

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

Arthur M. Glenberg<sup>1,2</sup>

Vittorio Gallese<sup>3</sup>

<sup>1</sup> Department of Psychology, Arizona State University

<sup>2</sup> Department of Psychology, University of Wisconsin

<sup>3</sup> Department of Neuroscience, University of Parma

## **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 exapted 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. A collateral benefit of our approach is that it uses motor control to link aspects of attention and working memory to language.

Keywords: Language, embodiment, mirror neurons, HMOSAIC model of action control

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). The phenomenological approach 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, meaning is amplified as it frees 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 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 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.

Data from psychology, psycholinguistics, and neuroscience have demonstrated the importance of action systems to perception (Wilson & Knöblich, 2005), social processes such as mentalizing (Gallese & Goldman, 1998; Gallese 2003a; Sommerville & Decety, 2006), and to language comprehension (Glenberg & Robertson, 1999; Pulvermüller, 1999, 2002, 2005; Gallese 2007, 2008). The action-related account of language and its intersubjective framing suggest 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 MOSAIC and HMOSAIC theories (Haruno, Wolpert, & Kawato, 2002) to produce our theory of action-based language (ABL). We apply the theory to language acquisition, comprehension, and some aspects of production including 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 & Craighero, 2004). This review is followed by a description of the MOSAIC and HMOSAIC models and how we modify them to apply to language phenomena. 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, Zeelenberg, & Barsalou, 2004; Rüschemeyer, Glenberg, Kaschak, & Friederici, under review) and emotional systems (Havas, Glenberg, & Rinck, 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.

### **Language and Action**

The Indexical Hypothesis (Glenberg & 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 the sentences: 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, or ACE. De Vega (in press) 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 had 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).

The Glenberg and Kaschak (2002) results are consistent with an alternative account: Perhaps sentence comprehension does not require any action simulation. However, after a sentence is understood, the meaning is used to prepare for action, and it is at this point that the ACE interaction arises. To test this account, Glenberg, Sato, Cattaneo, Riggio, Palumbo, and Buccino (2008) used transcranial magnetic stimulation (TMS). Sentences were presented

visually using a modified moving window technique. Participants responded with left-hand finger presses to indicate if the sentence was sensible. Note that unlike in Glenberg and Kaschak (2002), there were no movements other than the finger presses. Either at the verb or at the end of the sentence, a TMS pulse was delivered to the area of left motor cortex that controls the right hand, and muscle evoked potentials (MEPs) were recorded from the right-hand opponens pollicis (OP) muscle. (The TMS pulse activates the motor system so that subtle effects produced by linguistic and other stimuli can be measured, see for example, Fadiga et al., 2002).

If comprehension of sentences describing action requires an embodied simulation using the motor system (see Gallese, 2007), then there should be greater activation of OP with transfer sentences than control sentences describing similar events without transfer (e.g., “You and Andy smell the pizza”). Furthermore, if the simulation occurs during sentence processing, then the effect should be found at the verb, not just at the end of the sentence when motor imagery might be operating. Finally, if the motor system is used in processing both concrete and abstract transfer, then the activity in OP should also be found for the abstract transfer sentences relative to their abstract controls. These were exactly the effects we found (see Figure 1): A main effect of transfer sentences compared to no-transfer sentences; a main effect of pulse time (greater activity at the verb than at the end of the sentence); no interaction between any variable and concrete or abstract transfer.

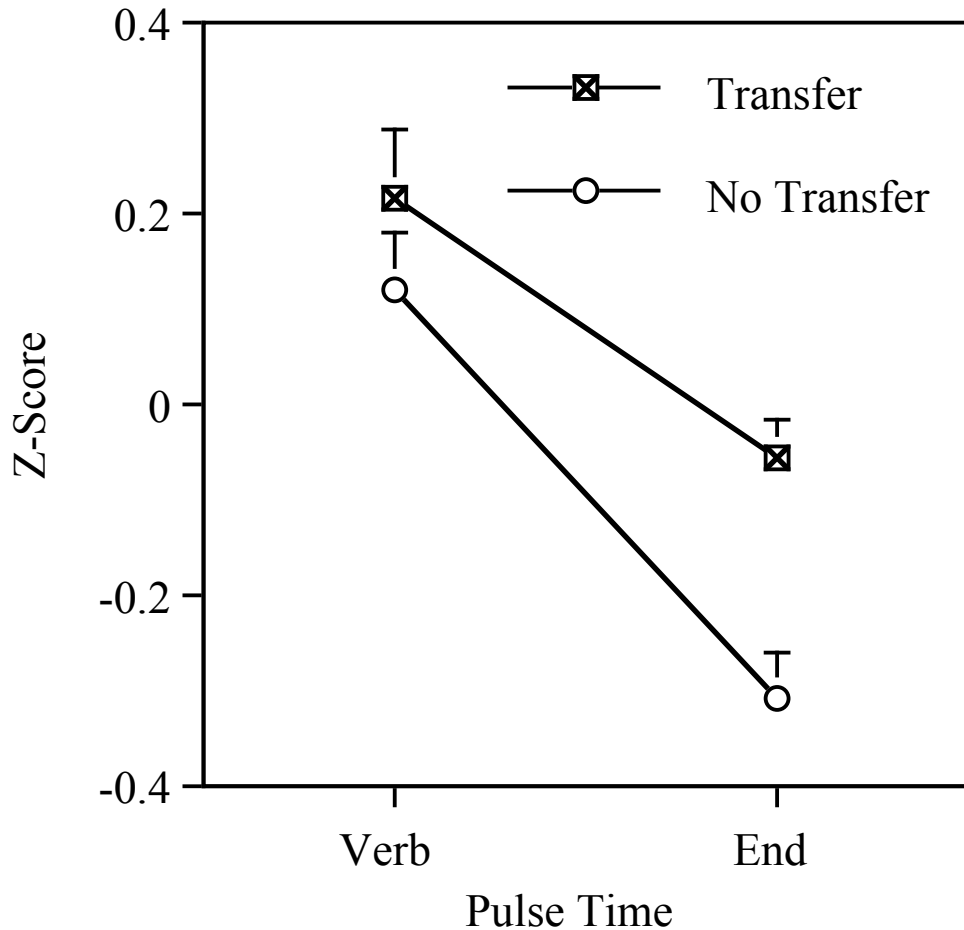


Figure 1. The average z-score of the motor evoked potential produced by transcranial magnetic stimulation over the left motor cortex. The magnetic pulse was delivered while the participant was reading the verb or at the end of a sentence describing either transfer or a static scene. Reprinted with permission from: Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., Buccino, G. (in press). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology*.

The conclusion that language comprehension calls on action systems during the comprehension process is consistent with data reported by Zwaan and Taylor (2006) 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.

These results have been mirrored in the imaging, neuropsychology, and kinematic 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 premotor 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 neurone disorder (ALS) affects comprehension of action verbs more than nouns. Glenberg, Sato, and Cattaneo (2008) 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 language material (Buccino et al., 2005; Boulenger, et al., in press). Furthermore, motor activation occurs very soon during after a stimulus is presented, and only 22 msec after peak activation in auditory temporal areas (Pulvermüller et al. 2003). 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).

### **Neurophysiology of the language-action connection**

The neurophysiological basis for this 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 the monkey premotor cortex (Gallese et al. 1996; Rizzolatti et al. 1996). These 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. A major step forward in the research on the mirror neuron systems (MNS) consisted in the discovery that parietal mirror neurons not only code the goal of an executed/observed motor act, like grasping an object, but they also code the overall action intention (e.g., bringing the grasped object to the mouth or into a container, Fogassi et al. 2005). The MNS maps integrated sequences of goal-related motor acts (grasping, holding, bringing, placing, the different “words” of a “motor vocabulary”, see Rizzolatti et al. 1988) to obtain different and parallel intentional “action sentences,” that is, temporally chained sequences of motor acts properly assembled to accomplish a more distal goal-state. The “motor vocabulary” of grasping-related neurons, by sequential chaining, reorganizes itself as to map the fulfillment of an action intention. The overall action intention (to eat, to place the food or object) is the goal-state of the ultimate goal-related motor act of the chain.

More recently, it has been shown in humans that the observation of actions done with different effectors (hand, foot, 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 led to the notion of a mirror neuron system involving areas in the frontal lobes (notably, Broca’s area) and parietal lobes. The mirror neuron system can also be activated by the typical sound of an action and even when actions are described verbally (for reviews see Rizzolatti & Craighero, 2004; Buccino, Binkofski, & Riggio, 2004; Buccino, Solodkin, & Small, 2006; Gallese 2007, 2008). 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.

One last point on mirror neurons relevant to the development of the ABL theory is a finding of Fadiga, Craighero, Buccino, and Rizzolatti (2002). When their Italian-speaking participants listened to words having a trilled double-r sound, they observed (using a TMS probe over motor cortex) more activation of the tongue muscles than when listening to a double-f sound. These findings have been complemented by a TMS study of Watkins et al. (2003), who showed that listening to and viewing speech gestures enhanced the amplitude of MEPs recorded from lip muscles. A recent fMRI study demonstrated the activation of motor areas devoted to speech production during passive listening to phonemes (Wilson et al. 2004). In addition, Watkins and Paus (2004) showed that during auditory speech perception, the increased size of the MEPs obtained by TMS over the face area of the primary motor cortex correlated with cerebral blood flow increase in Broca's area. This suggests that the activation of the MNS for facial gestures in the premotor cortex facilitates the primary motor cortex output to facial muscles, as evoked by TMS. Finally, Meister, Wilson, Delblieck, Wu, & Iacoboni (2007) demonstrated that repetitive TMS (which temporarily inhibits processing) to ventral premotor areas disrupts speech perception but not color perception or tone perception. Taken together, these results suggest that there are speech mirror neurons (cf. Galantucci, Fowler, & Turvey, 2006), that is, neural structures that respond both to heard and observed speech and speech production.

We will also make use of neurophysiological findings regarding canonical neurons, also found in area F5 of the macaque premotor 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 aimed to it (Rizzolatti, Fogassi and Gallese 2000; Rizzolatti & Luppino, 2001; Gallese 2003b). Unlike mirror neurons, however,

canonical neurons do not respond to observed action. The appearance of a graspable object in the visual space activates the appropriate motor program of the intended type of hand-object interaction. Interestingly enough, it has been shown that observation, silent naming, and imaging the use of man-made objects leads to the activation of the ventral premotor cortex (Perani et al. 1995; Grafton et al. 1997; Chao and Martin 2000; Martin and Chao 2001), a brain cortical region normally considered to be involved in the control of action and not in the representation of objects. The pragmatic properties of these objects (how they are supposed to be handled, manipulated, and used), that is, their *affordances*, appear to make a substantial contribution to their representational content. That explains why the perception of these objects leads to the activation of pre-motor regions of the brain controlling our interactions with those same objects.

In conclusion, there is strong behavioral and neurophysiological evidence pointing to a close connection between the motor system and language. Nonetheless, there are few formal accounts of the connection.

### **The MOSAIC and HMOSAIC theories of action control**

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 motor commands from a representation of goals and context. Thus a controller might produce the commands to control effector trajectory and forces in reaching for and lifting a cup. 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 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 to generate predictions (see Guillery, 2003, for anatomical data consistent with these claims). These predictions are useful for fast correction of movement before sensory feedback can be obtained, for determining if the movement was successful by comparing the prediction to actual sensory feedback, for enhancing perceptual accuracy (Grush, 2004; Wilson & Knöblich, 2005), and the predictions can serve as the basis for attention and working memory. Importantly, 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 cup. Each of the predictors and controllers is implemented as a recurrent neural network (see Wolpert & Kawato 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). Figure 2 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 of the environment, according to the classic cognitive account, amount to a mental model (Johnson-Laird, 1983) of the effects of actions, or expectations regarding how the body and the world will

change as a function of the actions.

In any particular context (e.g., lifting a can when there is uncertainty of 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 is represented by the circle in Figure 2, see Blakemore, Goodbody, & Wolpert, 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. 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.

Figure 2 shows some of the components of the Wolpert and Kawato (1998) theory, however, for simplicity, we have not shown several important aspects of the theory. Thus, in this illustration, a) the sensory/environment predictor should be considered to also compare sensory-motor feedback to the predictions to generate an error signal (hence the double headed arrow between the predictor and the Predictions module), 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.

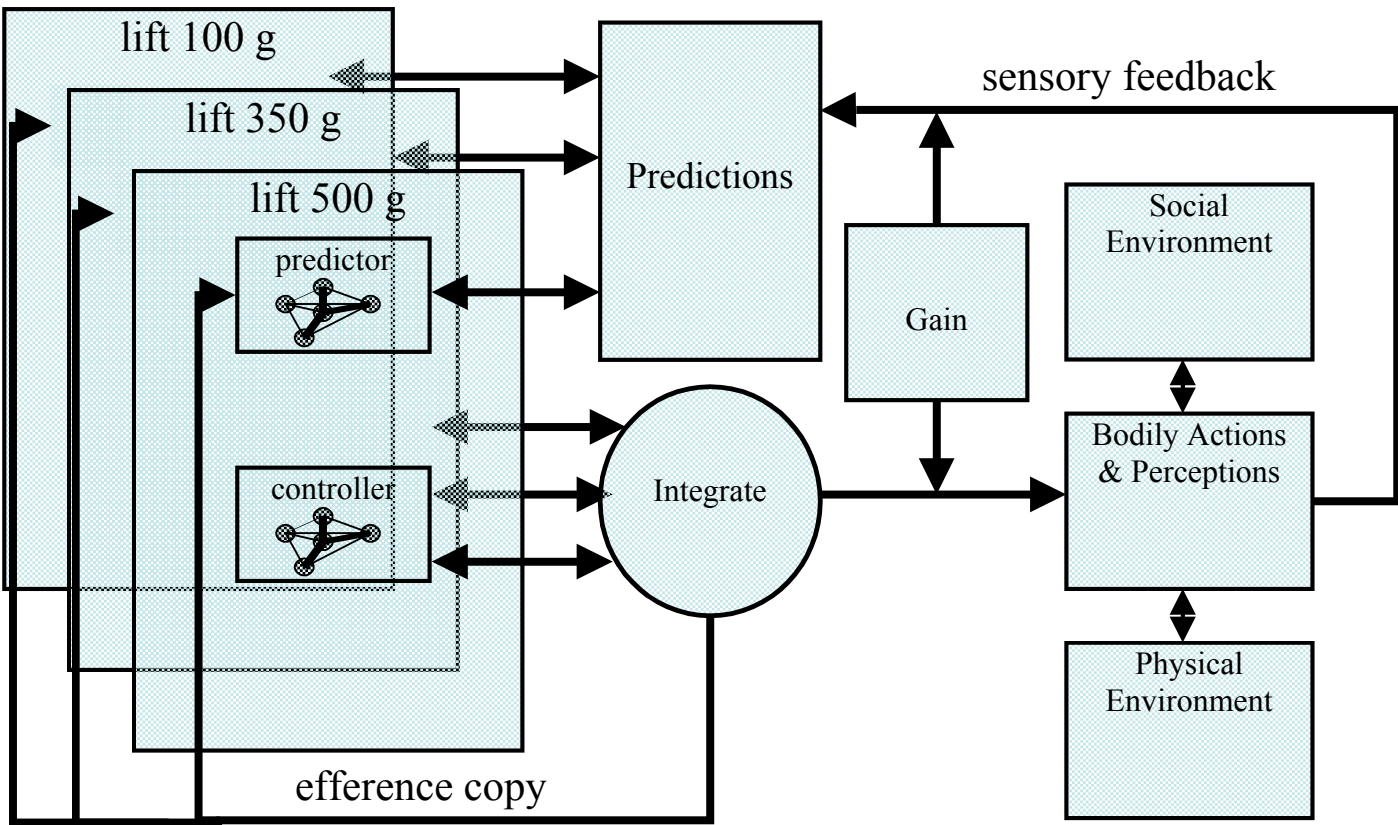


Figure 2. The modified MOSAIC model of movement control for lifting objects of three different weights.

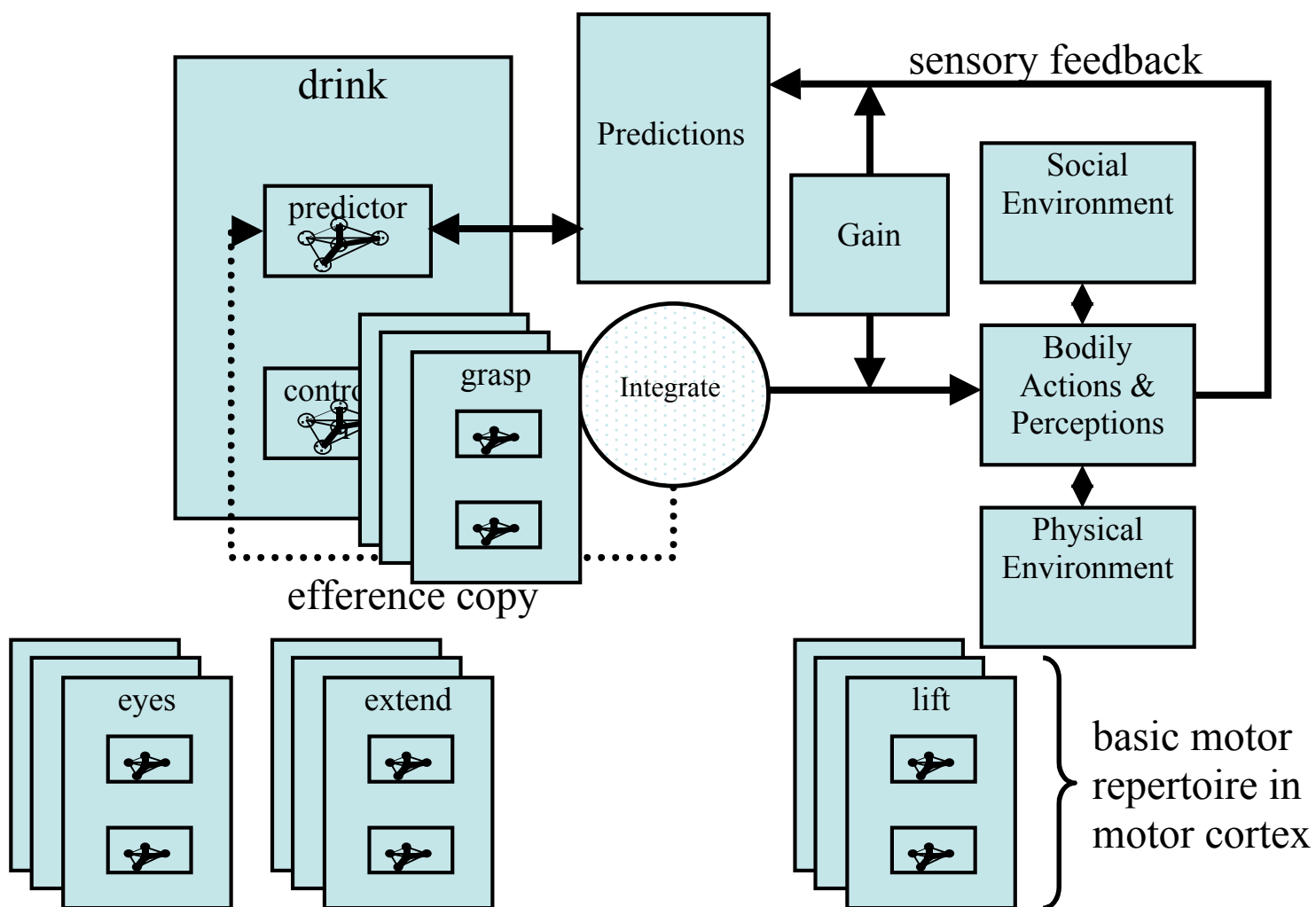
The predictors are generally learned faster than controllers because the output of the predictor and the sensory consequences can be coded using the same parameters (e.g., predicted proprioception and obtained proprioception). The relation between the output of the controller (e.g., a force to a particular 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. The mechanism by which motor commands can be suppressed is discussed next.

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 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 (see also, Glenberg’s, 1997, discussion of suppression of environmental input), 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 predictor is 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 (c.f., Grush, 2004).

Haruno, Wolpert, & Kawato (2003) introduce a hierarchical version of MOSAIC, HMOSAIC, as illustrated in Figure 3. 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 Figure 3 illustrates two levels of hierarchy, in fact, more can be added without any changes in the mathematics underlying HMOSAIC. Haruno, et al. (2003), demonstrate how the higher-level module can learn to select the basic motor acts and learn the appropriate temporal orderings.

Figure 3. The modified HMOSAIC model of action control for drinking.



Whereas the architecture of the lower and upper levels of the hierarchy are 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.) 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).

### **Linking HMOSAIC to language**

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 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 & 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. Consider this description from Wolpert and Kawato (1998, page 1317),

If we consider an example of lifting a can to one's lips, it may be that the desired output at a specific time is a particular acceleration of the hand as judged by sensory feedback. However, the motor command needed to achieve this acceleration will depend on many variables, both internal and external to the body. Clearly, the motor command depends on the state of the arm, i.e., its joint angles and angular velocities. The dynamic equations governing the system also depend on some relatively unvarying parameters, e.g., masses, moments of inertia, and center of masses of the upper arm and forearm. However, these parameters specific to the arm are insufficient to determine the motor command necessary to produce the desired hand acceleration; knowledge of the interactions with the outside world must also be known. For example the geometry and inertial properties of the can will alter the arm's dynamics. More global environmental conditions also contribute to the dynamics, e.g., the orientation of the body relative to gravity and the angular acceleration of the torso about the body. As these parameters are not directly linked to the quantities we can measure about the arm, we will consider them as representing the context of the movement. As the context of the movement alters the input-output relationship of the system under control, the motor command must be tailored to take account of

the current context.

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). A formal quantitative testing of this proposal was recently carried out by Umiltà et al. (2008). In this study, hand-related neurons were recorded from premotor area F5 and the primary motor cortex (area F1) in monkeys trained to grasp objects using two different tools: “normal pliers” and “reverse pliers.” These tools require opposite movements to grasp an object: With normal pliers the hand has to be first opened and then closed, as when grasping is executed with the bare hand, whereas with reverse pliers, the hand has to be first closed and then opened. The use of the two tools enabled the dissociation of neural activity related to hand movement from that related to the goal of the motor act.

All tested neurons in area F5 and half of neurons recorded from the primary motor cortex discharged in relation to the accomplishment of the goal of grasping - when the tool closed on the object - regardless of whether in this phase the hand opened or closed, that is, regardless of the movements employed to accomplish the goal. Goal coding is therefore not only an abstract, mentalist and experience-independent property, but it appears to be a distinctive functional feature upon which the cortical motor system of non-human primates is organized. Goal-directed motor acts are the nuclear building blocks around which action is produced, perceived, and understood.

Thus, the essence of our proposal is that 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).

Figure 4. The ABL model for understanding the verb “to drink.”

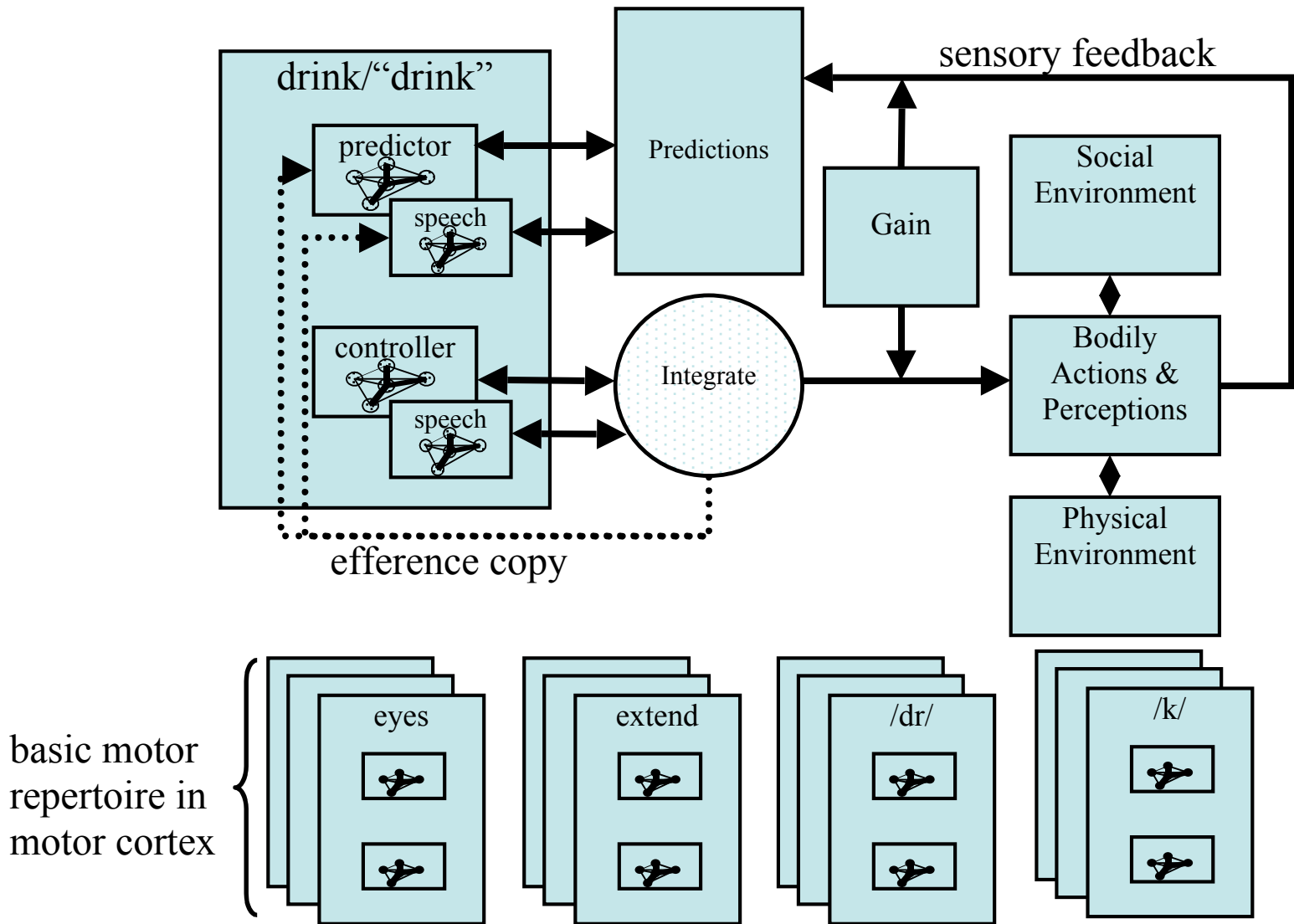


Figure 4 illustrates our version of neural exploitation, namely how the HMOSAIC theory can be linked to language. Along with others (e.g., Fadiga and Gallese 1997; Rizzolatti & Arbib, 1998; Guenther, Ghosh, & Tourville, 2006), we propose that this opportunistic sharing of action control and language was made possible by the development of mirror neurons. Recall that the frontal mirror neuron system overlaps with Broca's area, which controls both the speech articulators and the hand (see Fadiga, Craighero, & Roy, 2006). This overlap is noted in Figure 4 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 Figure 4 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, Fowler, & Turvey, 2006; Guenther, Ghosh, & Tourville, 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 Figure 4 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 & Kaschak, 2002), that is, the controller for

articulation of the word “drink” is associated 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 (sensory-motor feedback) of drinking. Thus, in its essence, knowledge is predictive and grounded in perception and action.

### **Two ABL-specific predictions**

As we discuss later, the ABL theory shares several components with other approaches to language. Nonetheless, it also makes unique predictions, some of which have been tested. Consider first the close connection between meaning and action systems. As illustrated in Figure 4, the meaning of a term (or more generally, a construction, as in Goldberg, 1995), is its associated action HMOSAIC including the predictor and controller. Thus, a strong prediction of the ABL theory is that adapting the action controller will produce a related effect on language. There is some neuropsychological evidence in this regard. For example, Bak and colleagues (e.g., Bak & Hodges, 2003) demonstrate that motor disorders produced by ALS has a selective effect on language.

The first behavioral test of the prediction is Glenberg et al. (2008a). In their experiments, participants first moved, one at a time, 600 beans from a wide-mouthed container to a narrow-mouthed container. This task takes about 20 minutes, a duration that is sufficient to induce measurable neural plasticity. In one condition, the narrow-mouthed container was near to the participant and the wide-mouthed container an arm’s length away. Thus, the movement conformed to a transfer toward action. In another condition, the locations of the containers were reversed, so that the movement conformed to a transfer away movement. After the bean task, participants read sentences presented on a computer screen and determined if they were sensible.

The participants responded by pressing keys on the computer keyboard, which required no movement other than depression of the index fingers, and the time needed to make the sensibility judgment was measured. The critical sentences described transfer of objects and information either toward the participant or away from the participant. The important finding was that direction of bean movement interacted with described direction of transfer. That is, adapting the action system (with the bean task) produced an effect on language processing. (A control experiment ruled out the alternative interpretation that the bean task adapted a symbolic representation of the concepts toward and away.)

A second set of experiments (Scorolli, Borghi, & Glenberg, in press) tested in a language context a feature of the MOSAIC architecture demonstrated by Hamilton, Wolpert, and Frith (2004). In Hamilton et al., observers lifted one weight while simultaneously judging the weight being lifted in a video tape. When the weight being lifted was similar to the weight being judged, there was a type of repulsion effect: When the observer lifted a weight lighter than that being observed in the video, observers judgments of the weight observed tended to be too heavy. Similarly, when the observer lifted a weight heavier than that observed, the judgments tended to be too light. Hamilton et al. explained this effect using the integration mechanism of the MOSAIC model and the assumption that if a module is being used in one task then it is unavailable to another task. Thus, they proposed that the judgment of the weight being observed is made by integrating the weights corresponding to modules primed by the video. However, the module being used by the observer to control literal lifting of a weight is unavailable for the judgment task. Consequently, if a light weight is being lifted, the judgments are biased on the heavy side, and if a heavy weight is being lifted the judgments are biased on the light side.

Scorolli, et al. (in press) used the Hamilton et al. (2004) logic in a language task.

Participants first heard a sentence describing the lifting of a light (e.g., pillow) or heavy (e.g., toolbox) object, and then the participant literally lifted one of two identical looking boxes. One of the boxes was heavy and the other light. Scorolli et al. demonstrated two effects on kinematic parameters during the lift of the box. First, early in the experiment, the kinematic parameters revealed an expectancy-like effect. After reading a sentence about a light object, participants tended to exert too little force and after reading a sentence about a heavy object, the participants tended to exert too much force. Second, and later in the experiment after participants came to realize that the weight described in the sentence was uncorrelated with the actual weight lifted, the repulsion pattern emerged. That is, reading about lifting a light object slowed the lifting of light boxes and speeded the lifting of heavy boxes; the reverse was found after reading of sentences describing the lifting of heavy objects. This sort of repulsion effect provides strong support for the ABL theory

### **Learning nouns**

Consider how a verbal label can be associated with the appropriate action module during language acquisition. In this analysis, we assume that the infant has already developed some skill in how to interact with objects (e.g., a baby bottle). We will present evidence in this regard later. 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 is 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 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 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,” the infant’s speech-mirror neuron system 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 premotor theory of attention (e.g., Awh, Armstrong, & Moore, 2006; Craighero et al., 1999; 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. 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 Figure 4 would control an infant’s behavior upon hearing a noun such as “bottle.” Hearing the noun activates speech mirror neurons for that 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 found, 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 to 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 & Kamide, 2004; Chambers, Tanenhaus, & Magnuson, 2004; see Thothathiri & 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., 2005) and work in brain imaging (e.g., Hauk, Johnsrude, & Pulvermüller, 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 & 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.

This account is closely related to Barsalou's (1999) notion of a simulator as the basis for concepts (see also Gallese 2003b, Gallese and Lakoff 2005). A simulator links neural activity associated with the attended components of an object such as a baby bottle. Activating the simulator, that is, contacting the conceptual information, generates a context-specific simulation. In the ABL theory, that context-specific simulation consists of the predictions generated by the predictor based on the efference copy of how to act on the object in the current context.

### **Learning Verbs**

A similar account can be given for verb learning. For example, consider how an infant who knows how to drink (that is the HMOSAIC module for controlling drinking is in place), might learn the verb "to drink." Suppose that the infant is drinking from a bottle. The parent might say, "good drinking!" The speech mirror system is activated by the parent's speech and a Hebbian learning process begins to establish connections between the action control for drinking and 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 mirror neuron system 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 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 a 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 PET 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. A recent analysis of the MacArthur Child Development Inventory (CDI, Angrave & 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, 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  $\text{Speech age} = .61 (\text{Gesture age}) + 17$  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, Woodward, & Brune, 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 & Tomasello, 2002; Goldin-Meadow, Seligman, & Gelman, 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, 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.

Maouene, Hidaka, & Smith (in press) have reported a remarkable finding confirming the ABL prediction that toddlers learn how to make particular actions before learning the associated verbs. Maouene et al. looked at the body parts associated with the 101 earliest learned verbs. They found that the verbs were learned in clusters related to body parts. Thus, verbs related to mouth actions tended to be learned first, the verbs that were next were associated with hand and arm actions, and finally verbs not strongly associated with any one body part were learned.

In natural situations, nouns are learned faster than verbs (e.g., D. 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, goal-directed actions often involve 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.

### **Learning determiners**

When children are first learning determiners, the use is context dependent (Tomasello, 2000). For example, “a” will occur with some nouns, but not others, and “the” may occur with a different set of nouns. It is also the case that correct usage of “the” and “a” is a late achievement compared to the acquisition of “this” and “that” (Moldyanova, 2006). Here, we discuss only the ABL theory account of “this” and “that.”

Linguistically, “this” can function as a determiner (e.g., “Grab this bottle”) or as a pronoun (e.g., “Grab this”). It always has the same action function, however, namely, attend to the speaker’s peri-personal space, such as the speaker’s hands. Thus the ease of learning “this” can be accounted for by several factors. The first is a consistent association with an action, namely, directing attention to hands. Second, the speaker is likely to call attention to his hands by presenting them, that is, lifting them and displaying them to the listener. This display will activate the listener’s hand-related mirror neurons. Finally, the common representational

substrate between speech and hand control in Broca's region sets the stage for Hebbian learning between controller for the production of the word "this" and the action of attending to the speaker's hands.

A similar account holds for the learning of "that." Functionally, "that" (used as a determiner or pronoun, but not when used as a relativizer) means that the listener should attend to whatever the speaker is attending to, and the object of the speaker's attention is often indicated by pointing or gaze direction. Thus, learning of "that" presupposes that the listener can follow a point and eye gaze.

### **Learning verb-argument constructions**

Our hypothesis regarding larger syntactic units has three parts. First, the basic function of syntax is to combine linguistic components in a way that produces a sensible outcome. Second, the basic function of motor control is to combine movements in a way that produces an effective action, that is, action that succeeds in reaching a goal. These goal-directed actions require hierarchical control. Third, syntax emerges from combining (using responsibilities) the descriptions (speech controllers) of modules for hierarchically controlled action. The effective action produced by the hierarchical control of action becomes the meaning of the syntactic construction. We first review the idea of meaning-bearing verb-argument constructions (e.g., Goldberg, 1995) and how they can coerce new interpretations of other linguistic units. We then demonstrate how the ABL theory accounts for the learning of verb-argument constructions and coercion of meaning.

Construction grammarians (e.g., Goldberg, 1995) propose that linguistic knowledge consists of a collection of constructions, where each construction consists of conjoining an arbitrary symbol<sup>1</sup> and a meaning. Examples at the morphemic level include "s+plural" and "cup

+ small hand-held object used for holding liquids.”

Verb-argument constructions relate the syntactic form of a simple sentence (the symbol) to a meaning. For example, the caused-motion construction relates the structure, “Subject-Verb-N1-Location phrase” to the meaning, “by means of the Verb, the subject causes N1 to move to the location.” An example is, “Art flicked the lint onto the floor.” Note that the claim is that part of the meaning comes from the form of the sentence, not simply from the meanings of the individual words. To illustrate this claim, consider “Vittorio sneezed the foam off the cappuccino,” a variant of one of Goldberg’s examples. Most people find this sentence to be grammatically acceptable and sensible. However, “to sneeze” is an intransitive verb, and thus should not be able to take a direct object such as “the foam.” Also, few people (or dictionaries) would ever include “to cause motion” as part of the definition of “to sneeze.” Goldberg uses examples such as this one to support the claim that the construction carries part of the meaning, and that the construction can coerce that meaning onto the verb.

Kaschak and Glenberg (2000) demonstrated a constraint on coercion using the double-object construction. The symbolic pole of the double-object construction is “Subject-Verb-N1-N2” and the associated meaning is “The subject transfers N2 (the object) to N1 (the recipient),” as in “The mailman handed Gloria the letters.” Kaschak and Glenberg studied coercion by inserting innovative denominal verbs into double-object constructions. Denominal verbs are verbs derived from nouns, such as “to bottle” and “to bicycle.” Innovative denominal verbs are not part of standard English, instead they are made up and interpreted on the fly. As an example, consider the following scenario and two endings:

A man with a broken foot was sitting on a park bench peeling his hard-boiled egg. A soccer ball rolled up to him from a nearby game being played by some

young girls. He crutched the girls the ball./ He egg-shelled the girls the ball.

Most people find the innovative denominal verb “to crutch” acceptable (Kaschak & Glenberg, 2000), and they interpret it as “to transfer something using a crutch.” That is, the double-object construction coerces a transfer meaning onto the verb. However, most people will reject the “egg-shell” sentence as ungrammatical and nonsense. Kaschak and Glenberg argue that coercion is successful when the noun underlying the denominal verb has the right affordances to effect the transfer of N2 to N1, as a crutch can be used to hit a soccer ball so that it is transferred from one person to another. However, when the object underlying the verb does not afford the transfer of N2 to N1 (as egg shells do not afford transfer of a soccer ball), then the coercion fails.

We propose that the mechanisms illustrated in Figure 4, with one additional assumption, are sufficient to learn verb-argument constructions that can coerce contextually appropriate meanings. The assumption is that whenever a predictor generates a high probability prediction that is disconfirmed, the modules that generated the prediction lose control over behavior, and control defaults to other modules (e.g., modules for attending to this unpredicted event).

As before, we assume that the child already knows how to interact with various objects (e.g., a bottle, a cup, a cookie), the names of those objects, and how to engage in some complex activities such as giving. Giving is a high-level module that includes a) attending to a recipient location, 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.

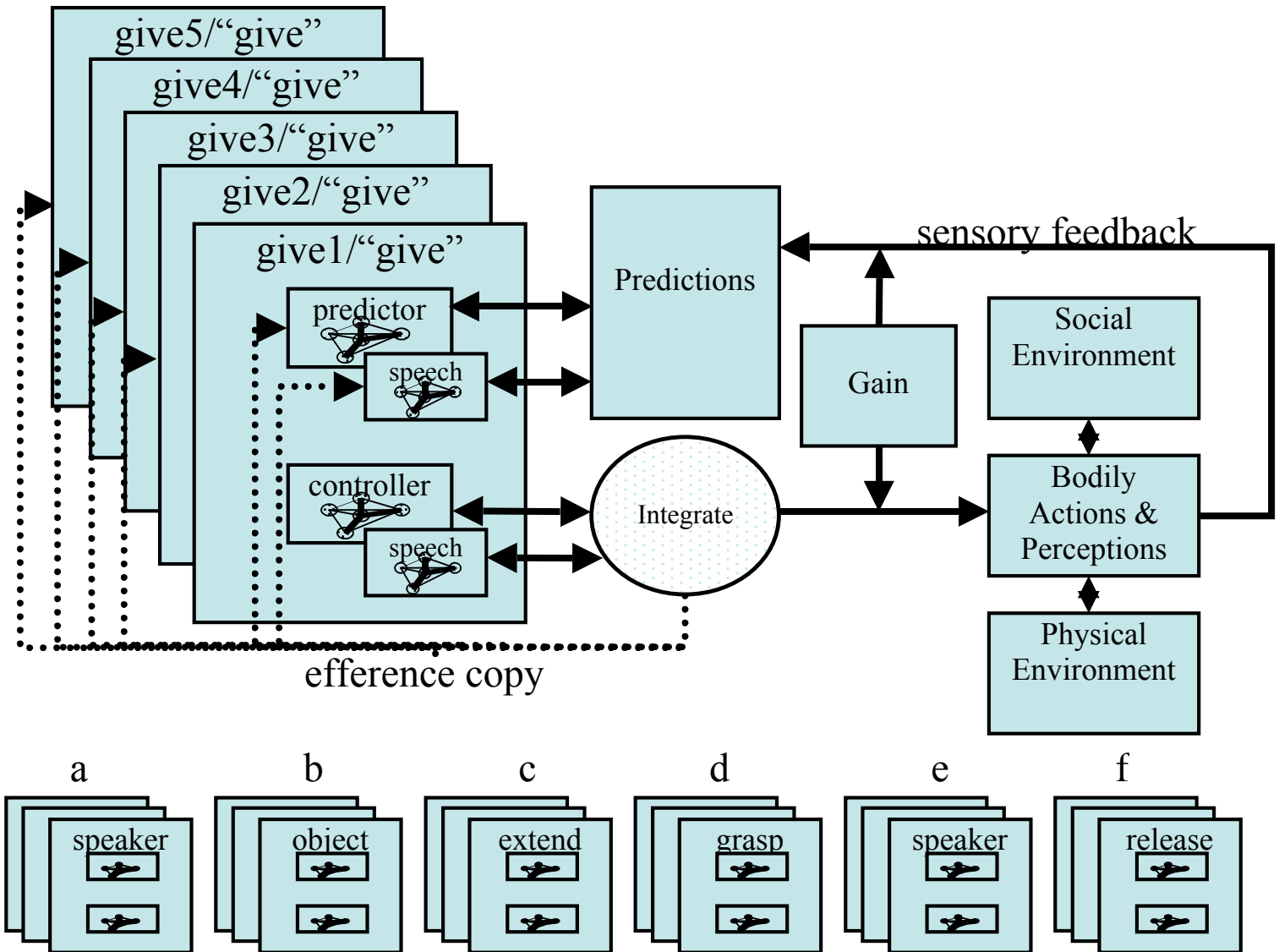
Imagine that the child hears his father say, “Give Dada the cup.” Upon hearing “Dada” the child will attend to Dada and prepare to interact with him. Upon hearing “the cup,” the child will attend to the cup and perhaps grasp the cup. But suppose that the child has not associated

the word “give” with the 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 mirror neuron system 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.

With repetition, the child learns the association between a specific higher-order module for giving a cup to Dada and the verbal signal “give-dada-the-cup.” The higher-level controller 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 Figure 5.

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 Figure 5. Similar disconfirmations of high probability predictions may result from hearing “Give Lavinnia the spoon” (resulting in *give3*), “Give Dada the bottle” (resulting in *give4*), and “Give Dada the spoon” (resulting in *give5*).

Figure 5. The ABL model for understanding five instances of double object sentences using the verb “to give.”



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 (on the comprehension side) to Tomasello’s (2003) 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. Although the infant has learned nothing like a grammatical rule, we propose that having learned a sufficient number of such give modules, the child can generalize and respond to many more sentences than those already learned (see below). That is, *behavior* consistent with having learned an abstract structure or rule emerges from the operation of the HMOSIAC theory (c.f., Hintzman, 1986). As we will demonstrate next, however, this generalization is sensitive to constraints on action.

As an example of generalization, consider the child hearing the novel sentence, “Give Dada the cookie.” 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 by the responsibilities. The responsibility-weighted efference copy that results from combining the outputs of 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” “bottle” 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 cookie,” there is a disconfirmation of low-probability predictions rather than a high-probability prediction, and control of behavior by the *give1*, *give4*, and *give5* modules is not disrupted. Instead, upon hearing “the cookie” the infant can execute the actions already associated with “the cookie,” namely, locate the cookie, extend the arm, and grasp the cookie. These three actions correspond (in kind) to actions (b) – (d) predicted by the *give1*, *give4*, and *give5* predictors, thus updating the *give* controllers and predictors. Now, the *give* controllers orient the child back to the last located person and extend the arm (e), and hand off the cookie (f). On this account, the child behaves appropriately, and it appears as if the child has used an abstract grammatical construction. Our claim, however, is that the appropriate behavior has emerged from the operation of the HMOSAIC mechanism for integrating particular representations, not from executing an abstract rule.

It should be noted that the generalization mechanism is not a simple blending. For example, it is not that the meaning of “cookie” is a blend of meanings of “cookie” “cup” “spoon” and so on. Instead, the ability to respond correctly to the sentence depends on having a well-articulated understanding of “cookie,” namely how to act on that sort of object. The generalization comes about because how one literally gives cookies 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 cookies.

This claim makes an important prediction that was recently confirmed by Bannard and Mathews (2008), namely that during language acquisition, children store in memory particular word sequences (that will later support generalization). In their research, children were asked to repeat four-word sequences that were played for them. These sequences were either frequent (e.g., “a drink of milk,” “when we go out”) or less frequent (e.g., “a drink of tea,” “when we go in”) in child-directed speech, although the frequencies of the fourth words (e.g., “milk” and “tea”) were matched. Bannard and Mathews determined 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.

Three characteristics of this instance-based generalization need to be noted. First, the theory describes a type of mesh of affordances as discussed in Glenberg and Robertson (2000). According to Glenberg and Robertson (see also Kaschak & Glenberg, 2000), sentences are only sensible when the various object affordances can be integrated, as guided by constructions, to result in goal-directed action. Here, the child’s understanding of how to interact with the cookie ( 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 were instead “Give Fido the cup,” the child and the model would reject the sentence as nonsense because although Fido can be located, Fido cannot take the cup. That is, at this point in the child’s learning about “give,” Fido does 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 the idea,” because ideas cannot be literally grasped and handed off. Thus, the generalization mechanism will not automatically produce a representation. Instead, generalization is limited (at this stage of acquisition) by constraints on action.

Second, the theory is consistent with Goldinger’s (1998) episodic theory of lexical access and Kaschak and Glenberg’s (2004) observations of episodic effects in verb-argument learning. In developing his theory, Goldinger discusses data demonstrating how idiosyncratic aspects of speech, such as details of the heard voice and ambient noise, affect speech perception. That is, perceptual details of experience are stored in memory and contribute to the perception and production of speech. Episodic effects such as these are expected at the lower-level modules that produce speech because of the learning done with each of the idiosyncratic inputs (cf., Guenther et al., 2006).

Kaschak and Glenberg (2004) tracked the learning of a grammatical construction novel for many American English speakers, the needs-construction, as in “The dog needs walked.” Participants were able to learn the construction after hearing just a few examples in context. Furthermore, Kaschak and Glenberg demonstrated that the particular episodic processing events accompanying those examples affected later use of the construction.

The final characteristic of note is that the model implements behavior that is at least partially consistent with a phrase structure grammar (PSG). That is, phrases such as “the cookie” can be embedded within the *give* hierarchical control structure. The ability to embed structures is thought to be characteristic of human language and an accomplishment beyond the ken of other primates (e.g., Fitch & Hauser, 2004).<sup>2</sup> Indeed, Friederici et al. (2006) have reported data demonstrating that processing such structures may call upon Broca’s area, whereas the processing of less complex grammars does not.

We prefer a different interpretation of the Fitch & Hauser (2004) and Friederici et al. (2006) data, however. Keep in mind that people have difficulty understanding sentences with more than one center-embedded clause. Thus, it is unlikely that people are implementing a true PSG with unlimited embedding. In our theory, the degree to which people can process such structures is directly related to the amount of practice with similar examples. Thus, the Friederici et al. finding may indicate that Broca's area deals with hierarchical action control with several levels in the hierarchy, whereas premotor areas deal with hierarchical control with fewer levels (see also Fiebach and Schubotz, 2006; Van Schie, Toni, & Bekkering, 2006). The ability to create additional levels of hierarchical control over action leads to a) more complicated actions, b) predictions farther into the future, and as we claim here c) structured language (see Gallese 2007, 2008).

The theory implies several developmental trends in regard to learning verb-argument constructions. First, the learning of new grammatical structures results from a base of simpler structures rather than the direct learning of the structure (Bates & Goodman, 1997; Tomasello, 2003). Second, at each level in the hierarchy, there exist multiple modules, rather than a generalized structure (Goldinger, 1998; Kaschak & Glenberg, 2004). Third, training should allow the relatively easy learning of new grammatical structures if and only if those structures can be mapped onto doable, hierarchical control of action (Goldberg, Casenhiser, & Sethuraman 2004).

### **Coercion and the understanding abstract language**

As noted above, the structures in Figure 5 would balk at the sentence "Give Lavinnia the idea" because ideas cannot be literally handed off. Eventually, however, children learn to deal with these sorts of abstract notions. Here we propose how at least some abstract language

related to transfer can arise from action-based understanding of “give.” We take up the notion of abstract language again in the discussion of perspective.

Consider the following sentences describing abstract transfer:

Delegate the responsibilities to Anna

Present the honor to Leonardo.

Convey the information to Adam.

In each of these cases, the verb has a transfer function that is closely related to “give.” In fact, “Give” can be substituted for all of the verbs without much change in meaning. Our proposal (see also, Goldberg, 1999) is that in these cases (and many others), children learn the meaning of the more abstract verbs of transfer in terms of giving. A child may be told explicitly that “delegate” means “give,” or the child might induce that “delegate” is a type of giving from the ditransitive structure of the sentence (Goldberg, 1995; Kaschak & Glenberg, 2000). In either case, the transfer modules (e.g., *give1*, *give2*, etc.) are used as templates for understanding the abstract verbs. This proposal has several implications. First, people should envision that nominally abstract actions such as delegating responsibilities, presenting honors, and conveying information involve hand and arm motion.<sup>3</sup>

Second, these sorts of sentences should produce interference when the described actions conflict with actual actions people might be performing, as found by Glenberg and Kaschak (2002). Third, in reading these sentences, motor structures controlling hand and arm movements should be activated, as found by Glenberg et al. (2008a, b).

### **Comprehension: language simulation using the motor system**

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

& Lakoff, 2005; Glenberg & Robertson, 2000; Zwaan & Taylor, 2006; Gallese 2007, 2008). 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, Meyer, & Lindem, 1986; Johnson-Laird, 1983).

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.<sup>4</sup>

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 specific predictions of the double-object modules are disconfirmed. Nonetheless, as discussed in the section *Learning verb-argument constructions*, 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 proposed sequence is consistent with the data reported by Glenberg et al. (in press). Using TMS, they were able to demonstrate activation of systems controlling the OP muscle while participants read sentences describing transfer of objects. The proposed sequence is also consistent with the analyses of Pickering and Garrod (2007) that production mechanisms are used to make predictions during language comprehension. Finally, the proposed sequence 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 such as giving.

### **Language production and gesture**

We do not attempt to describe a full model of language production. Nonetheless, the ABL theory provides insight into a number of salient findings in the production literature including gesture and structural priming.

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 & Golden-Meadow, 2001). Gesture can facilitate production (Kraus, 1998) and comprehension (Golden-Meadow, Nusbaum, & Wagner, 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 voice spectra 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 coded as a single signal by a unique communication system within the premotor cortex.

The involvement of premotor Broca's area in translating the representations of communicative arm gestures into mouth articulation gestures was recently 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 is part of the MNS, 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 premotor BA 44, “vehicle” and “content” of social communication are tightly interconnected (Gallese 2007, 2008).

ABL theory provides a computational account of the intimate relation between speech and action, as illustrated in Figures 3-5. Thus, speaking a word will activate the corresponding action, and it is only through active inhibition (by gain control) that the overt action is not taken. We suggest that gesture arises through the incomplete inhibition of action when speaking (Hostetter & Alibali, in press). That is, 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 exhibited, 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.

Figure 3 also makes clear why action can facilitate retrieval of pronunciation (Kraus, 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 neuron system will resonate, thereby activating the listener’s controllers and predictors for related concepts. This priming can aid speech perception (Galantucci, Fowler, & Turvey, 2006), as well disambiguate meaning (Kelly, Barr, Church, & Lynch, 1999).

Many gestures are schematic, and in the absence of speech they may not be readily interpretable. The ABL theory provides an explanation for how gestures can be both highly schematic and yet useful. McNeill and Duncan (2000) describe catchments as gestures with recurring form features such as a particular hand shape. For example, imagine relating a story about a car, and using a flat hand, palm down, swept from near the chest to farther away to indicate the car speeding away. Clearly, the gesture itself shares little with an actual car, so how can it be used repeatedly to refer to the car speeding away? We suggest that the gesture takes on a temporary meaning because the action control system is easily recalibrated. On hearing “the car sped away,” the listener’s predictors generate predictions of sensory information such as the visual image of a receding car, the reduction in noise, and perhaps a Doppler shift. Given that the gesture is made in close contiguity to the verbal description, the gesture becomes part of the sensory feedback. With learning, the listener’s predictor for “the car sped away” is recalibrated so that the gesture is predicted. Thus, the gesture becomes part of the meaning of “the car sped away.” Later invocations of the same gesture, the catchment, will now partially activate “the car sped away” module so that it can be used in interpreting the communication even if the words “the car sped away” are not uttered.

This explanation makes several predictions. First, when schematic gestures are first used, they are in the accompaniment of more complete verbal (or other) descriptions. Second, as the gesture is repeated, the amount of verbal support is reduced. Third, imagine the following scenario. A naïve listener watches a video tape of the description of the speeding car. However, the video (but not the audio) is blanked during the initial performance of the gesture. In this case, the naïve listener should derive little benefit from later repetitions of the catchment (in the reduced verbal contexts) because it will not have been defined for this listener.

The explication of gestures discussed here is closely aligned with “simulated action” account of gestures proposed by Hostetter and Alibali (in press). In brief, they propose that gestures arise because language producers engage in simulation during production, and they review a large amount of data consistent with this idea. They also propose three factors that determine the degree to which gestures will be made: the extent to which the content of the verbal message reflects action, the individual’s gesture threshold, and the simultaneous engagement of motor system for speaking. These factors are consistent with the ABL theory. For example, the gesture threshold corresponds to gain control: If the gain is low, few overt actions or gestures will be made. The operation of gain control also addresses the third factor identified by Hostetter and Alibali, namely that gesture are more likely when the motor system is also engaged for speaking. Note that if gain control is too low, then speech itself will be interrupted.

### **Structural priming**

Chang, Dell, and Bock (2006) suggest that structural priming is “One of the most convincing sources of evidence that people make use of abstract syntactic representations while speaking...” Consequently, it is important to demonstrate how the ABL theory, which does not have explicit abstract syntactic representations, can account for these data. Consider how a typical structural priming experiment is conducted (e.g., Bock & Loebell, 1990). The participant is presented with a mixture of sentences and pictures. When a sentence is heard (prime trials), the participant is to repeat it verbatim. When a picture is presented (target trials), the participant is to describe it. Thus, on a prime trial, participants might repeat a sentence such as, “The girl gave the man the book.” On a target trial, the participant might see a picture depicting a pirate, a crown, and a queen. The question of interest is whether the participant will

complete the description as, “The pirate gives the queen the crown” (which matches the double-object syntax of the prime) or “The pirate gives the crown to the queen” (the syntactically non-matching prepositional form). The basic finding is that participants are more likely to produce a description that matches the prime than the alternative form.

Chang et al. (2006) review a number of features of structural priming. These include a) that it occurs without lexical or conceptual repetition; b) it does not depend on prosodic pattern; c) it persists over time and the processing of other sentences; and d) it is insensitive to some, but not all, similarity in meaning between the prime and target sentences. This latter point is crucial to our interpretation of structural priming, and so it is worth examining in some detail.

The claim regarding meaning similarity uses particular notions of meaning, namely thematic roles, argument structures, and transitivity of the sentences. For example, consider the data from Bock and Loebell (1990, Experiment 2), which Chang et al. suggest provides a strong test of the independence of meaning and structural priming. A passive prime might be, “The 747 was alerted by the airport’s control tower.” A locative prime might be, “The 747 was landing by the airport’s control tower.” Although the passive prime and locative prime appear to share the same surface structure (noun phrase, verb phrase, by-phrase), they differ quite a bit in regard to putative meaning-related features such as thematic roles. Thus in the passive, the verb (“was alerted”) is transitive, whereas in the locative the verb (“was landing”) is intransitive; in the passive, the subject (“The 747”) is a patient, whereas in the locative, the subject (“The 747”) is an agent; in the passive, the by-phrases (“by the airport’s control tower”) is an agent, whereas in the locative, the by-phrase (“by the airport’s control tower”) is a location. Nonetheless, there was equivalent priming from the passive to the passive and from the locative to the passive. That is, the differences in thematic roles did not seem to affect structural priming.

Now consider a case in which elements of meaning appear to affect structural priming, as reported by Chang, Bock, and Goldberg (2003). This study used theme-locative structures such as “The man sprayed water on the wall,” in which the theme (water) is presented before the location (the wall), and locative-theme structures such as, “The man sprayed the wall with water.” Again, the surface structures match (noun phrase, verb phrase, prepositional phrase), but the order of the thematic roles differ. In this experiment, the theme-locative primes produced more priming of theme-locative targets than did the locative-theme primes. That is, the difference in the ordering of the thematic roles affected priming.

The ABL theory account for these effects is straightforward: they reflect planned eye movements and shifts in attention (Rizzolatti et al, 1987). As we describe later, shifts in attention correspond to the perspective taken on the situation described by the language. On the ABL theory, the proposed meaning of an object consists of first moving the eyes to locate the object and then extracting the affordances of the object represented by canonical neurons. These processes occur whether one is literally examining a scene or picture or constructing a mental model of the scene. Thus, consider the planned eye movements in comprehending the theme-locative, “The man sprayed water on the wall.” Upon hearing “The man,” an eye movement is planned to a likely location, such as directly in front of the participant. That is, in a spatial mental model of the sentence, the man is located directly in front of the listener. Upon hearing “sprayed water,” a saccade will be planned to a nearby location, perhaps to the right, to locate water emerging from an inferred hose in the man’s hand. Finally, upon hearing, “on the wall” a second saccade is planned in a similar, rightward direction. Thus, the pattern of planned eye movements driven by comprehension of the prime sentence is front->right->right. Planning these movements will lead to a learned increment in the probabilities of these movements

generated by the higher-level module.

In contrast, the locative-theme construction, “The man sprayed the wall with water” requires planning to move the eyes to a central location, then to the right (for the wall), and then back toward the middle (to locate the water emerging from an inferred hose). This pattern of planned eye movements is front->right->left. Given the increment in learning produced by the prime (either front->right->right or front->right->left) there will be a corresponding bias in the processing of the target. That is, when looking at a target picture, or when constructing a mental model from language (as in Chang et al.,) the sequence in which the target elements are encoded is biased toward the sequence induced by the prime.

On this account, consider the Bock and Loebell (1990) finding that a passive target (e.g., “The construction worker was hit by a bulldozer,” front->maintain attention to that location) is primed equally well by a passive prime (e.g., “The 747 was alerted by the airport’s control tower,” front->right scan, or perhaps front->maintain attention to that location) and a locative prime (e.g., “The 747 was landing by the airport’s control tower,” front->right scan or perhaps front->maintain attention to that location). The same scan pattern is used for the two types of primes, suggesting similar amounts of priming due to learning similar scan patterns.

Finally, consider the original finding that double-object primes such as “The girl gives the man the book” (front->right->left scan) elicit more double-object picture descriptions, such as “The pirate gives the queen the crown” (front->right->left scan) than prepositional descriptions such as, “The pirate gives the crown to the queen” (front->left->left scan). Here the double-object requires a different planned scan pattern than the prepositional. Thus, to the extent that processing the prime increments the learning of a scan pattern (which corresponds to an attention-driven perspective), the description of the picture is likely to use the same structural

form as used in the prime.<sup>5</sup>

Thomas and Lleras (2007) report data consistent with the scan-based interpretation of priming, albeit in a different paradigm. While solving the Duncker radiation problem, participants looked at a picture illustrating a tumor (a filled circle) surrounded by a circle representing the skin. A secondary task required participants to detect the appearance of a digit within a sequence of letters. The independent variable of interest was the location of the letters and digits. In one condition, the locations alternated between one location outside of the skin and one location near the tumor. Participants tracking these locations to detect the digit target will move their eyes repeatedly across the skin at one location. In another condition, multiple locations outside of the skin were used along with the location near the tumor. Tracking these locations requires the eyes to move repeatedly across the skin but at different locations, much like the path of multiple lasers needed to destroy the tumor without damaging the skin. Amazingly, 50% of the participants in the second condition solved the problem compared to only 19% in the first condition. Thus, controlling eye-movements, which we suggest induces a perspective on the situation, controlled the generation of a high-level solution strategy, much like the ABL theory posits that controlling eye movements will control the generation of speech conforming to one syntactic pattern or another.

### **Discussion**

This discussion briefly addresses the following issues: Non-motor processes; the implementation of symbols and perspective; the relation of the ABL theory to attention and working memory; similarities to other accounts; and a consideration of several remaining issues in development of the theory.

Before moving to those discussions, however, a common misunderstanding of our

position must be addressed. That misunderstanding is that we are attempting to update a thoroughly discredited Skinnerian account of language. Of course, there are similarities between ABL and Skinnerian accounts. We are focused on behavior and the control of behavior. Also, our invocation of the motor system seems similar to the notion of “responding” in an “S-R” psychology. Finally, at its core, 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 referent. 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 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. Also, as discussed below, prediction forms the very basis of conceptual knowledge. Third, the higher-order modules in ABL are explicitly symbolic, although the symbols are grounded, not arbitrary. Because of these important differences, ABL goes substantially beyond a simple associative system.

### **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, p15).<sup>6</sup> 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 exapted to control situation-specific language (see 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. 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 representing 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, in press) 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 (2005) linking processing in superior temporal areas with mirror neuron systems in parietal and frontal areas.

The predictions generated by the predictor models also link to other neural systems. Although a prediction can be modeled as a vector of probabilities, 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 sensory-motor 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, motor and pre-

motor cortex used in taking action, and so on. 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 example, Mandler (2008) proposes that the earliest concepts are redescrptions of innately salient spatial information such as the difference between biological and non-biological movement. The redescription makes use of an attentive mechanism called 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, p 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. This notion of state of the world is congruent with 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 the properties of concepts, which Mandler takes to be “declarative knowledge about object kinds and events that is potentially accessible to conscious thought” (p 207). Although these properties are likely to be components of human concepts, it is not clear how these properties imbue concepts with their most important characteristic, namely that they are used in thinking. The notion of prediction does just that. It allows one to go from 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 simulation and conceptualization.

### **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 accounts for the fact that language is productive (an infinite number of sentences can be generated from a finite number of elements) and compositional. Because symbols are inherently abstract, at first glance it would appear that a sensori-motor account of language could not succeed. On the other hand, 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 gaps between sensory-motor and symbolic accounts by virtue of the symbolic nature of the output of the high-level controller and predictor models. That is, these models generate vectors of probabilities, that is, partially completed simulations, rather than specific actions. As Barsalou (1999) has demonstrated, these simulations, or perceptual symbols can function as logical symbols in conceptual systems. On the other hand, the 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 often 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 (see Steels, in press). Or, as Roy (in press) claims, "...to account for ... meaning, the language interpreter must have its own autonomous interest/goals/purposes." These individual interests force a perspective on the interpretation of the world and language.

MacWhinney (1999) describes four perspectival systems used in effective communication and as a basis for embodied simulation. These systems are derivation of affordances (that is, what a particular person can do with objects), spatial-temporal reference frames (object-, speaker, and environment-centered), causal action frames (understanding events from the perspective of the actor or the object of the action), and adopting the perspective of others. A different account of perspective is provided by Tomasello (2003) who describes dimensions of perspectival construals 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” (page 13).

Whereas Tomasello may be correct that a simple sensory-motor account does not include multiple perspectives, the ABL theory does provide a straightforward account of at least some aspects of perspective. That is, the output of different predictor models for the same event can provide different perspectives on that event. Thus, when an event is described as an instance of giving, predictors will generate a sequence of expectations similar to, as described above, 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.

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, for example 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)<sup>7</sup>, democracy (a vast, intertwined set of relations between behaviors and consequences), 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 perspective that involves making predictions regarding 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 (as predictions) and thus serves to ground abstract language.

### **Language and emotion**

There is no question that language and emotion effect one another. Words may thrill, petrify, or arouse, and when listening to a bigot the words may invoke anger. Simulation accounts of language, and ABL in particular, have a natural explanation for this influence. Namely, just as understanding language about action requires simulation using neural systems that control action (e.g., Glenberg & Kaschak, 2002), and just as understanding language about perceivable events requires neural systems that contribute to perception (e.g, Kaschak et al., 2005), understanding language about emotions and emotional situations requires neural systems that contribute to emotion. Whereas investigation of the language/emotion link is just beginning, there are several findings that support this simulation account.

First, if understanding language about emotional situations requires a simulation using the emotion system, then activating the appropriate emotion should facilitate subsequent language understanding, and activating an inappropriate emotion should interfere with language understanding. To test this prediction, Havas, Glenberg, & Rinck (2007) used the Strack, Martin, and Stepper (1988) pen procedure to induce emotions: A participant was asked to hold a pen with the teeth without using the lips or to hold a pen with the lips without using the teeth. The former produces a smile which reliably brightens mood, whereas the latter inhibits smiling and darkens mood. The main result was that participants were faster to read sentences describing happy situations when smiling (pen in teeth) than not smiling. Similarly, participants were faster to read sentences describing sad situations when not smiling (pen in lips) than when smiling. Thus, emotional state affects language processing.

Second, emotions differentially motivate defensive and appetitive actions (Bradley, Codispoti, Cuthbert, & Lang, 2001). Thus, if understanding language about emotional events activates motivational systems, people should be differentially prepared to take defensive and appetitive actions depending on the language. To assess this prediction, Moulso, Glenberg, Havas, and Lindeman (2007) asked participants to judge if sentences were sensible or not using a large response lever. Half of the participants indicated “sensible” by moving (as quickly as possible) the lever away from their bodies; this response uses a motion that could be described as “striking out.” The other participants indicated sensible by moving the lever toward the body using an affiliative action. The main finding was that when using the striking out movement, participants were faster to read sentences describing angry situations than sentences describing sad situations, and this was particularly true for male participants. In contrast, when using the affiliative movement, participants were faster to read the sad sentences than the angry sentences,

and this was particularly true for female participants. (See Bradley, Codispoti, Sabatinelli, and Lang, 2001, for data pertaining to gender differences in reactivity to stimuli evoking different emotions.)

Third, we can put together two links in the causal chain. Suppose that understanding emotional language requires simulating the emotional state, and that simulating the emotional state is in part preparing to take emotion-related action. Then, if that action is made difficult, the difficulty should be reflected in language understanding. To test this prediction, Glenberg, Havas, Webster, Mouilso, & Lindeman (in press) used the lever to differentially fatigue action systems. In one condition, participants made the striking out motion approximately 300 times by pushing the lever against a force; in the other condition the participants used the lever to make the affiliative response (against a force) approximately 300 times. Then, participants judged angry, sad, and neutral sentences as sensible or nonsense. In this experiment, however, the sensible response was made by pushing buttons, not the lever. Thus the question is whether fatiguing a response will affect language comprehension even when that response is not logically required for the comprehension task. We found that the striking out motion differentially slowed the judgment of the angry sentences, especially for the men. That is, increasing the difficulty of the striking out response appeared to increase the difficulty of simulating anger and consequently the difficulty of understanding sentences about angry events.

Consider the ABL explanation for the third finding. An angry sentence taken from the experiment is, “The workload from your pompous professor was unreasonable; this course evaluation will make the jerk pay.” A number of words in the sentence (e.g., “pompous,” “unreasonable,” and “jerk”) as well as some phrases (e.g., “make the jerk pay”) may be associated with negative emotions and defensive motivation: That is, the simulations

engendered by the predictors will include activation of emotional and motivational systems. In addition, one component of the meaning of “make the jerk pay” includes physical action such as striking out. That is, high-level action controllers associated with “make the jerk pay” include the lower-level action of striking out. The responsibility (probability) associated with this striking out component is increased by activation of the defensive motivational system. Finally, operation of the striking out modules is slowed by the previous 300 repetitions of the striking out motion due to any of several mechanisms. For example, given the peripheral fatigue, predictors may predict that quick action will be physically painful and hence the execution of low-level modules may be slowed. Or, the previous 300 repetitions of the striking out motion may have increased the responsibility for this module in the experimental context to the extent that it is difficult to simulate an incompatible writing response implied by “this course evaluation.” Clearly, the many components and the circular causality of this explanation need to be tested. The point to take away, however, is that the ABL theory provides a framework for understanding complex interactions between language comprehension and emotion that may be at the heart of how language can be such a powerful emotional force.

### **Attention and working memory**

One function of predictors that we have not yet discussed is that of selecting relevant components of sensory feedback. For example, in walking down a set of wet steps, it is important to process sensory and proprioceptive information corresponding to leg movement, but not as important to react to information corresponding to an itch on one’s scalp. Of course, just the opposite is true while combing one’s hair. Because predictor models generate the expected consequences of action, they can be construed as a mechanism for controlling attention to particular features of the environment.

The idea that attention is related to the output of predictor models fits well with the premotor theory of attention (Craighero, Nascimben, & Fadiga, 2004; Moore & Fallah, 2001; Rizzolatti, Riggio, Dascola, & Umilta, 1987). According to this theory, visual attention is closely linked to eye-movement planning, that is, attending is planning a saccade in preparation for interacting with an object. A finding consistent with this approach is reported by Craighero et al. (2004). Using the Posner cuing paradigm, they demonstrate faster detection of targets in cued locations compared to non-cued locations. Consider what happens, however, when the fixation point is moved so that the eyes must be rotated temporally 40 degrees to keep the fixation point in the fovea. In this case, the eyes cannot saccade to a target located more temporally (although the target is visible and responded to accurately), and hence the saccade cannot be planned. In this condition, there is no benefit of cuing the temporal target location. However, if the cue specifies a target location more nasal from the fixation point (so that a saccade can be programmed), the usual cueing effect is obtained.

The operation of the controller/predictor cycle also provides a rough approximation to working memory when envisioned as a system for keeping available a limited amount of information for immediate access. In Baddeley's classic view, working memory consists of two, limited capacity peripheral systems, a visual-spatial sketchpad and a phonological store and articulatory loop, both controlled by an executive system. In the ABL theory, the operation of the speech controllers and predictors provide a first approximation to the articulatory loop and phonological store. That is, the motor program generated by a controller is an articulation, and an efference copy of the articulation is used by the predictor to generate expected feedback, such as the sound of the word. Even if gain control is set so that the articulation is never made, the expected feedback is generated, and this expected feedback may well serve as a phonological

store. We know that articulatory suppression (e.g., overtly pronouncing the syllable “the”) eliminates many of the effects attributed to the phonological store (e.g., reduces phonological errors in recall of target material). Similarly, we suppose that occupying the effectors with the pronunciation of irrelevant material (e.g., “the”) will reduce the ability of predictors too generate predicted feedback for the target material.

This motor-based account of working memory has the additional benefit of being able to address short-term retention of information other than phonological and visual. For example, Reisberg, Rappaport, & O’Shaughnessy (1984) demonstrated how active manipulation of the fingers could be used to store information, resulting in a digit digit-span. Because control of finger movements will involve controllers and predictors, the cycle of planning a finger movement (using a controller) and generating expected feedback (using a predictor) will create a finger-image store much like the phonological store generated by planned articulation.

In general, our theory is consistent with the emergentist view of working memory as discussed by MacDonald and Christiansen (2002) and Postle (2006). Postle describes two principles of this point of view, “First, the retention of information in working memory is associated with sustained activity in the same brain regions that are responsible for the representation of that information in non-working memory situations, such as perception, semantic memory, oculo- and skeletomotor control, and speech comprehension and production. Second, humans opportunistically, automatically, recruit as many mental codes as are afforded by a stimulus when representing that stimulus in working memory” (page 31). Our identification of the vector of predicted probabilities with partial simulations aligns Postle’s account with ours. For example, the partial simulation of a visual event (visual working memory) will involve visual cortex and the partial simulation of an acoustic even will involve auditory cortex. In

addition, Postle reviews the large literature demonstrating the contributions of PFC, including IFG and Broca's area, to working memory. In the emergentist account, the contribution of PFC is not storage of information but control and attention. We believe that these working memory functions reflect the operation of controller/predictor modules.

### **Related accounts of language**

Many scientist are suggesting that there is a the connection between motor control and language (e.g, Gallese 2007, 2008; Lakoff & Gallese, 2005; Iacoboni & Wilson, 2006 ; Wolpert, et al., 2003 ; Zwaan & Taylor, 2006). Here we note several accounts similar to ours in regard to brain mechanisms and several accounts similar to ours in regard to computational mechanisms.

Van Schie, Toni, & Bekkering (2006) suggest that “understanding the neural mechanisms underlying goal-directed actions may provide a general framework to understand language processes” (page 496). Their framework rests on several similarities between networks for language and action. One similarity is that both action control and language rely on frontal-posterior networks that integrate perceptual and motor functions connected by the arcuate fasciculus. Second, they note that Broca's area is involved with both action and language in complex ways. BA 45, the rostral portion of Broca's area appears to be involved with concrete semantics for both language and action. BA 44 is proposed as controlling abstract motor representations for production of verbal and manual actions. Third, they note the correlation in hierarchical structure between action control and language. For example, many mirror neurons appear to encode a mid-level representation of action (e.g., grasp with mouth or hand) that is between the very specific representation of muscle control and the representation of goals at the highest levels. They also note that, “syntax, at least in its early or primitive stages, may depend on the structure of natural actions...” (page 497). These three points are consistent with the

ABL theory that we have developed.

The basic position developed by Fiebach and Schubotz (2006) and Schubotz (2007) is that the ventral premotor cortex (PMv) and its neighbor, Broca's area, serve as "a highly flexible sequence processor, with a complexity gradient from PMv towards Broca's area" (page 501). The results from a series of studies investigating Schubotz's serial prediction task contribute to the development of their proposal. In that task, participants are exposed to sequences of stimuli and instructed to attend to the temporal pattern of specific stimulus properties (e.g., spatial location of the stimuli, the objects represented at the locations, or rhythm of stimulus presentations). A hypersequence contains several repetitions of the temporal pattern, and the task is to determine if any of the sequences in the hypersequence contain a violation of the temporal pattern. An important finding is that different stimulus dimensions recruit different portions of PMv: rhythm recruits inferior PMv, objects recruit middle PMv, and spatial sequences recruit dorsal PMv. These areas of PMv are also associated with control of lips and tongue (most inferior PMv), hand (middle PMv), and arm (dorsal PMv), as if the different types of temporal sequences are differentially encoded by these bodily control systems. In their scheme, PMv represents (or recruits) templates of predictable constituents of actions, that is, simple motor plans, whereas Broca's area deals with hierarchical representations. These suggestions are consistent with the framework in Figure 3, wherein the basic motor repertoire is encoded in motor and pre-motor cortex, and Broca's area represents control at a more abstract hierarchical level.

Schubotz (2007) takes these ideas one step further by suggesting that the label "premotor cortex" may be a misnomer. Instead, the area is a mechanism for making predictions both of the outcome of bodily actions as well as predictions for events such as the breaking of waves on the

shore that cannot be directly imitated. In this account, Schubotz explicitly uses predictor/controller networks. For example, in predicting the sequence of waves, she suggests that one uses the articulatory system to code auditory rhythms and hand/arm pointing systems to code spatial characteristics of the waves. Predictors generate predictions for the waves based on these codes.

Pickering and Garrod (2007) propose an account of the relationship between language production and comprehension that bears some similarity to the ABL theory. They note that the evidence is strong that people can use predictions to enhance perception (see Wilson & Knöblich, 2005 for a review). Furthermore, the evidence is strong that language comprehension mechanisms can use predictions to enhance comprehension. For example, people are quite good at predicting the next word in a sentence given constraints provided by the linguistic and referential contexts. But where do these predictions come from? Pickering and Garrod suggest that the production system incorporates a forward (predictor) model. Furthermore, they suggest that predictions a) arise at different levels in the system corresponding to phonology (c.f., the speech predictor in Figure 5), syntax, and semantics (c.f. the give predictor in Figure 5), and b) that the predictions are probabilistic, as in Figure 5. Pickering and Garrod discuss the evidence that similar brain areas (e.g., Broca's area) are activated during language production and comprehension tasks, and the extensive evidence for spontaneous imitation during conversation: people repeat each other's words, syntax, meaning, accent, and speech rate. This imitation could arise quickly and effortlessly, if person A's production system is primed during A's comprehension of B, much like the role we have assigned to speech mirror neurons and action mirror neurons in Broca's area.

The DIVA model of Guenther et al. (2006) matches our notion of predictor/controller

models of speech almost exactly. The model is designed to account for the learning and production of syllables and short sequences of syllables. DIVA is well attested in areas such as motor equivalence, contextual variability, coarticulation, speaking rate effects, and speaking skill acquisition, and it provides a persuasive account of phenomena such as recalibration in light of changes in vocal tract size with development, effects of auditory feedback, as well as the results of brain imaging during speech tasks.

One component of the model consists of a Speech Sound Map conceived of as consisting of speech mirror neurons (putatively in left ventral premotor cortex). This map is connected (by a matrix of teachable synaptic weights) to the Articulatory Velocity and Position maps (in motor cortex) which then controls a speech simulator. The matrix of weights connecting the Speech Sound Map and the Articulatory and Velocity Maps is directly analogous to the speech controllers in the ABL theory. The Speech Sound Map is also connected (by two matrices of teachable weights) to a Somatosensory Error Map (in inferior parietal cortex) and an Auditory Error Map (in superior temporal cortex). These two matrices correspond to predictor models for proprioceptive and auditory feedback, respectively. The predictions (the error maps) are compared to sensory feedback, and the errors are used to adjust the motor commands in the Articulatory Velocity and Position maps. The errors are also used to learn the matrix of weights from the speech sound map to the Articulatory Velocity and Position maps. Thus, the DIVA model is an exceptionally well-specified and tested account of speech motor control for small units such as syllables. Because the mechanisms specified by the model are directly analogous to those in the ABL theory, we take the success of DIVA as an indication that at least the speech control component of the ABL theory is reasonable.

Finally, we note several similarities between the ABL theory and a recent proposal of

Kemmerer and Castillo (in press). They note that many action verbs have two levels of meaning. The basic or root level is modality (motor) specific and distinguishes between verbs in the same class such as “dripped,” “poured,” and “spilled” as well as between classes. However, at this level, verbs in the same class are treated identically by grammatical processes. Kemmerer and Castillo propose that the root level of verbs are represented in somatotopically-mapped mirror neuron systems in premotor and primary motor cortices.

The second level of verb meaning, which they call the template level, specifically deals with grammatical processes. At this level, verbs in the same class are treated identically, as in “Carol dripped/poured/spilled water on the flowers,” but verbs in different classes are treated differently. For example, one cannot say in English that “Carol drenched/soaked/saturated water on the flowers.” Instead, a different template is needed, as in “Carol drenched/soaked/saturated the flowers with water.” Kemmerer and Castillo (in press) associate the template level with a more schematic mirror neuron system in Broca’s area.

Thus, both ABL and Kemmerer and Castillo propose that the motor system plays a crucial role in both meaning and grammar. Furthermore, both posit that to a large degree, the neurophysiology of this role can be captured by a mirror neuron system. Finally, both note the necessity of multiple levels, locate the higher levels in Broca’s area, and assign grammatical functions to those higher-levels.

### **What needs to be done**

We have not provided details of any of the processes involved in language acquisition, comprehension, and production. Instead, we have demonstrated how mechanisms of motor control can provide a framework for understanding language that is able to ground the meaning of linguistic terms. Nonetheless, those details need to be worked out. For example, we have

sketched how children might acquire the determiners “this” and “that,” but we have not addressed the far more frequently used determiners “the” and “a.” Similarly, we have sketched the acquisition of only one syntactic structure, the double-object structure. We have focused on the double-object because the account can be constrained by extant data, such as those reported by Glenberg et al. (2008a, b) demonstrating use of motor system in comprehension of these structures.

An important issue concerns the conditions under which the system builds a new hierarchical level. One solution depends on memory processes. For example, a predictor may generate for Action 1 (and Efference copy 1) Prediction 1, and that prediction is confirmed by Feedback 1. However, unbeknownst to the system, the context changes so that Action 1 now produces Feedback 2. The system uses the error signal (difference between Prediction 1 and Feedback 2) to learn to generate Prediction 2 from the Efference Copy 1. Again, unbeknownst to the system, the context changes back to the original context, and now the system must relearn to generate Prediction 1 from Efference Copy 1. Memory for fluctuations of this sort (“why does this work sometimes and not others?”) might prompt the search for changes in the context that can be encoded as a new hierarchical level: in Context 1, Action 1 leads to Prediction 1, but in Context 2, Action 1 leads to Prediction 2. An episodic account provides some of the prerequisites for this mechanism in that multiple modules, some encoding the prediction of Feedback 1 and others encoding the prediction of Feedback 2 would be available for analysis.

Another possibility is that hierarchical levels arise from the natural operation of HMOSAIC in a language context. For example, suppose that a child has learned multiple *give* HMOSAICs (as in Figure 5) which specify lower-level modules for determining the object to be given, the recipient, and so on. Now, suppose that a child has also learned the concept

corresponding to “the cup that Lavinnia gave to dada,” that is, the child has a *cup\_dada* HMOSAIC for controlling attention to and manipulation of a particular cup. Now, consider the processing that occurs when the child hears “Give that cup to momma” along with a pointing gesture to the cup that Lavinnia had given to dada. Hearing “give” will activate various *give* HMOSAICs predicting that a graspable object will be mentioned next. The gesture guiding attention to the cup will activate the *cup\_dada* HMOSAIC. Because *cup\_dada* will mesh with the control structure of the *give* HMOSAICs, the child will act appropriately. In this case, language comprehension has produced a complexly embedded structure with multiple levels.

### **Conclusions**

We propose that hierarchical, goal-directed mechanisms of action control, namely paired controller/predictor models, have been exapted 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 (see the “neural exploitation hypothesis”, 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, premotor cortex is more complex: It structures simple motor behaviors into coordinated motor acts. Thus, premotor 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 premotor cortex to specific regions of the primary motor cortex.

Similarly, as exemplified by the MNS, the same premotor circuitry controlling action execution instantiates the embodied simulation of the observed actions of others. Thus,

premotor cortex and the MNS instantiate hierarchical control structures that can be exploited by language. The HMOSAIC architecture of the ABL model demonstrates how this exploitation could work.

## References

- Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
- Angrave, L. C., & Glenberg, A. M. (2007, August). *Infant gestures predict verb production one year later*. Paper presented at the annual meeting of the American Psychological Association.
- Awh E, Armstrong KM, Moore T (2006). Visual and oculomotor selection: links, causes and implications for spatial attention. *Trends in Cognitive Sciences*, 10, 1240130.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., Iacoboni, M. (2006). Embodied semantics and the premotor cortex: Congruent representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, 16, 1818 – 1823.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation*, Vol. 8 (pp. 47-89). New York: Academic Press.
- Bak, T. H. & Hodges, J. R. (2003). The effects of motor neurone disease on language: Further evidence. *Brain and Language*, 89, 354-361.
- Bannard, C. & Mathews, D. (2008). Stored word sequences in language learning. *Psychological Sciences*, 19, 240-248.
- Bernardis P, Gentilucci M. (2006) Speech and gesture share the same communication system. *Neuropsychologia*, 44: 178-90.
- Bock, K., & Lobell, H. (1990). Framing sentences. *Cognition*, 35, 1-39.
- Borghi, A. M., Glenberg, A. M., and Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, 32, 863-873.

- Boulenger, V., Roy, A.C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 ms of processing. *Journal of Cognitive Neuroscience*, 18, 1606-1615.
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, 1, 276-298.
- Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion*, 1, 300-319.
- Bransford, J. & Schwartz, D. Rethinking Transfer: A Simple Proposal with Multiple Implications, (1999). *Review of Research in Education*.
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V. et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13, 400-404.
- Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, 89, 370-376.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., Rizzolatti, G.. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24, 355-363.
- Buccino, G., Solodkin, A., & Small, S. (2006). Functions of the mirror neuron system: implications for neurorehabilitation. *Cognitive Behavioral Neurology*, 19, 55-63.
- Buresh, J.S., Woodward, A., & Brune, C. W. (2006). The roots of verbs in prelinguistic action knowledge. In K. Hirsh-Paske & R. M. Golinkoff (Eds.) *Action Meets Word: How Children Learn Verbs*, 208-227. New York: Oxford University Press.
- Chambers, C. G., Tanenhaus, M. K., and Magnuson, J. S. (2004). Actions and affordances in syntactic

- ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 687-696.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234-272.
- Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, 90, 29-49.
- Chao LL, Martin A: Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 605, 2000, 12:478-484.
- Childers, J. B., Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38, 967-978.
- Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1998). Visuomotor priming. *Visual Cognition*, 5, 109-125.
- De Vega (in press). Levels of embodied meaning. From pointing to counterfactuals. In M. de Vega, A. M. Glenberg, and A. C. Graesser (Eds.) *Symbols, Embodiment, and Meaning*, Oxford: Oxford University Press.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15, 399-402.
- Fadiga, L., Craighero, L., & Roy, A. (2006). Broca's region: A speech Area? In Y. Grodzinsky (Ed.) *Broca's Region*. Oxford: Oxford University Press.
- Fadiga, L. and Gallese, V. (1997) Action representation and language in the brain. *Theoretical Linguistics*, 23: 267-280.
- Fitch, W. T. & Hauser, M. D. (2004). Computational Constraints on Syntactic Processing in a Nonhuman Primate. *Science* , 303, 377-380.

- Fiebach, C. J. & Schubotz, R. I. (2006). Dynamic anticipatory processing of hierarchical sequential events: A common role for Broca's area and ventral premotor cortex across domains? *Cortex*, 42, 499-502.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F. and Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, 302, 662-667.
- Friederici A.D., Bahlmann J., Heim S., Schubotz R.I., and Anwander A. (2006) The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U. S. A.* 103: 2458-63.
- Galantucci, B, Fowler, C A., & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13, 361-377.
- Gallese, V. (2003a) The manifold nature of interpersonal relations: The quest for a common mechanism. *Phil. Trans. Royal Soc. London B.*, 358: 517-528.
- Gallese, V. (2003b) A neuroscientific grasp of concepts: From control to representation. *Phil. Trans. Royal Soc. London B.*, 358: 1231-1240.
- Gallese V. (2007) Before and below Theory of mind: Embodied simulation and the neural correlates of social cognition. *Proc. Royal Soc. Biol. Biology*, 362: 659-669.
- Gallese, V. (2008) Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, in press.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493-501.
- Gallese, V., and Lakoff, G. (2005) The Brain's Concepts: The Role of the Sensory-Motor System in

- Reason and Language. *Cognitive Neuropsychology*, 22: 455-479.
- Gentilucci, M., Bernardis, P., Crisi, G., & Volta, R. D. (2006) Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *Journal of Cognitive Neuroscience*, 18: 1059-1074.
- Gentilucci M, Corballis MC. (2006) From manual gesture to speech: a gradual transition. *Neurosci Biobehav Rev.* 30: 949-60.
- Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Paske & R. M. Golinkoff (Eds.) *Action Meets Word: How Children Learn Verbs*, 544-564. New York: Oxford University Press.
- Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. (2006) Recursive syntactic pattern learning by songbirds. *Nature*. Apr 27;440 (7088):1204-7.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1-19.
- Glenberg, A.M. (in press). Language and action: Creating sensible combinations of ideas. To appear in G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.
- Glenberg, A. M., Havas, D. A., Webster, B. J., Mouilso, E., & Lindeman, L. M. (manuscript under review). *Gender, emotion, and the embodiment of language comprehension*.
- Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558-565.
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26, 69-83.
- Glenberg, A. M., Webster, B. J., Mouilso, E., & Lindeman, L. M. (in press). Gender, emotion, and the embodiment of language comprehension. *Emotion Review*.
- Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, 28, 1-26.

- Glenberg, A. M. & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379-401.
- Glenberg, A. M., Sato, M., Cattaneo, L. (2008a). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology*, 18, R290-R291.
- Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., Buccino, G. (2008b). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology*, 61, 905-919.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In B. MacWhinney (Ed.), *The emergence of language*, 197-212. Mahwah, NJ: Lawrence Erlbaum Associates.
- Goldberg, A. E, Casenhiser, D. M., Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15, 289-316.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.
- Goldin-Meadow, S., Seligman, M. E. P., & Gelman, R. (1976). Language in the two-year-old: Receptive and productive stages. *Cognition*, 4, 189-202.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12, 516-522.
- Grafton ST, Fadiga L, Arbib MA, Rizzolatti G. (1997) Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*, 6:231-236.

- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27,377-442.
- Guenther, F. H., Ghosh, S. S., Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain & Language*, 96, 280-301.
- Haruno, M, Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. *International Congress Series* 1250, 575-590.
- Hauk, O., Johnsrude, I., Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301-307.
- Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, 14, 436-441.
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your Own Action Influences How You Perceive Another Person's Action. *Current Biology*, 14, 493-498.
- Hintzman, D. C. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gesture as simulated action. *Psychonomic Bulletin & Review*.
- Hurley, S. (2004). The shared circuits hypothesis: A unified functional architecture for control, imitation, and simulation. In S. Hurley & N. Chater (Eds.), *Persepectives on Imitation: From Neuroscience to Social Science*. Cambridge, MA: MIT Press.
- Huttenlocher, J., Smiley, P., and Charney, R. (1983). Emergence of action categories in the child: evidence from verb meanings. *Psychological Review*, 90, 72-93.
- Iacoboni, M., & Wilson, S. M. (2006). Beyond a single area: motor control and language within a neural architecture encompassing Broca's area. *Cortex*, 42, 503-506.

- Iverson, J. M., and Goldin-Meadow, S. (2001). The resilience of gesture in talk: gesture in blind speakers and listeners. *Developmental Science*, 4, 416-422.
- Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
- Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, 43, 508-529.
- Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133, 450-467.
- Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A., and Zwaan, R. A. (2005). Perception of Motion Affects Language Processing. *Cognition*, 94 (3), B79-B89.
- Kelly, S. D., Barr, D. J., Church, R. B., and Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40, 577-592.
- Kemmerer, D. M., & Castillo, J. G. (in press). The two-level theory of verb meaning: An approach to integrating the semantics of action with the mirror neuron system. *Brain and Language*.
- Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7, 54-60.
- Llinas, R. (2001). *The i of the vortex*. Cambridge, MA: MIT Press.
- Masur, E. F. (1997). Maternal labeling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language*, 24, 427-439.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney. (Ed.), *The Emergence of Language*

- (pp. 177-196). Mahwah, NJ: Erlbaum.
- MacDonald, M. C. & Christiansen, M. H. (2002). Reassessing working memory: A reply to Just & Carpenter and Waters & Caplan. *Psychological Review*, 109, 35-54.
- MacWhinney, B. (1999). The emergence of language from embodiment. In B. MacWhinney. (Ed.), *The Emergence of Language* (pp. 177-196). Mahwah, NJ: Erlbaum.
- Martin, A., and Chao, L.L. (2001) Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, 11:194–201.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, D. & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.). *Language and gesture: Window into thought and action* (pp. 141-161). Cambridge: Cambridge University Press.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17, 1692-1696.
- Moore, T., & Fallah, M. (2001). Control of eye movements and spatial attention. *Proceedings of the National Academy of Sciences*. 98, 1273–1276.
- Maouene, J. & Hidaka, S. & Smith, L. B. (in press) Body Parts and Early-Learned Verbs. *Cognitive Science*.
- Mouilso, E., Glenberg, A. M., Havas, D. A., & Lindeman, L. M.. (2007). Differences in action tendencies distinguish anger and sadness after comprehension of emotional sentences. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1325-1330). Austin, TX: Cognitive Science Society
- Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2003). Verifying different-modality properties for

- concepts produces switching costs. *Psychological Science*, 14, 119-125.
- Perani D, Schnur T, Tettamanti M, Gorno-Tempini M, Cappa SF, and Fazio F. (1999) Word and picture matching: a PET study of semantic category effects. *Neuropsychologia*, 37:293-306.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science*, 11, 105-110.
- Plunkett, K., & Marchman, V. ((1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 1-60.
- Postle, B. R. (2006). Working memory as an emergent property of the mind and the brain. *Neuroscience*, 139, 23-38.
- Pulvermüller F. (2002) The neuroscience of language. Cambridge University Press, Cambridge, UK.
- Pulvermüller, F. (2005) Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576-582.
- Pulvermüller, F. (in press). Grounding language in the brain. In M. de Vega, A. M. Glenberg, and A. C. Graesser (Eds.) *Symbols, Embodiment, and Meaning*, Oxford: Oxford University Press.
- Reisberg, D., Rappaport, I., & O'Shaughnessy, M. (1984). Limits of working memory: The digit digit-span. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 203-221.
- Roy, D. (in press). A computational model of three facets of meaning. In M. de Vega, A. M. Glenberg, and A. C. Graesser (Eds.) *Symbols, Embodiment, and Meaning*, Oxford: Oxford University Press.
- Richardson, D.C. & Spivey, M. J. (2000). Representation, space, and Hollywood squares: looking at things that aren't there anymore. *Cognition*, 76, 269-295.
- Rizzolatti G., Camarda R., Fogassi M., Gentilucci M., Luppino G. and Matelli M. (1988) Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal

- movements. *Exp. Brain Res.*, 71: 491-507.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-192.
- Rizzolatti, G., & Craighero, L. (2007). Language and mirror neurons. In G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.
- Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. *Cog. Brain Res.*, 3: 131-141.
- Rizzolatti G, Fogassi L, Gallese V. (2000). Cortical mechanisms subserving object grasping and action recognition: a new view on the cortical motor functions. In Gazzaniga MS, (Ed.), *The Cognitive Neurosciences*, Second Edition. Cambridge, MA: MIT Press, pp. 539-552.
- Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, 31, 889-901.
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory. *Neuropsychologia*, 25, 31-40.
- Rueschemeyer, S-A., Glenberg, A. M., Kaschak, M. P. & Friederici, A. (under review). Listening to sentences describing visual motion activates MT/V5.
- Scorolli, C., Borghi, A. M., & Glenberg, A. M. (in press). Language-induced motor activity in bi-manual lifting. *Experimental Brain Research*.
- Sommerville & Decety (2006). Weaving the fabric of social interaction: Articulating developmental psychology and cognitive neuroscience in the domain of motor cognition. *Psychonomic Bulletin & Review*, 13, 179-200.
- Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality & Social Psychology*, 54, 768-777.

- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P. Fazio, F., Rizzolatti, G., Cappa, S. F., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17, 273-281.
- Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, 14, 663-668.
- Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language*, 58, 188-213.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, pp 209-253.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Umiltà, M.A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., Gallese, V., and Rizzolatti, G. (2008). How pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences*, 105: 2209-2213.
- VanSchie, H. T., Toni, I., Bekkering, H. (2006). Comparable mechanisms for action and language: neural systems behind intentions, goals, and means. *Cortex*, 42, 495-498.
- Vigliocco, G., Warren, J., Siri, S., Aruili, J, Scott, S., & Wise, R. (2006). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex*, 16(12)
- Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience*. 16, 978-987.
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 48, 989-994.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London, B*.

358, 593-602.

Wolpert, D.M., Kawato, M. (1998). Multiple paired forward and inverse models for motor control.

*Neural Networks*, 11, 1317-1329.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics.

*Psychological Bulletin*, 131, 460-473.

Zwaan, R. A.& Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language

comprehension. *Journal of Experimental Psychology: General*, 135, 1-11.

#### 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.

## Footnotes

<sup>1</sup>The notion that linguistic symbols are arbitrary can be challenged, but we will not pursue the challenge in this paper.

<sup>2</sup>It should be added, though, that T. Gentner et al. (2006) recently showed that European starlings do possess the capacity to recognize acoustic patterns defined by a recursive, centre-embedding, context-free grammar

<sup>3</sup> In fact, an informal experiment reveals that people do envision some arm motions in comprehending these sentences. The three abstract transfer sentences noted in the text were mixed with three sentences describing literal arm movements (e.g., “Hand the pencil to Sally”) and three no-arm-action sentences (e.g., “Look at the moon in the twilight”). Participants were asked to rate on a four-point scale the likelihood that they would use a literal arm or hand motion during the described activity. The 12 participants gave an average rating of 3.97 to the literal arm sentences, and an average of 1.64 for the no-arm-action sentences. The mean rating for the abstract transfer sentence was 2.24, significantly greater than the rating for the no-action sentences,  $t(11) = 3.46, p < .01$ . Nine of the twelve participants rated the abstract sentences as more likely to involve a hand or arm action than the no-arm-action sentences.

<sup>4</sup> Although, 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.” One would predict that 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 (McDonald, 1999).

<sup>5</sup>In these experiments, the left to right order of objects depicted in the picture is counterbalanced so that the pirate, crown, and queen might be depicted in a left to right order or

a right to left order. The ABL theory predicts more priming when the order in the picture matches the prime scan pattern.

<sup>6</sup>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.

<sup>7</sup>Thanks to Chad Mortensen for this observation.