology needed to receive an apple (e.g., a hand or mouth). In contrast, the prepositional modules will be predicting that the next-named object will afford giving (e.g., an apple).

Upon hearing "the horse", the predictions made by the prepositional modules are disconfirmed, and will no longer be considered in detail.[1] Lower-level modules controlling action with horses become activated by the speech mirror neurons, eye movements are planned to a new location in space in which the mental model of a horse will be constructed, and that model is constructed from the predicted sensory feedback from the *horse* module. Because the *horse* module was activated, the various low-probability predictions (e.g., "dada", "momma") are disconfirmed. Nonetheless, because learned actions in regard to a horse can fit the double-object control structure, comprehension can proceed. Namely, upon hearing "an apple" the speech mirror neurons activate the *apple* module and the double-object module plans eye movements back to the agent (the girl) and predicts the sensory feedback of an apple in the agent's hand. Finally, the double-object modules direct attention (planned eye movements) from the apple-in-hand to the horse.

This account of the processes of comprehension offers a take on what it means to comprehend. Namely, comprehension is the process of fitting together actions suggested by the linguistic symbols so that those actions accomplish a higher-level goal

---

[1] One might trace out processing implications of when objects that are usually recipients (e.g., people) are actually the objects transferred, as in "The girl gave the boy to his mother". In this case, hearing "the boy" would confirm predictions from double-object modules and disconfirm the prepositional modules, which would then lead to difficulty when the sentence continues as a prepositional (MacDonald, 1999).

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

such as giving. Furthermore, as discussed before, in the ABL theory the motor system guides comprehension, but it is not the sole contributor. For example, predicting the sensory consequences of locating a girl are tantamount to generating a visual image of a girl, which uses non-motor processes.

These proposed comprehension processes are consistent with the data reviewed in the introduction. For example, Glenberg et al. (2008b) used TMS to demonstrate activation of motor systems controlling the hand while participants read sentences describing transfer of objects. Because in the ABL theory the word “give” is associated with the action of giving, we would expect just this sort of muscle activity. As another example, Pickering and Garrod (2007) reviewed data showing that language production mechanisms are used to make predictions during language comprehension, much as proposed by the ABL theory.

## 8. Gesture as one component of language production

The ABL model provides insight into several features of language production such as syntactic priming (Chang et al., 2006), interactive alignment in dialog (Pickering and Garrod, 2004), and gesture. Due to space constraints, here we only consider production of co-speech gesture, and even more specifically, representational gestures (McNeill, 1992). These gestures pertain to speech content by pointing to objects, by depicting with the hands object shapes and motion paths, and by using particular locations to metaphorically represent ideas such as the past.

Most people gesture while speaking (McNeill, 1992), and the gestures occur in tight temporal synchrony with speech. Gestures occur even when a blind person is speaking to another blind person (Iverson and Goldin-Meadow, 2001). Gesture can facilitate production (Krauss, 1998) and comprehension (Goldin-Meadow et al., 2001). Several elegant studies by Gentilucci and co-workers have shown a close relationship between speech production and the execution/observation of arm and hand gestures (for a review, see Gentilucci and Corballis, 2006; Gallese, 2007, 2008). In particular, Bernardis and Gentilucci (2006) showed that word and corresponding-in-meaning communicative arm gesture influence each other when they are simultaneously emitted: the second formant in the speech spectrum is higher when the word is pronounced together with the gesture. No modification in the second formant is observed when executing a meaningless arm movement involving the same joints. Conversely, the second formant of a pseudo-word is not affected by the execution of meaningful arm gestures. The same effects occur when gestures are observed rather than performed. In sum, spoken words and symbolic communicative gestures are tightly linked in the communication system.

The involvement of Broca’s area in translating the representations of communicative arm gestures into mouth articulation gestures was confirmed by transient inactivation of BA 44 with repetitive TMS (Gentilucci et al., 2006). Why are speech and gesture so closely related? Since BA 44 exhibits mirror properties, it is likely that through embodied simulation, the communicative meaning of gestures is fused with the articulation of sounds required to express it in words. It appears that within BA 44, “vehicle” and “content” of social communication are tightly interconnected (Gallese, 2007, 2008).

In the ABL theory, speaking a word will activate the corresponding action, and it is only through inhibition (by gain control) that the overt action is not performed. In agreement with Hostetter and Alibali (2008), we propose the following reason for why gestures are not completely inhibited: speaking requires action of the articulators, and hence, gain control cannot be set to inhibit all activity. Thus when speaking, some actions may be partially performed, and those actions in effectors other than the speech articulators are classified as gesture. Given that Broca’s area controls both speech articulation and hand action (e.g., Fadiga et al., 2006), it may be particularly difficult to inhibit hand action during speech.

Fig. 3 also makes clear why gestures can facilitate retrieval of pronunciation (Krauss, 1998). Namely, taking action (or the partial action corresponding to gesture) corresponds to running an action controller. This in turn will stimulate the controller for the production of the corresponding word.

The ABL theory also suggests how gesture can facilitate language comprehension. On seeing a speaker make a gesture, the listener’s mirror neurons will resonate, thereby activating the listener’s controllers and predictors for related concepts. Note that it is unlikely that a gesture will activate just one module. Nonetheless, given the structure of the ABL model, activation from the gesture can summate with activation from speech and contextual information to substantially reduce uncertainty as to what needs to be simulated. This priming can aid speech perception (Galantucci et al., 2006) and disambiguate meaning (Kelly et al., 1999).

## 9. Discussion

This discussion briefly addresses three issues that would seem to present difficulties for the ABL theory. The first is that any account of language centered on the motor system is already discredited by arguments against Behaviorism. The second issue is that language must take into account processes that are not motoric, and in particular that the representation of space plays a critical role in cognitive development (e.g., Mandler, 2008). Finally, we briefly address how a sensorimotor account can deal with the complexities of perspective and abstract concepts.

### 9.1. *ABL and Behaviorism*

A common misunderstanding of our position is that we are attempting to update a discredited Skinnerian account of language. To be sure, there are similarities between ABL and Skinnerian analyses: we are focused on behavior and the control of behavior; our invocation of the motor system may seem similar to the notion of “response” in an “S-R” psychology; our proposed mechanism for grounding linguistic knowledge is associative, that is, associations are formed between the controller for the production of the word and the controller for the production of actions relevant to the word’s meaning.

Nonetheless, ABL goes substantially beyond a simple S-R approach in three respects. First, ABL is based on hierarchical structures rather than serial structures, thus allowing for

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

more complex dependencies in action and speech. Second, because the learning mechanism is based on prediction and errors between prediction and sensory feedback, the learning can be more effective, nuanced, and independent of traditionally-conceived reward mechanisms. Third, the higher-order modules in ABL are explicitly symbolic (e.g., can stand in for direct experience, can act as a unit of thought, and can serve in predication) although the symbols are grounded, and not arbitrary. Fourth, an important component of language understanding in ABL consists of the predictions made by the predictor models. For example, understanding a description of drinking includes the predictions about what drinking feels like and how it quenches thirst. Because of these important differences, ABL goes substantially beyond a simple associative system, and the differences belie the criticism that the ABL account of language is simply the association of words to specific actions.

### 9.2. *Non-motor processes*

We have focused on motor processes for two related reasons. First, we believe that the basic function of cognition is control of action. From an evolutionary perspective, it is hard to imagine any other story. That is, systems evolve because they contribute to the ability to survive and reproduce, and those activities demand action. As Rudolfo Llinas puts it, "The nervous system is only necessary for multicellular creatures-…that can orchestrate and express active movement" (Llinas, 2001, page 15).[2] Thus, although brains have impressive capacities for perception, emotion, and more, those capacities are in the service of action. Second, we propose that systems that have evolved for control of situation-specific action have been exploited to control situation-specific language (see Anderson, 2010; Gallese, 2007, 2008).

Nonetheless, just as effective action requires coordination with other essential systems, so language requires coordination with perceptual and emotional systems found throughout the brain.[3] For example, we have discussed how hearing speech activates mirror neurons in predictors and controllers found in Broca's area. Clearly, however, this mirror neuron activation requires contributions from auditory and speech processing systems in the temporal cortex. Similarly, we have invoked the operation of action mirror neurons in recognizing the actions of others and canonical neurons in coding the affordances of objects. Both of these functions require contributions from visual and parietal cortices. Thus, we see our account as consistent with the model of word processing developed by Pulvermüller (e.g., Pulvermüller, 2002, 2008) in which cortical processing of action words begins in peri-sylvian areas, spreads to pre-motor areas including Broca's area, and then to motor cortex. Similarly, we see our account as consistent with models of the action mirror neuron system such as that proposed by Iacoboni and Wilson (2006) linking processing in superior temporal areas with mirror neuron systems in parietal and frontal areas.

The predictions generated by ABL predictor models also link to other neural systems. Although a prediction can be modeled as a vector of probabilities (Haruno et al., 2003), we believe that it is useful to relate this vector to the notion of simulation as discussed in Barsalou (1999). These simulations are built from neural activity in the sensorimotor areas that were utilized during initial perception and learning. For example, the simulation of eating an apple (or the prediction that an apple will be eaten) will invoke activity in visual cortex used in processing shape and color, and activity in motor and pre-motor cortex used in taking action, and so on. In this case, a high-probability prediction corresponds to a simulation that is mostly completed, whereas a low-probability prediction corresponds to a simulation that is only partially completed. The vector of probabilities corresponds to multiple simulations in various stages of completion.

At first glance, our emphasis on the importance of action for conceptualization would seem to conflict with theory and data showing the importance of spatial information for conceptualization. For example, Mandler (2008) proposes that the earliest concepts are redescriptions of innately salient spatial information such as the difference between biological and non-biological movement. The redescription makes use of an attentive mechanism Mandler calls "perceptual meaning analysis" that, for example, converts attention to spatial primitives (e.g., the biological motion arising from "an apple being put into a bowl") into a conceptualization such as "thing into container" (Mandler, 2008, page 212).

If one maps perceptual meaning analysis onto the MOSAIC notion of learning a predictor, then the two ideas can be brought into correspondence. First, the function of the predictor is to generate predictions regarding sensory feedback and the state of the world after action is taken, and the predictions often include spatial information. Second, because predictors can be learned more easily than controllers, it is indeed the case that spatial concepts can emerge prior to action. Finally, consider properties of concepts, which Mandler takes to be "declarative knowledge about object kinds and events that is potentially accessible to conscious thought" (page 207). Although declarative knowledge is a component of human concepts, it is not clear how declarative knowledge imbues concepts with their most important characteristic, namely that they are used in thinking and action. In contrast, the notion of prediction does just that. It allows one to go from

---

[2] An illustration of this principle is the life cycle of the sea squirt, a tunicate member of the phylum chordata. After hatching, the sea squirt spends a few hours to a few days as a tadpole-like creature with a primitive spinal cord, a brain, a light-sensitive organ, and the ability to express active movement. After it finds a suitable attachment site, such as a rock, the sea squirt never again expresses active movement. Almost as soon as the sea squirt stops moving, it begins to ingest its brain. No action, no need for a brain.

[3] Glenberg et al. (2009) develop and test the hypothesis that because emotions literally change the body (through physiological processes) that there is a change in perceived affordances. For example, when one is angry, physiological processes make forceful actions (e.g., slamming a car door) more possible than otherwise. Thus, a sentences such as "Reeling from the fight with that stubborn bigot, you slam the car door" is more readily understood when angry than when not (Havas et al., 2007, 2010). ABL could model some changes in affordances as a change in responsibilities. Another possibility is that predictor models directly access current bodily states, such as fatigue or anger, in deriving predictions. This mechanism fits well with research demonstrating that bodily states play a role in visual perception and action production (e.g., Proffitt et al., 2003; Witt et al., 2010).

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

mere sensation to a prediction of what is likely to happen next given that sensation. To say it bluntly, it is the ability to predict that is the essence of conceptualization.[4]

### 9.3. *Symbol manipulation, perspective, and abstract language*

A powerful description of language is that it consists of symbols (e.g., words) and rules for manipulating them (e.g., syntax). This type of description easily accounts for the facts that language is productive and compositional. On the other hand, the symbols and rules accounts of language have a difficult time explaining meaning (how the abstract symbols are grounded), language use, and development (see Tomasello, 2003, for data and arguments). The ABL theory bridges at least some of the gap between sensorimotor and symbolic accounts by virtue of the symbolic nature of the output of the high-level controller and predictor models. That is, these predictor models generate vectors of probabilities, or partially completed simulations, rather than specific actions. As Barsalou (1999) has demonstrated, these simulations can function as logical symbols in conceptual systems. At the same time, the ABL perceptual symbols are explicitly grounded in motor commands and the predicted sensory consequences of those commands.

Another important component of language use is that it forces a perspective. This need for perspective may be related to the fact that we have bodies so that we always experience events from a given perspective (e.g., from the side, at eye-level) rather than a god's eye view. Tomasello (2003) describes three dimensions of perspectival construal of a scene. The granularity dimension reflects the ability to describe objects coarsely (e.g., a thing) or in fine grain (e.g., a kitchen chair); the perspective dimension captures point of view, such as describing an exchange as buying or selling; and the function dimension corresponds to different construals of the same object according to different functions, such as a person being a father or a doctor. Tomasello goes on to note that, "The way that human beings use linguistic symbols thus creates a clear break with straightforward perceptual or sensory-motor cognitive representations—even those connected with events displaced in space and/or time—and enables human beings to view the world in whatever way is convenient for the communicative purpose on hand" (2003, page 13).

Whereas Tomasello may be correct that a simple sensorimotor account cannot include multiple perspectives, the ABL theory does provide a straightforward account of at least some aspects of perspective. Namely, the output of different predictor models for the same event can provide different construals of that event. Thus, when an event is described as an instance of giving, predictors will generate a sequence of expectations such as the possessor attending to the recipient, the possessor attending to the object, the possessor grasping the object, and so on. In contrast, when an event is described as an instance of taking, the predictors will generate a sequence of expectations such as the recipient attending to the object, the recipient receiving the object, and so on. In brief, language can force different construals of a scene that correspond to the outputs of different predictor models invoked by the language.

Central to Tomasello's analysis of perspective is the idea that humans, compared to other great apes, are specialists in social interaction and intention-sharing (Herrmann et al., 2007; see also Castiello et al., 2010). The ABL theory can capture this specialization by the assumption that the human MM, compared to other great apes, is more finely tuned to others. But note also our previous speculation (in the section on Neurophysiology and the language-action connection) as to why humans have unique language capabilities.

These ideas regarding perspective also provide a handle on the understanding of abstract ideas. To a first approximation, abstract ideas correspond to relations (cf. Barsalou, 1999). Consider some prototypically abstract ideas such as truth (a relation between a perceptible situation and a description of it), beauty (a relation between perceptible situations and emotions such as awe),[5] ownership (a relation between objects and what one can do with them), and kinship terms such as "aunt", "nephew", and "cousin." In all of these cases, understanding the abstract concept requires the ability to make predictions much like taking a perspective on a situation. Thus, in viewing a person as a "nephew" one takes a (culturally-determined) perspective that involves making predictions regarding behavior as well as people who will be aunts, uncles, and cousins. Our claim is that the HMOSAIC mechanism of learning predictors is sufficient to encode these sorts of relations and thus serves to ground abstract language. As another example, consider causal relations. A substantial component of understanding that A causes B is the prediction that in situation A, B is likely to follow. These relation-encoding predictions need not be the result of complex cognition. For example, Rovee-Collier and Capatides (1979) have shown that 3-month-old infants easily learn the causal relation between their own behavior (kicking) and events (a mobile moving).

### 9.4. *Conclusions*

We propose that hierarchical, goal-directed mechanisms of action control, namely paired controller/predictor modules, have been exploited for language learning, comprehension, and production. That is, the neural circuits for controlling the hierarchy of goal-related actions were "exploited" by selection pressure to serve the newly acquired function of language (Anderson, 2010; Gallese, 2007, 2008).

Motor cortex, to a large extent, controls individual synergies — relatively simple movements like extending and flexing the fingers, turning the wrist, flexing and extending the elbow, etc. In contrast, pre-motor cortex is more complex: it structures simple motor behaviors into coordinated motor acts.

---

[4] Nonetheless, how can the appreciation of music or visual artistic expression arise from action control? At the risk of becoming a theory of all things, we briefly note that musical appreciation can be enhanced by motor simulation (e.g., humming along) and prediction (using predictor models), whereas the appreciation of some visual art may depend on activation of the MM to induce empathetic understanding. Thus, although esthetic appreciation may not serve action control, esthetic appreciation may depend on action control.

[5] Thanks to Chad Mortensen for this observation.

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

Thus, pre-motor cortex provides a "phase structure" to actions and specifies the right parameter values in the right phases, e.g., by activating the appropriate clusters of corticospinal neurons in the appropriate temporal order. This information is conveyed through neural connections by the pre-motor cortex to specific regions of the primary motor cortex.

Similarly, as exemplified by the MM, the same pre-motor circuitry controlling action execution instantiates the embodied simulation of the observed actions of others. Thus, pre-motor cortex and the MM provide hierarchical control structures that can be exploited by language. The HMOSAIC architecture of the ABL model demonstrates how this exploitation could work.

## Author note

Arthur Glenberg, Department of Psychology, Arizona State University (glenberg@asu.edu); Vittorio Gallese, Department of Neuroscience, University of Parma. Support for this work was provided to AMG by NSF grants BCS 0315434 and BCS 0744105 and to VG by MIUR (Ministero Italiano dell'Università e della Ricerca) and by the EU grants DISCOS, NESTCOM and ROSSI. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

## REFERENCES

Altmann GTM and Kamide Y. Now you see it, now you don't: Mediating the mapping between language and the visual world. In Henderson JM and Ferreira F (Eds), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*. New York: Psychology Press, 2004: 347–386.

Anderson ML. Neural re-use as a fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33: 245–266. doi:10.1017/S0140525X10000853, 2010.

Angrave LC and Glenberg AM. *Infant Gestures Predict Verb Production One Year Later*. Paper presented at the annual meeting of the American Psychological Association.

Awh E, Armstrong KM, and Moore T. Visual and oculomotor selection: Links, causes and implications for spatial attention. *Trends in Cognitive Sciences*, 10(3): 124–130, 2006.

Aziz-Zadeh L, Wilson SM, Rizzolatti G, and Iacoboni M. Embodied semantics and the premotor cortex: Congruent representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, 16: 1818–1823, 2006.

Bak TH and Hodges JR. The effects of motor neurone disease on language: Further evidence. *Brain and Language*, 89: 354–361, 2003.

Bannard C and Mathews D. Stored word sequences in language learning. *Psychological Science*, 19: 240–248, 2008.

Barsalou LW. Perceptual symbol systems. *The Behavioral and Brain Sciences*, 22: 577–660, 1999.

Bernardis P and Gentilucci M. Speech and gesture share the same communication system. *Neuropsychologia*, 44: 178–190, 2006.

Blakemore SJ, Goodbody SJ, and Wolpert DM. Predicting the consequences of our own actions: The role of sensorimotor context estimation. *The Journal of Neuroscience*, 18: 7511–7518, 1998.

Bonini L, Rozzi S, Serventi FU, Simone L, Ferrari PF, and Fogassi L. Ventral premotor and inferior parietal cortices make distinct contribution to action organization and intention understanding. *Cerebral Cortex*, 20: 1372–1385, 2010.

Borghi AM, Glenberg AM, and Kaschak MP. Putting words in perspective. *Memory and Cognition*, 32: 863–873, 2004.

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Freund HJ, et al. Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, 13(2): 400–404, 2001.

Buccino G, Binkofski F, and Riggio L. The mirror neuron system and action recognition. *Brain and Language*, 89(2): 370–376, 2004.

Buccino G, Riggio L, Melli G, Binkofski F, Gallese V, and Rizzolatti G. Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24(3): 355–363, 2005.

Buccino G, Solodkin A, and Small S. Functions of the mirror neuron system: Implications for neurorehabilitation. *Cognitive Behavioral Neurology*, 19: 55–63, 2006.

Buresh JS, Woodward A, and Brune CW. The roots of verbs in prelinguistic action knowledge. In Hirsh-Paskek K and Golinkoff RM (Eds), *Action Meets Word: How Children Learn Verbs*. New York: Oxford University Press, 2006: 208–227.

Castiello U, Becchio C, Zoia S, Nelini C, Sartori L, et al. Wired to be social: The ontogeny of human interaction. *PLoS ONE*, 5(10): e13199. doi:10.1371/journal.pone.0013199, 2010.

Chambers CG, Tanenhaus MK, and Magnuson JS. Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3): 687–696, 2004.

Chang F, Dell GS, and Bock K. Becoming syntactic. *Psychological Review*, 113(2): 234–272, 2006.

Childers JB and Tomasello M. Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6): 967–978, 2002.

Classen J, Liepert J, Wise SP, Hallett M, and Cohen LG. Rapid plasticity of human cortical movement representation induced by practice. *Journal of Neurophysiology*, 79(2): 1117–1123, 1998.

Craighero L, Fadiga L, Rizzolatti G, and Umiltà C. Visuomotor priming. *Visual Cognition*, 5(1): 109–125, 1998.

D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, and Fadiga L. The motor somatotopy of speech perception. *Current Biology*, 19(5): 381–385, 2009.

De Vega M. Levels of embodied meaning. From pointing to counterfactuals. In de Vega M, Glenberg AM, and Graesser AC (Eds), *Symbols, Embodiment, and Meaning*. Oxford: Oxford University Press, 2008: 285–308.

Dinstein I, Thomas C, Behrmann M, and Heeger DJ. A mirror up to nature. *Current Biology*, 18(1): R13–R18, 2008.

Fadiga L and Gallese V. Action representation and language in the brain. *Theoretical Linguistics*, 23(3): 267–280, 1997.

Fadiga L, Craighero L, Buccino G, and Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15(2): 399–402, 2002.

Fadiga L, Craighero L, and Roy A. Broca's region: A speech Area? In Grodzinsky Y (Ed), *Broca's Region*. Oxford: Oxford University Press, 2006: 137–152.

Fiebach CJ and Schubotz RI. Dynamic anticipatory processing of hierarchical sequential events: A common role for Broca's area and ventral premotor cortex across domains? *Cortex*, 42(4): 499–502, 2006.

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, and Rizzolatti G. Parietal lobe: From action organization to intention understanding. *Science*, 308(5722): 662–667, 2005.

Friederici AD, Bahlmann J, Heim S, Schubotz RI, and Anwander A. The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings*

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010

*of the National Academy of Sciences of the United States of America*, 103(7): 2458–2463, 2006.

Galantucci B, Fowler CA, and Turvey MT. The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, 13(3): 361–377, 2006.

Gallese V. The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*, 358(1431): 517–528, 2003a.

Gallese V. A neuroscientific grasp of concepts: From control to representation. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*, 358(1435): 1231–1240, 2003b.

Gallese V. Before and below theory of mind: Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*, 362(1480): 659–669, 2007.

Gallese V and Goldman A. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12): 493–501, 1998.

Gallese V and Lakoff G. The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, 22(3): 455–479, 2005.

Gallese V, Fadiga L, Fogassi L, and Rizzolatti G. Action recognition in the premotor cortex. *Brain*, 119(2): 593–609, 1996.

Gallese V. Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, 3: 317–333, 2008.

Gentilucci M and Corballis MC. From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews*, 30(7): 949–960, 2006.

Gentilucci M, Bernardis P, Crisi G, and Volta RD. Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *Journal of Cognitive Neuroscience*, 18(7): 1059–1074, 2006.

Gentner D. Why verbs are hard to learn. In Hirsh-Paske K and Golinkoff RM (Eds), *Action Meets Word: How Children Learn Verbs*. New York: Oxford University Press, 2006: 544–564.

Glenberg AM. Language and action: creating sensible combinations of ideas. In Gaskell G (Ed), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 2007: 361–370.

Glenberg AM and Kaschak MP. Grounding language in action. *Psychonomic Bulletin and Review*, 9(3): 558–565, 2002.

Glenberg AM and Robertson DA. Indexical understanding of instructions. *Discourse Processes*, 28: 1–26, 1999.

Glenberg AM and Robertson DA. Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43(3): 379–401, 2000.

Glenberg AM, Meyer M, and Lindem K. Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26(1): 69–83, 1987.

Glenberg AM, Sato M, and Cattaneo L. Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology*, 18(7): R290–R291, 2008a.

Glenberg AM, Sato M, Cattaneo L, Riggio L, Palumbo D, and Buccino G. Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology*, 61(6): 905–919, 2008b.

Glenberg AM, Webster BJ, Mouilso E, Havas DA, and Lindeman LM. Gender, emotion, and the embodiment of language comprehension. *Emotion Review*, 1(2): 151–161, 2009.

Goldberg AE. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press, 1995.

Goldin-Meadow S, Seligman MEP, and Gelman R. Language in the two-year-old: Receptive and productive stages. *Cognition*, 4(2): 189–202, 1976.

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

Grush R. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27: 377–442, 2004.

Guenther FH, Ghosh SS, and Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3): 280–301, 2006.

Haruno M, Wolpert DM, and Kawato M. Hierarchical MOSAIC for movement generation. *International Congress Series*, 1250: 575–590, 2003.

Hauk O, Johnsrude I, and Pulvermüller F. Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41: 301–307, 2004.

Havas DA, Glenberg AM, and Rinck M. Emotion simulation during language comprehension. *Psychonomic Bulletin and Review*, 14: 436–441, 2007.

Havas DH, Glenberg AM, Gutowski K, Lucarelli M, and Davdison R. Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychological Science*, 21: 895–900, 2010.

Herrmann E, Call J, Hernandez-Lloreda MV, Hare B, and Tomasello M. Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, 317: 1360–1366, 2007.

Hintzman DC. "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93(4): 411–428, 1986.

Hirsh-Pasek K and Golinkoff RM. *Action Meets Word: How Children Learn Verbs*. New York: Oxford University Press, 2006.

Hostetter AB and Alibali MW. Visible embodiment: Gesture as simulated action. *Psychonomic Bulletin and Review*, 15(3): 495–514, 2008.

Hurley S. The shared circuits hypothesis: A unified functional architecture for control, imitation, and simulation. In Hurley S and Chater N (Eds), *Perspectives on Imitation: From Neuroscience to Social Science*. Cambridge: MIT Press, 2004.

Huttenlocher J, Smiley P, and Charney R. Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review*, 90(1): 72–93, 1983.

Iacoboni M and Wilson SM. Beyond a single area: Motor control and language within a neural architecture encompassing Broca's area. *Cortex*, 42(4): 503–506, 2006.

Iverson JM and Goldin-Meadow S. The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science*, 4(4): 416–422, 2001.

Johnson-Laird PN. Mental models. In Posner MI (Ed), *Foundations of Cognitive Science*. Cambridge: MIT Press, 1989.

Kaschak MP and Glenberg AM. Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, 43(3): 508–529, 2000.

Kaschak MP, Madden CJ, Therriault DJ, Yaxley RH, Aveyard M, Blanchard A, et al. Perception of motion affects language processing. *Cognition*, 94(3): B79–B89, 2005.

Kelly SD, Barr DJ, Church RB, and Lynch K. Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40(4): 577–592, 1999.

Kraskov A, Dancause N, Quallo MM, Shepherd S, and Lemon RN. Corticospinal neurons in macaque ventral premotor cortex with mirror properties: A potential mechanism for action suppression? *Neuron*, 64(6): 922–930, 2009.

Krauss RM. Why do we gesture when we speak? *Current Directions in Psychological Science*, 7(2): 54–60, 1998.

Llinas R. *The I of the Vortex:From Neurons to Self*. Cambridge, MA: MIT Press, 2001.

MacDonald MC. Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In MacWhinney B (Ed), *The Emergence of Language*. Mahwah: Erlbaum, 1999: 177–196.

Mandler JM. On the birth and growth of concepts. *Philosophical Psychology*, 21(2): 207–230, 2008.

Masur EF. Maternal labeling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language*, 24(2): 427–439, 1997.

McNeill D. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press, 1992.

Meister IG, Wilson SM, Deblieck C, Wu AD, and Iacoboni M. The essential role of premotor cortex in speech perception. *Current Biology*, 17(19): 1692–1696, 2007.

Mukamel R, Ekstrom AD, Kaplan J, Iacoboni M, and Fried I. Single-neuron responses in humans during execution and observation of actions. *Current Biology*, 20(8): 750–756, 2010.

Pecher D, Zeelenberg R, and Barsalou LW. Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, 14(2): 119–125, 2003.

Petrides M and Pandya DN. Comparative architectonic analysis of the human and the macaque frontal cortex. In Boller F and Grafman J (Eds), *Handbook of Neuropsychology*, Vol. 9. Amsterdam: Elsevier, 1994: 17–58.

Pickering MJ and Garrod S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2): 169–226, 2004.

Pickering MJ and Garrod S. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3): 105–110, 2007.

Proffitt DR, Stefanucci J, Banton T, and Epstein W. The role of effort in distance perception. *Psychological Science*, 14(2): 106–112, 2003.

Pulvermüller F. *The Neuroscience of Language*. Cambridge, UK: Cambridge University Press, 2002.

Pulvermüller F. Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7): 576–582, 2005.

Pulvermüller F. Grounding language in the brain. In de Vega M, Glenberg AM, and Graesser AC (Eds), *Symbols, Embodiment, and Meaning*. Oxford: Oxford University Press, 2008: 85–116.

Richardson DC and Spivey MJ. Representation, space, and Hollywood squares: Looking at things that aren't there anymore. *Cognition*, 76(3): 269–295, 2000.

Rizzolatti G and Arbib M. Language within our grasp. *Trends in Neurosciences*, 21(5): 188–194, 1998.

Rizzolatti G and Craighero L. The mirror-neuron system. *Annual Review of Neuroscience*, 27: 169–192, 2004.

Rizzolatti G and Craighero L. Language and mirror neurons. In Gaskell G (Ed), *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 2007.

Rizzolatti G and Luppino G. The cortical motor system. *Neuron*, 31(6): 889–901, 2001.

Rizzolatti G and Sinigaglia C. *Mirrors in the Brain. How Our Minds Share Actions, Emotions, and Experience*. Oxford: Oxford University Press, 2007.

Rizzolatti G, Riggio L, Dascola I, and Umiltà C. Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory. *Neuropsychologia*, 25(1): 31–40, 1987.

Rizzolatti G, Camarda R, Fogassi M, Gentilucci M, Luppino G, and Matelli M. Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. *Experimental Brain Research*, 71(3): 491–507, 1988.

Rizzolatti G, Fadiga L, Gallese V, and Fogassi L. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2): 131–141, 1996.

Rizzolatti G, Fogassi L, and Gallese V. Cortical mechanisms subserving object grasping and action recognition: a new view on the cortical motor functions. In Gazzaniga MS (Ed), *The Cognitive Neurosciences*. 2nd ed. Cambridge: MIT Press, 2000: 539–552.

Rochat MJ, Caruana F, Jezzini A, Escola L, Intskirveli I, Grammont F, et al. Responses of mirror neurons in area F5 to hand and tool grasping observation. *Experimental Brain Research*, 204(4): 605–616, 2010.

Rovee-Collier CK and Capatides JB. Positive behavioral contrast in 3-month-old infants on multiple conjugate reinforcement schedules. *Journal of the Experimental Analysis of Behavior*, 32(1): 15–27, 1979.

Saygin AP, McCullough S, Alac M, and Emmorey K. Modulation of BOLD response in motion-sensitive lateral temporal cortex by real and fictive motion sentences. *Journal of Cognitive Neuroscience*, 22(11): 2480–2490, 2010.

Scorolli C, Borghi AM, and Glenberg AM. Language-induced motor activity in bi-manual lifting. *Experimental Brain Research*, 193(1): 43–53, 2009.

Sommerville and Decety. Weaving the fabric of social interaction: Articulating developmental psychology and cognitive neuroscience in the domain of motor cognition. *Psychonomic Bulletin and Review*, 13(2): 179–200, 2006.

Taylor LJ and Zwaan RA. Language modulates response force. in preparation.

Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, Scifo P, et al. Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17(2): 273–281, 2005.

Thothathiri M and Snedeker J. Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language*, 58(2): 188–213, 2008.

Tomasello M. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press, 2003.

Umiltà MA, Escola L, Intskirveli I, Grammont F, Rochat M, Caruana F, et al. How pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences*, 105(6): 2209–2213, 2008.

Vigliocco G, Warren J, Siri S, Aruili J, Scott S, and Wise R. The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex*, 16(12): 1790–1796, 2006.

von Bonin G and Bailey P. *The Neocortex of Macaca Mulatta*. Urbana, IL: University of Illinois Press, 1947.

Watkins K and Paus T. Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience*, 16(6): 978–987, 2004.

Watkins KE, Strafella AP, and Paus T. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 48(8): 989–994, 2003.

Wilson M and Knoblich G. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3): 460–473, 2005.

Wilson SM, Saygin AP, Sereno MI, and Iacoboni M. Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7): 701–702, 2004.

Witt JK, Proffitt DR, and Epstein W. How and when does action scale perception? *Journal of Experimental Psychology: Human Perception and Performance*, 36(5): 1153–1160, 2010.

Wolpert DM and Kawato M. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7): 1317–1329, 1998.

Wolpert DM, Doya K, and Kawato M. A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences*, 358(1413): 593–602, 2003.

Zwaan RA and Taylor LJ. Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, 135(1): 1–11, 2006.

Please cite this article in press as: Glenberg AM, Gallese V, Action-based language: A theory of language acquisition, comprehension, and production, Cortex (2011), doi:10.1016/j.cortex.2011.04.010