SPECIAL ISSUE:

**EMBODIMENT AND DEVELOPMENT**

Guest editors: Josita MAOUENE, Thea IONESCU

**Editorial**

403  • Josita MAOUENE, Thea IONESCU

**Articles**

409  • Aarre LAAKSO
   EMBODIMENT AND DEVELOPMENT IN COGNITIVE SCIENCE

427  • Gwenden L. DUEKER, Sandra PORTKO, Megan ZELINSKY
   MEANINGFUL TOUCH IN NATURALISTIC CONTEXTS:
   HAPTIC INPUT AS A CUE TO THE REFERENT OF INFANT DIRECTED SPEECH
• Josita MAOUENE, Nitya SETHURAMAN, Aarre LAAKSO, Mounir MAOUENE
THE BODY REGION CORRELATES OF CONCRETE AND ABSTRACT VERBS IN EARLY CHILD LANGUAGE

• Karin Harman JAMES, Paroma BOSE
SELF-GENERATED ACTIONS DURING LEARNING OBJECTS AND SOUNDS CREATE SENSORI-MOTOR SYSTEMS IN THE DEVELOPING BRAIN

• Elizabeth M. WAKEFIELD, Karin H. JAMES
EFFECTS OF SENSORI-MOTOR LEARNING ON MELODY PROCESSING ACROSS DEVELOPMENT

• Hanako YOSHIDA, Joseph M. BURLING
A NEW PERSPECTIVE ON EMBODIED SOCIAL ATTENTION

• Susan Wagner COOK
ABSTRACT THINKING IN SPACE AND TIME: USING GESTURE TO LEARN MATH

• Larissa K. SAMUELSON
ABSTRACT THINKING IN SPACE AND TIME: USING THE ENVIRONMENT TO LEARN WORDS
In the last two decades, we have observed an expansion of the definition of cognition (Gentner, 2010a). While the dominant view until the 1980s considered cognition as a separate entity from the body (and for some it still is), that view is declining as growing evidence indicates that cognition cannot be separated easily - or for some not at all – from the body's morphology and its sensorimotor systems nor from the context it is embedded in. As Gentner (2010a) shows in her recent analysis of the history of psychology in cognitive science, embodiment research is growing.

What about it in cognitive developmental psychology? Is the embodiment perspective growing? It certainly is in domains concerned with unveiling certain sensorimotor processes such as those involved in learning how to crawl, walk, avoid obstacles, etc. (Adolph, 2008), learning eye-hand-object coordination (Street, Jones & Smith, 2011), searching for hidden objects (Perry, Samuelson, & Spencer, 2009), or with the processes involved in feedback about the consequences of infants’ actions and how they could entice infants to start reaching (Needham & Libertus, 2011). The growth is also perceptible in domains that go beyond the boundaries of what Piaget called the sensorimotor stage of intellectual development. As such it can be extended to the domains of categorization, language, mathematics, and prelinguistic communication. Some of us study now the relation between social visuo-motor processes and higher level-cognition such as word learning (Pereira, Smith & Yu, 2008), how visual motor movement is connected with statistical word learning (Smith and Yu, in press), and how visual intelligence connects to symbolic process (Smith & Jones, 2011). It is also being investigated in the work on tactile categorization of object properties in infants (Sheya and Smith, 2010), in how gestures influence thought (Beilock & Goldin-Meadow, 2010) and mathematics (Goldin-Meadow, Cook & Mitchell, 2009), how purposeful intentions of others are grasped by very young infants through the perception of body movements, facial
expressions and gesture (Gallagher & Hutto, in press). However, there are still domains where an embodiment perspective is, to our knowledge, inexisten

however not unthinkable). For example, the role of the sensory-motor processes in the development of knowledge about the physical world and causal reasoning (Luo & Bailleaigeon, 2010; Shutts, Condry, Santos, & Spelke; 2009), the role of sensory-motor processes in representational insight (Troseth, Bloom, and Deloache, 2007) and the role of embodiment in analogical reasoning and structural mapping (Gentner, 2010b). We wish to argue that there is also room for embodied and situated cognition for “declining areas” such as representation of knowledge, conceptual semantics, and computational semantics (Gentner, 2010a), in particular in the modeling of developmental semantic networks (Steyvers & Tenenbaum, 2005; Hills, Maouene, Maouene, Sheya & Smith, 2009a,b). Ultimately embodied and situated cognition perspectives will have to propose whether there are such a thing as modal networks or not (Schubert & Semin, 2009) and we predict that the question of the code will reemerge (Pylyshyn, 1981).

All the contributions to this Special Issue reflect developments that are of central relevance and interest to the topic of learning in cognitive science from the point of view of embodiment. The special issue starts with an introductory paper on Embodiment and Development in Cognitive Science from Laakso. Laakso’s introduction aims to give the reader a sense of the main issues regarding the role of embodiment in cognition and development. The paper begins with the history of the “disembodied” perspective that dominated cognitive science from the origins of the discipline through the 1980s, and the history of the “embodied” perspective. Then, it surveys some arguments for the embodied cognition hypothesis and describes some of the empirical evidence that weighs in its favor, with a focus on the developmental literature in particular. In the end it points out important challenges for supporters of the embodied cognition hypothesis.

After this introductory paper, experimental and correlational papers will follow with chronological age of the tested population as the organizational principle. The second contribution is that of Dueker, Portko & Zelinksy on 6- to 11-month-old infants and their mothers entitled Meaningful touch in naturalistic contexts: Haptic input as a cue to the referent of Infant Directed Speech. They report that infants are more likely to be touched, particularly held or held-up when adults speak about topics for which there are no immediately useful physical referents (other than themselves) than when they speak about objects that are present in the immediate environment. Further, infants are likely to experience more simultaneous touch contact points when adults are speaking about a topic that does not have an immediate physical referent than when speaking about a nearby object. But, the location of touch did not vary by topic. They hypothesize that haptic input could affect infant attention to various aspects of the context and also, that haptic input could facilitate multimodal coordination. The third paper is a correlational study of Maouene, Sethuraman, Laakso & Maouene examining the speech of 12
to 23 month-olds entitled *The body region correlates of concrete and abstract verbs in early child language*. The authors report evidence that embodiment is an important factor to consider in children’s acquisition of verbs, with two major findings: first, those verbs that are more highly associated with a single region of the body, according to adult judgments, are acquired among the first in vocabulary development; second, associations with a specific region of the body may be a better predictor of whether a verb is learned early than the concreteness or abstractness of the verb.

The fourth contribution is that of James and Bose with 5- to 7-year-olds entitled *Self-generated actions during learning objects and sounds create sensori-motor systems in the developing brain*. These authors report four important findings that contribute to enrich the embodied perspective on learning. First, in children, learning about objects actively results in very different brain system recruitment than observing others act. Second, sounds learned actively result in motor, auditory, and visual system recruitment more than passive observation upon subsequent auditory presentation. Third, videos of objects that are learned actively result in motor and visual system recruitment more than passively learned objects. And fourth, when given all learned information about objects through multi-sensory presentation, active learning results in greater visual system recruitment than passive observation. The fifth contribution is that of Wakefield and James with 4- to 7-year-olds and adults and the *Effects of Sensori-Motor Learning on Melody Processing across Development*. They present two studies on how abstract actions affect perception, and how this may change across development. They address this question by teaching children (4-7 year-olds) and adults to sing melodies, with or without an abstract motor component, and by using functional Magnetic Resonance Imaging (fMRI) to determine how these melodies were subsequently processed. Results show that developmental change occurs during melody processing. Children, contrary to adults, do not process motor signs as meaningful, although associations among sensori-motor systems are still created. These associations lead to motor system recruitment during purely perceptual tasks only when movement is incorporated into the learning episode. Later, in adulthood, motor movement is coded as meaningful to the learning episode. Therefore, as we mature, abstract motor movement become meaningful to the learning event.

Review papers follow, although they are not what one might call "classic style" reviews. The sixth contribution is that of Yoshida and Burling, a paper at the crossing between experimental and theoretical purposes entitled *A new perspective on embodied social attention*. Yoshida and Burling present a perspective paper, where they first discuss the role of early social input in language learning as it has been treated in the literature so far and then introduce their recent theoretical and empirical embodied direction. The authors hypothesize a link between sensori-motor experiences and embodied attention—specifically how different bodies produce different kinds of attention. Understanding the role of bodily events (the child’s and the child’s social partners’) in early visual
experiences provide insight into the development of learning mechanisms and the processes involved in learning disabilities. They share with us their first data on normally developing hearing children, deaf children of deaf families, and children with autism who were observed in a social context using a new child-centered technology. The final contributions are a review of Wagner Cook on Abstract thinking in space and time: Using gesture to learn math and a commentary from Samuelson entitled Abstract thinking in space and time: using the environment to learn words. Cook first briefly discusses some of the evidence for embodied cognition in general, focusing on language and motor processes as these are most relevant for thinking about gesture as an embodied representation. She then discusses evidence that hand gestures are functionally involved in mathematical thinking in both children and adults. Finally, she addresses some possible criticisms and data that are difficult to explain from this perspective. Samuelson’s comments on Cook’s arguments, highlighting how this view of math as embodied offers new insights for our understanding of classic developmental themes, in particular, the continuity versus discontinuity dichotomy. In addition, she presents a brief summary of recent work on how children use their bodies in another realm typically thought of as abstract—understanding referential intent. She presents an embodied account of how children disambiguate speaker intent in novel naming situations and argues that, as in the case of embodied math, an embodied view of cognition can help elucidate developmental mechanisms.

In surveying some of the latest elements brought to the debate about embodied cognition by studies with adult subjects, the contributions of the authors in this special issue help delineate which objections we can discard and which ones we cannot (see Laakso, this issue). For example, arguments according to which abstract concepts like beauty cannot be accounted for by the embodied theory of cognition are not problematic (Mahon & Caramazza, 2008). Cook and Samuelson remind us that very abstract entities like numbers are experiences as are beauty and justice and thus pertaining to the spatiotemporal context and the perception-action systems involved in their processing. However, we do see issues with the mirror neurons account (Arbib & Bota, 2006) since some development data (James & Swain, 2011 and James & Bose, this issue) as well as lesion data in adults (Arevalo, Balso & Dronkers, in press) seem to cast doubt on their existence in humans. An associative theory of embodiment might be a better option. Similarly, for embodied semantics, the results of James and Maouene (2009) and Arevalo, Balso & Dronkers (in press) do not support a strict interpretation of the homuncular topology view where language would be restricted to the network responsible for execution and action, but supports the view of a greater motor-language network of associations. Developmental science will help in that regard - the question is whether adult cognitive development research will take into account this compelling evidence from research on child development.
REFERENCES


Cognition, Brain, Behavior. An Interdisciplinary Journal
15 (2011) 403-408


EMBODIMENT AND DEVELOPMENT IN COGNITIVE SCIENCE

Aarre LAAKSO*

Department of Behavioral Sciences, University of Michigan-Dearborn, USA

ABSTRACT

This paper is a survey of the main issues surrounding the role of embodiment in cognition and development. For most of the history of cognitive science, a “disembodied” model of cognition—according to which cognitive processes are independent of sensorimotor processes—dominated the field. More recently, the field has taken a turn toward embodiment, the view that sensorimotor processes routinely influence (and perhaps even constitute) cognitive processes. This view has diverse historical origins in American pragmatism, phenomenology and ecological psychology. However, in recent years, researchers have adduced many arguments and a great deal of empirical evidence in favor of the embodied cognition hypothesis. Developmental psychologists have played an important role in this transformation of cognitive science, adopting and adapting Piaget’s views as a means of explicating the role that embodiment plays in development, as well as collecting developmental data that support the embodied cognition hypothesis. Despite the enthusiasm for embodiment among cognitive scientists generally and developmentalists in particular, however, the embodied cognition hypothesis still faces formidable challenges.

KEYWORDS: cognition, development, embodiment, sensorimotor systems

* Corresponding author:
E-mail: aarre@umd.umich.edu
Over the last 20 years or so, an enormous amount has been said in the cognitive science literature about the topic of embodiment, some of it specifically about the role that embodiment plays in development. A comprehensive review of the evidence and arguments is not practical in this forum. Instead, the goal of this paper is to give the reader a sense of the main issues regarding the role of embodiment in cognition and development. The first section traces the history of the “disembodied” perspective that dominated cognitive science from the origins of the discipline through the 1980s. The second section traces the history of the “embodied” perspective. In addition, it surveys some arguments for the embodied cognition hypothesis and describes some of the empirical evidence that weighs in its favor. The third section examines some of the evidence and theoretical contributions that have come out of the developmental literature in particular. The fourth section points out important challenges for supporters of the embodied cognition hypothesis.

THE DISEMBODIED PERSPECTIVE IN COGNITIVE SCIENCE

For many years, the orthodox view in cognitive science was that the study of cognition, including the study of conceptual processes and mechanisms, could proceed independently from the study of sensorimotor processes and mechanisms. Researchers assumed that it was sufficient to use discrete, amodal symbols to represent not only the contents of thoughts but also the sensory activities that provide information about the environment and the motor activities that constitute behavior. They considered these sensorimotor processes irrelevant to the core project of cognitive science—explaining cognitive processes, which were thought to take place in an abstract information-processing device. Sensorimotor processes, on this view, were relegated to peripheral input and output devices (M. Wilson, 2002).

The view that cognition can (and should) be studied independently of perception and action had its roots in the computational metaphor of mind, beginning with Turing’s concept of a universal computing machine (Turing, 1936). Turing essentially reduced the problem of creating a thinking machine to the problem of calculating mathematical functions (Rohrer, 2007). From its origins in Turing’s work, the universal computing machine became the preferred metaphor for the mind through the 1980s (Johnson & Rohrer, 2007).

There was a natural affinity between the computational metaphor of mind and the philosophical position known as functionalism (Putnam, 1960), the view that mental states inherit their contents from their functional roles, that is, their roles in a computational system. Moreover, according to this view, because different kinds of media (such as vacuum tubes or transistors) can compute the same function, cognition can take place in a variety of different media, whether biological
or manufactured. Functionalists regard the mind as a “black box” whose inputs and outputs are symbols. The precise mechanisms inside the black box are irrelevant to the function it computes. From this perspective, explaining cognition is a matter of getting the black box to transform the input symbols into the correct output symbols. The physical body disappears (Rohrer, 2007). This point of view was advocated forcefully by Jerry Fodor (Fodor, 1975, 1983) and Zenon Pylyshyn (Pylyshyn, 1984, 1999) and dominated cognitive science until recently.

**THE EMBODIED COGNITION HYPOTHESIS**

Although the “disembodied” perspective governed cognitive science for many years, there is a very different view that also has an impressive, albeit different, intellectual heritage. This view emphasizes the importance of sensorimotor processes for successful interaction with the environment (M. Wilson, 2002). According to this view, which is often called the “embodied cognition hypothesis,” the specific details of how the brain and body embody the mind do matter to cognition (Rohrer, 2007).

The embodied cognition hypothesis has origins in American pragmatism, particularly the writings of William James and John Dewey. In the *Principles of Psychology*, James wrote “My thinking is first and last and always for the sake of my doing” (W. James, 1890, p. 333). This observation, which one often hears paraphrased as “thinking is for doing,” suggests that thought is intimately entwined with the motor system. Similarly, Dewey writes in *Experience and Nature*:

> To see the organism in nature, the nervous system in the organism, the brain in the nervous system, the cortex in the brain is the answer to the problems which haunt philosophy. And when thus seen they will be seen to be in, not as marbles are in a box but as events are in history, in a moving, growing never finished process (Dewey, 1925, p. 224).

In this passage, Dewey not only claims that embodiment is crucial for understanding the mind but also acknowledges the developmental, dynamic nature of cognition.

A further root of the embodied view of mind lies in J. J. Gibson’s ecological psychology, which took possible interactions with the immediate environment (“affordances”) to be the basis of perception. As Andy Clark and others have noted (Clark, 1999; Kirsh, 1991b; Van Leeuwen, 1998), however, ecological psychology runs into problems when one considers what many take to be the most characteristic form of cognition—cognition in the absence of its objects.

In philosophy, the phenomenologists, particularly Heidegger and Merleau-Ponty, also remarked upon the centrality of embodiment. In *Being and Time* (1927), Heidegger pointed out that human beings, and all of the things humans do, are
always in the world. One cannot study being independently of “being-in-the-world,” (a Heideggerian term of art) and therefore, one cannot bracket the world when studying human activities.

Merleau-Ponty, the most scientifically informed of the phenomenologists, argued that embodiment in the three-dimensional, physical space of the actual world is a critical fact about human thought. Thus, Merleau-Ponty writes, in the Phenomenology of Perception:

Insofar as, when I reflect on the essence of subjectivity, I find it bound up with that of the body and that of the world, this is because my existence as subjectivity is merely one with my existence as a body and with the existence of the world, and because the subject that I am, when taken concretely, is inseparable from this body and this world (Merleau-Ponty, 1962, p. 475).

These points were taken up again by the Dreyfus brothers in their criticisms of artificial intelligence methodology in the 1990s (Dreyfus & Dreyfus, 1999).

In contemporary philosophy, even many of those who had endorsed functionalism came to the realization that psychology cannot give a complete account of meaning and that, therefore, it is necessary to study mind and language from an embodied, social point of view (Putnam, 1988). Some philosophers also argued that the idea of non-embodied perception is unintelligible (New, 1976) or that there is no reason to believe in the possibility of disembodied existence (Tye, 1983).

A variety of commentators inside and outside of philosophy have offered other a priori arguments for the embodied cognition hypothesis. Some (e.g., Birnbaum, 1991; Kirsh, 1991a; B. C. Smith, 1991) emphasize the connection between embodiment and the “symbol grounding problem.” In reflecting upon traditional “disembodied” computational systems, they point out that an agent cannot grasp a concept merely by tokening an arbitrary symbol. To count as grasping a concept, a symbol tokening needs to provide the agent with a variety of abilities, including some that involve sensorimotor processes. For example, no agent could genuinely be counted as having grasped the concept *blue* if it was unable to discriminate blue objects from those that are not blue. Thus, for a symbol to be truly counted as representing a concept, there must be mechanisms that connect it to the outside world. These mechanisms are, of course, the sensorimotor systems.

Other commentators (Evans, 1982; Kirsh, 1991a) have instead focused on the fact that some of our referring expressions (e.g., “over there”) are inherently indexical. That is, they have meaning only from the egocentric perspective from which human beings perceive and act in space. That being the case, it is not possible to fully simulate or explain cognition, including language, without appeal to the body and its attitude in space.
Another argument appeals to evolution (M. Wilson, 2002). Human beings evolved from organisms with limited nervous systems wholly dedicated to sensorimotor processes. Their behavior was limited to direct, real time interconnectivity with their surroundings. Thus, cognition in humans (instead of being amodal, symbolic, and separate from sensorimotor processes) may have similarly rich connections with the sensorimotor periphery.

Arguments aside, there is an enormous amount of empirical evidence for the embodied cognition hypothesis. It would not be possible to go through all of it in this forum, but a sampling will be instructive. The following paragraphs discuss seminal early work in cognitive linguistics and (briefly) robotics. The remainder of the section is then devoted to describing a few important illustrative experimental results.

The resurgence of the embodied cognition hypothesis in cognitive science began in the early 1980s, when cognitive linguists started explaining abstract concepts as metaphorical extensions of concrete concepts, particularly those related to the body (Lakoff & Johnson, 1980). This line of work begins with the claim that humans conceive of the world through metaphorical reasoning. That is, humans learn nearly all concepts through metaphor: human beings acquire new concepts almost exclusively by grasping them in terms of previously mastered concepts. To avoid a vicious regress, however, humans must be able to grasp a small number of atomic concepts non-metaphorically. Such atomic concepts, according to Lakoff and Johnson, derive from interactions between the human body and the external world—grasping these concepts depends upon having a body.

Certain spatial concepts, including up and down, are atomic concepts on this view. Human beings acquire the concepts up and down because moving requires holding or altering a normally erect posture. Subsequently, bodily concepts like up and down scaffold the acquisition of more complex concepts. Up and down, in particular, can be used to structure metaphorical conceptions of happiness and sadness, as Lakoff and Johnson (1980) illustrate with examples like: “I’m feeling up” and “I’m feeling down.” In other words, metaphorical reasoning, an essential component of concept learning, depends upon contingent properties of the human body. The body is thus directly involved in cognitive processes.

Also in the 1980s, roboticists began demonstrating that sophisticated behavior could be achieved by combining simple environmental interactions instead of handcrafting complex internal representations. Brooks (1986), in particular, advocated the idea of a “subsumption architecture,” which involves building robots by successively adding limited control mechanisms over raw sensory and motor systems in order to progressively obtain more sophisticated behavior. Such an approach highlights the role of the sensorimotor systems in driving complex behavior while minimizing—or possibly eliminating—the role of central representations (e.g., concepts) in doing so, as Brooks (1991) emphasizes.
More recently, a wide variety of experimental findings in psychology have provided empirical support for the embodied cognition hypothesis. Richardson, Spivey, Barsalou, and McRae (2003), for example, demonstrated that comprehending verbs with horizontal or vertical image schemas interacted with perceptual and memory tasks. Thus, for example, participants who heard a sentence and saw two pictures, of the agent and the patient of the sentence, were later faster to recognize the picture pairs if they were presented in the same orientation as the image schema of the verb in the associated sentence. Richardson et al. argue that such effects suggest that linguistic representations have a sensorimotor character.

Zwaan, Madden, Yaxley, and Aveyard (2004) found that participants were faster to judge that two sequentially presented visual objects were the same when the implied motion between the stimuli matched the movement described in an accompanying sentence. For example, if the sentence was “The pitcher hurled the softball to you” and the second visual stimulus, depicting a softball, was slightly larger than the first, which was also a softball (implying motion toward the participant), then responses were facilitated. The result, they argue, suggests that language comprehension routinely involves the activation of visual motion (i.e., sensorimotor) representations. Moreover, they suggest, it provides converging evidence for the hypothesis that language comprehension generally involves perceptual simulation of the described situation, contrary to theories postulating that language is represented propositionally, that is, amodally (e.g., Pylyshyn, 1984).

These sorts of findings are not limited to aspects of visual perception (orientation, implied motion), nor are they limited to linguistic stimuli. For example, Simmons, Martin, and Barsalou (2005) found, using event-related fMRI, that pictures of appetizing foods not only activate visual shape areas but also activate areas of cortex specialized for taste and food reward. That is, they showed that a visual stimulus activates gustatory areas. To give another example, N. L. Wilson and Gibbs (2007) showed that moving the body in a particular way (e.g., shaking), or even just imagining doing so, primes comprehension of metaphorical phrases related to that action (e.g., “shake off a feeling”). In this case, a motor act facilitates processing of a linguistic stimulus. Simmons et al. (2007) found, again using event-related fMRI, that the same cortical area (the left fusiform gyrus) is active both when perceiving color and when processing color property words such as “blue” and “green.” That is, there appears to be a common neural substrate for perceiving colors and conceptualizing (retrieving knowledge about) colors.

In light of these arguments and evidence over the last 30 years, the embodied cognition hypothesis has increasingly become the mainstream position in cognitive science.
EMBODIMENT IN DEVELOPMENT

Developmental researchers have made crucial contributions of evidence and arguments supporting the embodied cognition hypothesis. In addition to the origins discussed in the previous section, the embodied cognition hypothesis can be traced to Jean Piaget’s developmental psychology, specifically to his emphasis on the way that cognitive skills emerge out of a foundation of sensorimotor abilities (M. Wilson, 2002). Piaget thought the earliest stage of development consisted primarily if not exclusively of sensorimotor capabilities and saw the issue of how adult cognition arises from those limitations as the core question for developmental psychology.

Thelen (2000), however, argues that development of mature cognition does not require overcoming or abandoning these early sensorimotor skills but rather refining them and making them more flexible. In line with that view, Thelen, Schöner, Scheier, and Smith (2001) describe a dynamic systems model of the “A-not-B” error. This is the phenomenon, documented by Piaget (1963), that 7- to 12-month-old infants who have been successful in finding a toy at one location, A, continue to reach for that location even after they see the toy hidden in another location, B. Thelen et al. (2001) dispute standard accounts, according to which the A-not-B error results from a deficit in infants’ knowledge about objects. Instead, they claim, it is best explained in terms of operations like perceiving, planning, deciding, reaching and remembering. Some of those operations (i.e., planning, deciding and remembering) sound characteristically cognitive, but the model supports the embodied cognition hypothesis by implementing them in the same dynamic field as that used to capture the reaching movement. The authors maintain that this dynamic engagement between the mental and the sensorimotor is characteristic of cognition at all ages (Thelen et al., 2001).

The notion of embodiment in use in the developmental literature differs slightly from that in the adult literature. Linda B. Smith, for example, defines the “embodiment hypothesis” as “the idea that intelligence emerges in the interaction of an organism with an environment and as a result of sensory-motor activity” (L. B. Smith, 2005b; L. B. Smith & Gasser, 2005). Note the emphasis on how intelligence “emerges.” This is because a developmental perspective focuses inquiry on the issue of explaining change. The disembodied approach, besides assuming that cognition can be explained independently of sensorimotor processes, also makes the assumption that David Kirsh (1991a) called “learning can be added later.” This is the assumption that adult human cognition (the only kind that matters) can be explained independently of development and learning. In the developmental literature, advocates of the embodied approach challenge both assumptions. L. B. Smith and Gasser (2005), for example, argue that starting out embodied and embedded in the world is necessary for the development of human cognition. L. B. Smith and Sheya (2010) argue that the very reason that mainstream cognitive
science failed to explain developmental change for so many years was the fact that it was based on a disembodied approach.

The most important empirical evidence for embodiment to come out of the developmental literature has been a series of findings showing that developmental changes in body posture and motor activity have surprising and wide-ranging effects on putatively cognitive tasks. Clearfield (2004), for example, showed that infants who learned to find a hidden goal location in a large space while crawling could no longer do so after the transition to walking, suggesting that infants’ spatial learning is linked to their mode of locomotion. In another example, Broaders, Cook, Mitchell, and Goldin-Meadow (2007) demonstrated that telling children to gesture while explaining their (incorrect) answers to novel arithmetic problems improves their ability to solve the problems correctly later, suggesting that the children’s mathematical learning is linked to their motor activity.

Karin Harman James has documented similar effects using fMRI. K. H. James (2009) showed that preschool children who practiced printing letters subsequently had greater activity in visual association cortex during letter perception than those who had merely practiced visual letter recognition. In other words, the motor experience associated with printing letters augmented processing in the visual system.

The developmental neuroimaging results are not limited to effects of the motor system on perceptual systems such as vision. K. H. James and Maouene (2009) found that listening to verbs—but not adjectives—activated motor systems in the frontal cortex of preschool children. Moreover, verbs associated with different body parts by adults (Maouene, Hidaka, & Smith, 2008) activated different parts of motor cortex in children. An additional study using novel verbs (K. H. James & Swain, 2010) showed that the motor system in preschoolers is only activated during auditory perception of a verb if the verb has been learned by self-generated (active) exploration of an object with which the verb was paired. Passively viewing manipulation of an object paired with a verb resulted in substantially lower activation of motor regions.

Indeed, theorists of child language have emphasized for some time that language emerges from action (Bates, Camaioni, & Volterra, 1975; Bruner, 1975; Iverson & Thelen, 1999; Kelly et al., 2002). Recently, Iverson (2010) has argued that the acquisition of motor skills gives infants the chance to hone skills that will later be needed for language acquisition. For instance, rhythmic arm movements (such as hand banging) provide opportunities for the infant to practice the sort of temporally precise, recurrent patterns of movement that are necessary for babbling. In addition, Iverson argues, the achievement of certain motor development milestones alters infants’ relations with the world in ways that contribute to language acquisition. For instance, the onset of crawling, by increasing the infant’s proximity to dangerous objects and contexts while simultaneously decreasing the infant’s proximity to the caretaker, increases the frequency and intensity of distal
communication from the caretaker. These developments motivate the infant to find the object of the caregiver’s communication, and this motivation contributes to improvements in the infant’s ability to follow gaze and pointing gestures. That is, crawling indirectly leads to improvements in the ability to maintain joint attention, which is critical for language acquisition among other things.

In a commentary on Iverson’s paper, Adolph, Tamis-LeMonda, and Karasik (2010) argue that Iverson may not go far enough. Whereas Iverson claims only that aspects of motor development are “normally participatory” in language acquisition but neither necessary nor sufficient, Adolph et al. (2010) suggest that motor development should be central not only to developmental psychology but to psychology in general. Similarly, L. B. Smith and Sheya (2010) offer a synthesis of the developmental research that places action at the center. In their view, bodily actions routinely bring an individual’s various sensorimotor systems into dynamic couplings with each other, changing the sensorimotor systems themselves. These changes, over time, transcend single modalities and particular tasks, leading ultimately to the sort of complex, flexible behavior that characterizes adult cognition.

CHALLENGES FOR THE EMBODIED COGNITION HYPOTHESIS

The previous sections have shown that there is a wide variety of evidence that sensorimotor processing accompanies conceptual processing of various sorts, both in adults and in children. Most researchers have taken these results as evidence for the embodied cognition hypothesis, that is, as evidence that sensorimotor processes influence or constitute cognitive processes. However, as Mahon and Caramazza (2008) note, that inference is invalid on two grounds. One is that most of the empirical evidence to date can be explained just as well by a “disembodied” theory supplemented by hypotheses about spreading activation between the conceptual system and sensorimotor systems. Another is that there is some neuropsychological evidence that specifically weighs against the embodied cognition hypothesis. The next few paragraphs discuss some of the neuropsychological evidence for and against the embodied cognition hypothesis; the final paragraphs of this section address the point that the evidence usually marshaled for the embodied cognition hypothesis is also consistent with a kind of disembodied theory.

Mahon and Caramazza (2008) consider the disorder apraxia as an example of neuropsychological evidence against the embodied cognition hypothesis. Apraxia consists of an inability to use objects; that is, it is an impairment of motor processes. The embodied cognition hypothesis requires that motor processes causally influence cognitive processes. Strong versions require that motor processes (together with sensory processes) are constitutive of cognitive processes. If the embodied cognition hypothesis is true, then, one might expect that apraxia would also impair
object recognition or action recognition. However, apraxia affects neither object recognition nor action recognition. Patients with apraxia can name objects that they are unable to use, and they can recognize pantomimes of actions that would be appropriate for objects they are unable to use. Therefore, apraxia demonstrates that certain motor processes are not necessary for corresponding cognitive processes. Indeed, even paraplegia does not interfere with object recognition, so clearly many motor processes are not necessary for object recognition! By contrast, if sensorimotor processes were actually constitutive of cognitive processes (like object or action recognition), one would expect that the inability to use objects would cause failures in the abilities to recognize the objects and the actions appropriate to them. The phenomenon of apraxia (not to mention paraplegia) therefore rules out extreme forms of the embodied cognition hypothesis.

However, both apraxia and paraplegia are consistent with weaker forms of the embodied cognition hypothesis. Sensorimotor processes (such as using an object) may influence central processes (such as recognizing the object) without being necessary for them. The burden is then on the proponent of the embodied cognition hypothesis to explain in exactly what sense sensorimotor processes influence cognitive processes short of being necessary for them. In addition, the weaker versions of the embodied cognition hypothesis that are consistent with apraxia and paraplegia are less radical and therefore, perhaps, less interesting. Explaining apraxia and paraplegia seems to require postulating that central processes like recognition can sometimes operate independently of relevant sensorimotor processes. That is, explaining apraxia and paraplegia seems to require leaving room for cognitive processing independent of sensorimotor processing. Doing so raises questions about the centrality of sensorimotor processing in cognition. In other words, if sensorimotor processing is not necessary for cognitive processing, then sensorimotor processing may not be so interesting after all.

There is evidence that relevant sensorimotor processing is sometimes necessary for cognitive processing, including object recognition. For example, preventing gesture results in speech disfluencies, particularly for spatial content (Rauscher, Krauss, & Chen, 1996). In some sense, then, gesture (i.e., a sensorimotor activity) is sometimes necessary for fluent speech (i.e., producing language, a characteristically cognitive activity).

Further evidence that sensorimotor processing is sometimes necessary for certain kinds of cognitive processing comes from the fact that people at different levels of expertise (e.g., novice as opposed to expert musicians or chess players) perceive stimuli differently. For instance, professional basketball players are better than college students at recognizing basketball dribbles from point-light displays, and are better at identifying their own dribbles than those of others from such displays (Hohmann, Troje, Olmos, & Munzert, 2011). That is, the particular sensorimotor experiences accumulated in the course of attaining expertise at something like basketball seem to be necessary for the task of recognizing an expert.
activity such as dribbling and identifying who is performing that activity from informationally impoverished stimuli. Note, however, that the necessity in this case is on a different time scale than that involved in apraxia. Apraxia demonstrates that relevant sensorimotor processes in real time are not always necessary for cognitive processes like recognition. Basketball expertise, by contrast, demonstrates that relevant sensorimotor processes active during the development of expertise are sometimes necessary for later cognitive processes like recognition and identification. Moreover, the phenomenon of expertise is consistent with weaker versions of the embodied cognition hypothesis. That is, the fact that earlier sensorimotor processes are sometimes necessary for later cognitive processes does not impugn the claim that cognitive processes sometimes occur in the absence of relevant sensorimotor processes acting in real time.

Along similar lines, it is worth considering object agnosia, which may be conceptualized as the converse deficit to apraxia. Object agnosia consists of an inability to visually recognize familiar objects, but object agnosics can mimic actions appropriate to objects they cannot recognize (Magnié, Ferreira, Giusiano, & Poncet, 1999). For instance, there is the famous example of a man who cannot recognize a picture of a combination lock but pantomimes opening a combination lock while looking at the picture. Indeed, there are many reports (e.g., Brain, 1941) of severe object agnosics who can still move through the world and act appropriately on objects. Arguably, such cases reinforce the notion that the motor system stores certain kinds of object-specific information as motor programs that can be accessed independently of explicit visual recognition. Object agnosia is therefore also evidence against strong versions of the embodied cognition hypothesis. That is, to the extent that motor processes constitute cognitive processes, one might expect that the availability of an object-specific motor program would be sufficient for recognizing the object. However, object agnosia shows that the availability of an object-specific motor program is not sufficient for recognizing an object.

At the same time, object agnosia does not tell against weaker versions of the embodied cognition hypothesis. Object-specific motor programs may causally affect cognitive processes such as recognition without thereby being sufficient for them. Once again, however, the weaker versions of the embodied cognition hypothesis that are consistent with the phenomenon are less radical and, perhaps, less interesting than the strong version of the hypothesis. Explaining object agnosia, in particular, seems to require postulating that sensorimotor processes can sometimes operate independently of relevant cognitive processes. That is, as with apraxia and paraplegia, explaining object agnosia seems to require leaving room for cognitive processing independent of sensorimotor processing. Doing so raises questions about the centrality of sensorimotor processing in cognition. Once again, if sensorimotor processing is not sufficient for cognitive processing, then sensorimotor processing may not be so interesting after all.
The neuropsychological literature is full of interesting interactions and dissociations like these between sensorimotor processes and cognitive processes. However, it is important to keep in mind that the relations between sensorimotor processes and cognitive processes are multifaceted and graded. Even visual agnosics can recognize some objects, and apraxics have not been exhaustively tested on learning to recognize novel objects without typical motor interaction. Many of the more interesting effects that neuropsychologists have discovered have only emerged clearly from exhaustively testing the abilities and inabilities of patients, but this sort of testing is by no means routine. Hence, it is likely that neuropsychology will continue to provide surprises in the coming years. Given the current state of knowledge, however, it does not make sense to place a lot of theoretical weight on particular neuropsychological observations.

The second challenge to the embodied cognition hypothesis that is raised by Mahon and Caramazza (2008) is the claim that the empirical data that have been taken to support embodied cognition can equally well be explained by a version of a disembodied cognition hypothesis that postulates spreading activation between sensorimotor areas and conceptual areas. This is a version of an argument that advocates of embodiment often hear—that embodied cognition amounts essentially to associative learning between sensorimotor experiences and cognitive processes. That is, associative learning could easily explain all of the examples cited in the previous two sections of this paper. The Simmons et al. (2005) experiment showing that gustatory regions are activated by viewing picture of appetizing foods, for example, could be explained in terms of associative learning. The explanation would posit that viewing images of tasty foods activates regions of gustatory cortex because people have previously learned to associate the objects depicted in the images with good tastes. In general, people learn associations among descriptions or representations of things, the things to which these descriptions or representations refer in the physical world, and the effects typically caused by such things. These learned associations may explain the existing data thought to favor the embodied cognition hypothesis.

One way to respond to such concerns is to grant that embodiment describes associative learning, but argue that is not really the point. Rather, the point is that words, images and concepts in general, in order to have meanings, must be associated with real things in the world with which a person has had prior experience. Those prior experiences ground the concepts in the world via sensorimotor systems. This kind of response, though, begs the argument put forth by Mahon and Caramazza (2008). The foe of embodiment may grant that the meanings of ordinary, everyday concepts are grounded in previous sensorimotor experiences in much the same way that a basketball professional’s concept of a dribble is grounded in previous sensorimotor experiences. Doing so, however, is not the same as granting that sensorimotor processes are causally active in the processing of such concepts in real time, let alone that sensorimotor processes...
happening in real time constitute the processing of such concepts. The defense provided by the grounding argument is therefore only a defense of a weak version of the embodied cognition hypothesis. On the weak version, sensorimotor processes participate in grounding concepts by being active simultaneously with (and so associated with) both the referents and the nascent concepts during concept acquisition. However, on the weak version of the hypothesis, sensorimotor processes are not causally involved with, let alone constitutive of, mature conceptual processes happening in real time. The sensorimotor associations remain, of course, but they manifest themselves in both neuroimaging studies and behavioral studies as inert byproducts of the learning process rather than as causes or constituents of conceptual processing.

For the embodiment theorist to address this challenge, it is necessary to demonstrate not merely that sensorimotor processing is associated with cognitive processing, nor even that cognitive processing affects sensorimotor processing, but that sensorimotor processing causally affects (or constitutes) cognitive processing in real time. Taking up this challenge, Pezzulo et al. (2011) cite several behavioral studies demonstrating that particular motor actions affect responses on cognitive tasks. For example, L. B. Smith (2005a) showed that repeatedly performing actions on an object changes 2.5-year-olds’ cognitive representation of the object. Specifically, children who moved a symmetrical object horizontally tended to extend its name to similar but wider objects, whereas those who moved the symmetrical object vertically tended to extend its name to similar but taller objects. As Pezzulo et al. (2011) acknowledge, however, there is at present a relative paucity of evidence that sensorimotor processing causally affects cognitive processing, compared to the many results demonstrating a mere association or a causal link in the other direction. Thus, an important remaining challenge for the embodied approach is to empirically document phenomena that are consistent with the embodied cognition hypothesis but not with the hypothesis that “disembodied” cognition spreads activation to sensorimotor systems as a mere side-effect of previously learned associations.

While Pezzulo et al. (2011) clearly support the embodied approach, they also describe in detail a number of additional challenges that it must meet if it is to live up to its ambitions. These include producing more empirical results demonstrating that sensorimotor processing causally influences characteristically high-level cognitive processes, such as language, decision making, reasoning and problem solving. Of particular note is their challenge for the robotics community to implement embodied computational models explicitly demonstrating how sensorimotor processes may comprise cognitive processes. As they note, such models will improve theories of embodied cognition by elucidating and examining their essential features, including the degree to which sensorimotor processes causally affect cognitive processing. In so doing, the models will thereby also open new frontiers for research on embodied cognition.
CONCLUSION

For most of the history of cognitive science, the “disembodied” view that cognition and development can be studied independently of sensorimotor processes has dominated the field. Recently, however, cognitive science has assimilated several other intellectual traditions (such as American pragmatism, phenomenology, Piagetian cognitive development, and ecological psychology) that emphasize the important roles that the body plays in cognition. The result has been a flourishing theoretical and empirical research effort on the embodied cognition hypothesis, one that has produced many novel and interesting findings. The ultimate fate of the embodied cognition hypothesis, however, depends on the extent to which it can meet several outstanding challenges. Among these are the need to document more thoroughly via empirical studies the causal role that sensorimotor processes may play in cognitive processes, particularly higher cognitive processes, and the need to produce and study mechanistic implementations of the theory in robots. Developmental scientists will continue to play critical roles in answering both of these challenges, as well as the other challenges that will undoubtedly arise as the embodied cognition hypothesis is tested, revised and elaborated.

REFERENCES


MEANINGFUL TOUCH IN NATURALISTIC CONTEXTS: HAPTIC INPUT AS A CUE TO THE REFERENT OF INFANT DIRECTED SPEECH

Gwenden L. DUEKER*, Sandra PORTKO, Megan ZELINSKY

Department of Psychology, Grand Valley State University, Michigan, USA

ABSTRACT

Caregivers structure the contexts in which infants develop and learn about their world. The naturalistic interactions of twenty caregiver-infant dyads in a laboratory setting were recorded and aspects of caregiver touch and speech were coded and examined for relationship. Analysis revealed a significant relationship between caregiver touch and topic of caregiver communication. Overall touch and number of simultaneous touch contact points varied by topic of conversation with adults much less likely to touch infants while speaking about an object in the immediate physical environment than when speaking about other topics. Location of touch did not vary by topic. These results suggest that haptic input is an embodied cue that could direct infant attention to the referent of adult speech and help to align input from multiple sensory modalities.

KEYWORDS: embodied cognition, haptic input, intersensory redundancy, lexical development, infant-caregiver system

In recent years, the focus of developmental researchers has shifted from asking whether language is innate or learned through environmental influence to studying how language emerges in human infants in their natural contexts of development. From this perspective, the environment does not merely provide a backdrop for the infants and caregivers interactions, rather the entire event is actually part of a

* Corresponding author:
E-mail: duekerg@gvsu.edu
system. Viewed in this way, the “environment” of language is not as impoverished as some earlier theorists have suggested (Pinker, 1984) and recent researchers emphasized that all facets of a language context perceptually available to young children developing language must be considered part of overall language learning.

In this experiment, the researcher seeks to clarify the role of the caregiver-infant interaction system as part of the context in the development of language. Specifically, this experiment is designed to explore the interactions between caregivers and infants in a naturalistic setting to determine whether caregiver touch varies with the topic of caregiver speech to the infant.

**EMBODIED COGNITION AND DYNAMIC SYSTEMS**

Investigation of the contextual influences on development has contributed greatly to the understanding of human behavior. New insights into context effects on children’s cognitive, social-emotional, and language development have spurred research with infants in the first year of life in an attempt to identify some of the specific influences. As the impact of context has been investigated, the concept of *embodiment* has been gaining prominence (Clark, 2008; Smith, 2005; Smith, Chen, & Pereira, 2011). There are varying definitions of this term, but they all center on the fact that cognition is a process that involves the entire active organism and its interactions with its entire context. As such, the physical body of the human organism is assumed to be actively engaged in the overall problem-solving processes of the organism and the emerging skills of the infant are both a result of and stimulus for continuous exchanges of information with the context (Thelen, Schöner, Scheier & Smith (2001). The organism and its context constitute an organized system. While a systems approach did not originate in the field of psychology, Bronfenbrenner adapted the ideas and introduced ecological systems theory in 1979, thus enlarging the arena of potential influences on the developing organism. Gibson and Gibson expanded the research by focusing on the interactions of the individual with the affordances of the objects in his/her world (E. Gibson, 1969; J. Gibson, 1979). In 1994, Thelen and Smith introduced the concept of dynamic systems theory which, in a sense, integrated and expanded the two earlier theories by studying the capabilities of the human body and incorporating these effects as part of the overall system in which human infants develop. Dynamic systems theory provided a language for discussing change and development across a variety of domains. In this system, the characteristic of emergence provided a means of discussing the variations in development of the human infant that arose out of the reciprocal interactions of the organism and its context.
SOCIAL COMMUNICATION AND BODY CUES

Obviously, there are variations in the quality of information a given context can provide. For younger infants, these informational contexts are primarily structured by the interactions with the caregiver that rely heavily on sensory stimulation and feedback. The caregiver could convey meaning through many combinations of contextual cues: gazing, touching, holding, changing postures, and the production of different sounds. One way that caregivers share contextual information with infants is through the establishment of what Tomasello (1999) has labeled “joint attentional scenes”. Essentially, this is when an infant and caregiver are both focused on each other and on another factor, such as an object or an event. By nine months of age, the infant can begin to detect intention on the part of another and indicates this in several ways (Tomasello, 1999). Intention as an aspect of a “joint attentional scene” can involve gazing at the partner or at an object or pointing at someone or something while maintaining interaction with the partner. It is more than the partner simply watching the infant’s actions; the partner is clearly part of the ongoing event, the communication is a social event (Tomasello, 1999).

In social communication, the participants rely on cues from each other to direct the course of the interaction. This is especially important with infants who are in the process of acquiring language so they are able to understand the communication. Some researchers have studied infants’ abilities to determine the referent of speakers’ utterances by examining how infants begin to differentiate relevant from non-relevant cues (Sabbagh & Baldwin, 2005) in learning novel words. In those studies, infants between 12 to 18 months of age learned novel names for unfamiliar objects that coincided with the speakers’ focusing attention on those objects while speaking the name. Random touching of articles, without simultaneous gazing, did not result in learning the new names. Baldwin and Saylor (2004; Saylor & Baldwin, 2004) studied infants’ nonverbal and verbal behaviors in response to utterances about present and absent caregivers. They observed that infants’ between 15-months and 24-months attended visually to the speaker (experimentator) when the utterances concerned the absent caregiver. In addition, the 18 and 24 month-old infants displayed search behaviors and pointing gestures. The 30 month-olds in their study demonstrated fewer pointing and search behaviors but more verbal responses to the experimenter’s utterances. They interpreted this sequence of actions as demonstrating the infants’ developing understanding of speaker intentionality, but acknowledged there was no satisfactory explanation of the emergence of this behavior. The intentionality pattern of behaviors described by the researchers definitely began with a visual orientation to the speaker during the caregiver-absent trials. When the caregiver was present during the utterances, the infant gazed at the caregiver and not the speaker. The utterances about absent caregivers were considered an example of increasing abstractness of the referent.
From a systems perspective, the whole organisms (caregiver and infant) are intimately involved in this experience and the involvement is very physical, requiring coordination of multiple sensory and motor inputs. Thus, the understanding of objects, people and events that emerges is thoroughly grounded in the infant’s physical experiences (Thelen & Smith, 1994). It is in this sense that cognitive processes can be described as embodied.

**Multimodal Presentation of Information**

Recent studies of aspects of contexts typical in infant experiences have focused on the role of intersensory redundancy or multimodal presentation of information (especially vision and motor-movement coordination and visual-auditory coordination) to help direct the infants’ attention to important aspects of the environment. According to Bahrick and Lickliter (2000), intersensory redundancy occurs when stimuli conveying the same information are presented to two or more sense modalities simultaneously. Essentially, the more modalities involved in the presentation of stimuli, the greater the likelihood that the infant will attend to and process the information presented (Bahrick, Netto & Hernandez-Reif, 1998; Flom & Bahrick, 2010; Lewkowicz, 1988a; 1988b; 1992; 1996; Morrongiello, Fenwick & Chance, 1998; Pickens & Bahrick, 1995). Information synchronously presented in audio-visual modalities is discriminated and attended to significantly better than information presented in a unimodal fashion or an asynchronous bimodal fashion (Bahrick & Lickliter, 2000). Such processing would enhance the aspects of language, memory, attention and cognition represented in embodiment (Thelen et al., 2001). In spite of this growing literature on the relationship between bodily actions and sensations and cognition, one sensory modality has received comparatively less attention: the sense of touch or haptic perception.

**Touch, Cognition and Language**

The relationship of bodily actions and the development of cognition have been explored in various ways (e.g. Needham, Barrett & Peterman, 2002; Pereira, Smith & Yu, 2008; Smith et al., 2011). Needham and her colleagues (2002) concluded that infants’ own early object manipulation provided infants with salient information that would enhance their ability to attend to more features of the object. Another aspect of touch is the actual motor activity of the infants themselves. Iverson (2010) reviewed the language development literature and proposed that many of the results previously attributed solely to visual and auditory aspects of the environment could be reinterpreted as evidence for the importance of infant motor activity in language acquisition. Basically, she suggests that the changing motor abilities of the child
provide a constant source of new information about the world that contributes to all the developing aspects of perception, cognition and language calling the relationship “complex and multifaceted” (p. 258).

A number of researchers have addressed the specific relationships between produced gestures and lexical development in children, beginning at two years of age (Skarabela, 2007; Kita, Ozyurek, Allen, Brown, Furman & Ishizuku, 2007; Ozyurek, Kita, Allen, Brown, Furman & Ishizuku, 2008). These researchers discovered robust relationships between types of gestures and types of verb phrases produced in children from ages 3-9 and adults across two very different language typologies (Ozyurek et al., 2008). Other researchers have explored the role of children’s motor activity in the development of specific verbs and have shown that many common and early-learned verbs often refer to highly specific actions by specific body parts (e.g. Maouene, Hidaka and Smith, 2008).

Caregiver-infant interactions are characterized by physical activity by both partners and include frequent touching of the infant by the caregiver (Ferber, Feldman & Makhoul, 2008; Stack & Muir, 1990). Most investigations of adult touching of infants have explored how emotions can be communicated through touch and the role touch plays in developing emotional self-regulation and promoting secure attachment (e.g., Field, Grizzle, Scafidi, Abrams, Richardson, Kuhn, et al 1996; Field, Healey, Goldstein & Guthertz, 1990; Hertenstein & Campos, 2001; Tronick, 1989, 1995; Weiss, Wilson, Hertenstein & Campos, 2000; see Field, 2010 for a recent review).

However, physical aspects of dyadic interactions may provide cues that play a greater role in linguistic and cognitive development than previously realized. These cues are embedded in the social interactions of the infant-caregiver system from which the baby must become capable of gleaning meaning which eventually comes to be represented by the referents of the caregiver utterances. Kelly (2001) and others (Hoff, 2006; Iverson, Capirici, Longobardi & Caselli, 1999; Trautman & Rollins, 2006) have suggested that one function social interaction plays in communication is to serve as the link between the language produced and the particular aspects of the physical environment that are necessary for decoding the meaning of that language. Indeed, infants do treat the social partner as a privileged source of information about that link (Baldwin, Markman, Bill, Desjardins & Irwin, 1996; Campbell & Namy, 2003). These relationships serve to help the infants and caregivers learn to focus their attention on the same stimuli and help create the “joint attentional scenes” described by Tomasello (1999).

Other investigators have begun to specifically address the role of adult motor activity in providing intersensory redundancy and the role it plays in infant language development (e.g., Brand & Shallcross, 2008; Gogate & Bahrick, 1998; Gogate, Bahrick & Watson, 2000; Gogate & Walker-Andrews, 2001). Gogate and her colleagues (2000) discovered that when mothers were asked to teach their pre-lexical infants labels for novel objects, 99.9% of the mothers’ attempts were
multimodal and 60% of these involved temporally synchronous object motion and verbal labels. Goldin-Meadow and Beilock, who have focused their work on identifying the relationship between gesture and thinking in children and adults maintain that “gestures convey substantive information….that play a role in changing thought.” (2010, p. 664; Bielock & Goldin-Meadow, 2010).

Some interesting research has been conducted in a population of hearing impaired individuals (Koester & Lahti-Harper, 2010; Waxman & Spencer, 1997). Waxman & Spencer (1997) conducted a longitudinal study of communication patterns in deaf and hearing mothers of deaf and hearing infants. They discovered that deaf mothers of deaf children used touch to gain attention significantly more often than hearing mothers of hearing children did. Additionally, deaf mothers of hearing infants also touched the child’s body to attract attention significantly more often than hearing mothers of deaf infants. In comparing the groups of mothers, they found that deaf mothers followed the attention strategies with a signed communication more than twice as often as hearing mothers of deaf infants, thus explicitly connecting touch with language. When hearing mothers touched their deaf infants, they were significantly less likely to follow the touch with communication, suggesting that their infants might be missing out on important contextual cues that the deaf mothers provided for their deaf children. The deaf mothers of hearing children did not touch their children as frequently as the deaf mothers of deaf children, but they used touch significantly more frequently than the hearing mothers of deaf infants and they followed the touch with communication, just as the deaf mothers of hearing infants did. The deaf mothers were clearly using touch to influence their infants’ language development.

It is possible that caregivers of typically developing infants modify their touches during daily interactions in ways that vary systematically and carry information, thus changing the context and the “joint attentional scene” for the infant. Such consistent variations in touch could provide additional contextual information to help disambiguate the referent of adult speech, for example, communicating the referent or overall topic of the utterance. The current experiment seeks to investigate the ways through which caregivers of preverbal infants provide haptic information in the context of naturalistic interaction. Is there a relationship between the topic of caregivers’ infant directed speech and whether the caregiver is touching the infant? If so, does the type and location of caregiver touch vary by the topic of caregiver speech? Specifically the researchers investigated if touch patterns differed by whether adults were speaking about an object in the immediate physical environment or speaking about a topic for which there was no immediate physical reference.
EXPERIMENT 1

Methods

Participants

Twenty infants between the ages of 6 months, 11 days and 11,28 (M = 8,14; SD = 43 days) and their female primary caregivers participated in Study 1. Nine of the infants were female (M = 8,6; SD = 42 days) and eleven were male (M = 8,20; SD = 45 days). Two infants were identified as Bi-racial, sixteen as Caucasian and two caretakers did not identify their infant’s ethnicity. Participants volunteered through a web-site describing the study, or were identified through local birth announcements and recruited via mail and phone solicitations. Infants included in these analyses represent a randomly selected subset of participants from a larger study of adult-infant dyadic interactions.

Design & Procedure

The procedure’s design simulated a naturalistic, home-like play setting and took place on a 3’ X 3’ colorful floor mat in the laboratory. An infant seat was placed beside the mat for caretakers who wished to use it for any part of the interaction. The stimuli included four sets of objects (plastic fruits, stuffed toy dogs, baby clothing, and blocks of different shapes) and four topic cards. Topic cards read: Please talk to your baby about (a) how he/she is feeling today, (b) the weather or the season, (c) his or her family, and (d) what he or she did yesterday. Each object set was in a small basket and all of the materials were placed in a large crate with the topic cards stacked beside the baskets. To avoid order effects, cards and object sets were shuffled prior to presentation to dyads. Caretakers were instructed to interact with their infant like they would at home and asked to use the objects and cards as part of that interaction, switching from object set to card and back again until they had covered all the topics. Dyads were digitally recorded while they interacted for as long as they wanted. Later, all caretaker utterances were transcribed verbatim.

Data coding procedures

Each caretaker utterance was coded for topic and for touch.

Topic coding. All utterances were coded as having occurred during interaction about objects (anything said about the fruit, dogs, clothes, blocks or the baskets the objects were in), as about more abstract topics (utterances about feelings, yesterday, the weather/season or family), or as off-topic (speech unrelated to the task). Off-topic utterances were excluded from the analyses.
**Touch coding.** Each caretaker utterance was coded for whether the caretaker was touching the infant or not. If the caretaker was touching the infant, then that touch was coded on a number of dimensions including whether and how the infant was being supported by the caretaker, what the caretaker was touching the infant with, the number of overall support and non-support touch points the infant was experiencing, and the location of those touches on the infant’s body. Please see Table 1 for operational definitions of each coding category. Only the first instance of touch that occurred during an utterance was included in the analysis. Touch that occurred while caretakers were not speaking was not included in this analysis. Two coders who were blind to the hypothesis of this study coded touch. Twenty percent of the dyads included in the study were coded by both coders. Reliability across the various categories ranged from 86% to 100% with an average of 94%.

Table 1.
Operational Definitions of Touch Coding Categories

<table>
<thead>
<tr>
<th>Was infant touched? Operational Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes Touch is initiated by caregiver with or without object in hand.</td>
</tr>
<tr>
<td>No Caregiver does not initiate touch.</td>
</tr>
<tr>
<td>Type of touch</td>
</tr>
<tr>
<td>Support Touch bears part or all of infant’s weight.</td>
</tr>
<tr>
<td>Non-support Touch does not bear weight.</td>
</tr>
<tr>
<td>Type of support touch</td>
</tr>
<tr>
<td>Held Caregiver supports the majority of infant’s weight off ground.</td>
</tr>
<tr>
<td>Held-up Infant’s weight is stabilized by caregiver but not fully supported.</td>
</tr>
<tr>
<td>Location of touch</td>
</tr>
<tr>
<td>Head/Face Touch occurs to infant’s head and/or face; above the neck.</td>
</tr>
<tr>
<td>Limb Touch occurs to infant’s hand, arm, shoulder, feet and/or leg.</td>
</tr>
<tr>
<td>Torso Touch occurs to infant’s front, back and/or side in between the hips and neck.</td>
</tr>
<tr>
<td>Mixed Touch simultaneously occurs across two categories for location.</td>
</tr>
</tbody>
</table>
Results

Descriptive Analyses

In total, 5232 usable utterances across participants were coded for topic and touch. The average number of useable utterances per caregiver-infant dyad was 262 (91-436). Seventy-nine percent of the utterances were about an object and 21% were about an abstract topic.

Regarding touch, 1461 instances of touch were recorded, meaning during 28% of utterances there was some form of caregiver-infant touch. The average number of utterances that involved touch per caregiver-infant dyad was 74 and the mean proportion of utterances that coincided with adult utterances was 27% (SD = .20). Sixty-five percent of utterances that involved touch included at least one nonsupport touch point (946 instances) and 43% of utterances involved at least one touch point that provided some form of physical support to the infant (622 instances). Since adults could touch infants with more than one hand or with their face or with objects, there were also instances where infants were simultaneously experiencing support and non-support touching during one utterance. However, this was relatively rare (4% of touches for which this could be assessed†). Of the support touches, 30% provided balance to the infant and 70% involved the infant being held.

Eight hundred and ninety-three occurrences of utterances with nonsupport touch were coded for location. This total excludes 53 instances of touch for which coders could not specify location. Thirty-five percent of touches were to the infant’s torso, 32% were to the limbs, 24% involved the infant’s face and/or head. Nine percent of utterances that co-occurred with non-support touch simultaneously occurred across two categories for location (e.g. limb and torso).

Final Analyses

To account for the statistical dependence of utterances by the same caretaker, a GEE (Generalized Estimating Equations) repeated measures model was fit for all analyses using the exchange correlation structure. Various aspects of touch were tested as predictors of topic (Object or Abstract). Because there are reports of adult touch to infants varying by both age (Ferber et al, 2007; Jean, Stack & Fogel, 2009; Kaye & Fogel, 1980; Koester, Brooks & Traci, 2000) and sex (Stepakoff, 2000) of the baby, infant sex and age category (using a median split) were considered for

† Of the 1461 utterances that included some type of touch, there were 51 utterances for which coders could not determine if both hands were engaged in support touch or not so they were excluded from this particular analysis.

Cognition, Brain, Behavior. An Interdisciplinary Journal
15 (2011) 427-448
inclusion as covariates in the final analyses, but neither variable significantly predicted either instances of touch nor topic of interaction, so these variables were dropped from the final models‡. The first analysis was to learn if the presence of any caregiver touch during an utterance was related to the topic of conversation. Touch was significantly associated with topic ($\chi^2(1)=10.14, p<0.002$) with caregivers being more likely to touch infants during speech about abstract topics than during speech about objects in the immediate environment and this pattern held across all of the specific abstract topics. Infants were also more likely to be held ($\chi^2(1)=5.3, p<0.02$) or held-up ($\chi^2(1)=6.6, p<0.01$) during speech about abstract topics than during speech about objects. See Figure 1.

**Figure 1. Probability that an utterance is about a nearby object given presence or absence of any touch or support.**

‡ Neither age category ($\chi^2(1)=0.34; \text{n.s.}$) nor infant sex ($\chi^2(1)=0.08; \text{n.s.}$) significantly predicted whether the infant was being touched. Age category ($\chi^2(1)=1.6; \text{n.s.}$) and sex ($\chi^2(1)=0.5; \text{n.s.}$) were also unrelated to the topic of conversation. All analyses were run twice, including age and sex as covariates and without these covariates. In no analysis did either age or sex significantly predict the topic of conversation nor modify the effect of the independent variable being assessed, so the analyses presented in the paper are from models that do not include these two covariates.

The actual number of distinct touch points an infant was experiencing also predicted topic, with more touch points associated with abstract topics and fewer touch points associated with utterances about objects in the immediate vicinity ($\chi^2(1)=10.12, p<0.002$). An additional analysis investigated whether the number of different types of touch points, support or non-support predicted topic of conversation, and the pattern was the same for both types. Regardless of whether the touch point was providing support ($\chi^2(1)=8.9, p<0.003$) or not providing support ($\chi^2(1)=8.0, p<0.005$) the more touch points of any type that were present, the less likely it was that the concurrent utterance was referring to an object. In fact, if an infant was experiencing three touch points, the probability of that utterance being about an object was only 12%. See Figure 2. Overall, touch or its absence predicted the type of referent of adult speech.

**Figure 2. Probability that an utterance is about a nearby object varies by number of simultaneous touch points**

In order to assess the possible meaning of various patterns of touch location, a separate analysis was conducted including only utterances that co-occurred with at least one non-support touch point. Location of touch did not significantly predict topic of conversation ($\chi^2(3)=2.8, p>0.43$).

**Discussion**

This study has two significant results. First, infants are more likely to be touched, particularly held or held-up when adults speak about topics for which there are no immediately useful physical referents (other than themselves) than when they speak about objects that are present in the immediate environment. Second, infants are likely to experience more simultaneous touch contact points when adults are
speaking about a topic that does not have an immediate physical referent than when speaking about a nearby object. The location of touch did not vary by topic. Clearly, adults’ haptic input to infants varied with the topic of conversation. When speaking about an object in the immediate vicinity, adults touched and held infants much less than when speaking about a topic for which there was no physical referent available.

Hereafter we discuss two possible effects of the differential patterns of touch associated with speech about objects and more abstract topics. We hypothesize first that haptic input could affect infant attention to various aspects of the context and second, that haptic input could facilitate multimodal coordination. We review the literature supporting those interpretations and address the limits of our experiment.

**Touch cues two broad classes of referents**

Haptic input could serve as a cue to distinguish two broad classes of referents: internal (emotional, abstract) and external referents (objects) the caregiver is likely speaking about, and could help infants to begin to disambiguate the referential intent of the accompanying language. Touch cues could be a part of adult “scaffolding,” or framing, of infant experiences to support infant’s eventual acquisition of linguistic competence.

How might this happen? In a recent study, mothers of infants under the age of one reported consciously using touch to communicate love (emotions), to complete practical tasks (e.g., diaper changes - object manipulation) and to support and position the infant (e.g., holding) (O’Brien & Lynch, 2010). Mothers did not report consciously using touch to affect infant attention or to communicate linguistic meaning. However, adults do not need be aware of their consistent use of touch to signal the topic of conversation for it to be a meaningful cue to infants. Hertenstein (2002) has argued that infants abstract meaning from caregiver touch whether caregivers are intending to communicate haptically or not.

We propose that the adult caretakers of the typically developing babies in this study were using touch to orient infant attention similarly to how deaf parents used tapping to orient infant’s attention to the signing hands (Waxman & Spencer, 1997). Adults routinely adjusted their haptic input to the infant according to the topic of conversation which might have been an attempt to manage infant attention and to constrain the number of possible referents infants might associate with the language being used. Perhaps, by touching the infant, the adult drew the infant’s attention to focus on the caregiver and away from scanning the immediate environment. The fact that adults were less likely to touch an infant while speaking about a nearby object could have also resulted from the adults not wanting to distract the infant from attending to the immediate surroundings and the object or
objects that are being spoken about or from the adults being likely to touch the objects being spoken about, thereby reducing the amount of touch they are giving the infants. Whether this pattern was an attempt to affect infant attention, or a by-product of adult’s handling objects while speaking about them, adults were literally making some of the meaning of the language physical for the infants. “Attend to the environment ” or “Do NOT attend to the surroundings” was communicated through the haptic input. The message was literally “embodied” for the infant.

The results of this investigation show that during naturalistic interactions, caregivers are providing consistent haptic cues that could scaffold infant’s understanding of the interaction by helping to guide the infants’ attention to the aspects of the environment most appropriate for beginning to understand the language being used. This interpretation is congruent with what little prior research exists on the effect of adult touch on infant attention in the first year of life. For example, Fogel and colleagues have described mothers using touch to attract the attention of 6- & 13- week-old infants during naturalistic dyadic interactions (Kaye & Fogel, 1980). Other researchers have documented that adult touch increases infant visual attention to another person (Gusella, Muir & Tronick, 1988; Stack & Muir, 1990; 1992) and to pictures of objects (Arditi, Feldman & Eidelman; 2006). So, there is evidence that touch can affect infant attention, but exactly how it does so in naturalistic dyadic interactions and how these interactions change over time remains to be elucidated.

Our interpretation is also consistent with Baldwin & Saylor’s (2004; Saylor & Baldwin, 2004) finding that infants attended to a person speaking about an absent caregiver rather than to other aspects of the experimental situation or environment during that speech. In that instance, the researchers interpreted this “looking at speaker” behavior as a means of establishing joint attention about an absent referent. Those infants ranged between 15 & 30 months, when joint attention is well established.

In any case, it seems plausible to claim that haptic input could be helping to focus infant attention to various aspects of the environment by serving to “point” infant attention in the direction of the topic of conversation. Our results show that this pattern is present with infants as young as six months of age, before the age at which infants can reliably engage in joint attention with adults (Butterworth, 2000; Butterworth & Ikatura, 2000; Carpenter, Nagell & Tomasello, 1998). Perhaps early haptic input helps to scaffold the eventual attainment of participation in “joint attentional scenes” by creating contexts early in development where infants and adults are more likely to be visually attending to the same aspect of the environment, in this case, the appropriate referent for what the adult is speaking about.

We suggest that for infants in the second half of the first year of life, haptic input could serve as part of a general “frame” for communicating two broad classes of referents of speech, allowing infants to adjust their attention accordingly.
This attentional shift could affect what aspects of the environment infants choose to further explore visually or manually and make it more likely that infants are attending to the correct referent of adults’ speech.

**Touch helps multimodal coordination**

Another effect of haptic input that varies with the topic of conversation could be to help the infant coordinate information from other modalities. If haptic input from adults makes it more likely that infants are attending to the actual referent of adult speech during that speech then infants will be more likely to be **visually** attending to the object while hearing **auditory** input about it. Basically, the haptic information could be serving to align input from the visual and auditory modalities so that the infants experience more instances of intersensory redundancy, a particularly powerful learning situation for infants (Bahrick, Flom & Lickliter, 2002). As adults interact with infants, they are providing rich multimodal and highly structured contexts full of meaningful information communicated visually, haptically and auditorily, and there is growing evidence that these sorts of rich, multimodal contexts are important for infant language learning (Gogate & Hollich, 2010; Smith & Sheya, 2010).

In future work it would be interesting to examine the link between speech referent and haptic input for dyads speaking other languages and in other cultural contexts as descriptive studies of parent-infant dyadic interactions have documented unique interaction patterns of voice, gaze and touch in different cultures (e.g., Feldman, Masalha & Alony, 2006).

The results of this study make clear that adults are providing meaningful patterns of haptic input that infants could use to learn about the referents of adult speech. However, we do not yet know how infant attention and verbal processing are actually affected by these cues. Future studies that vary haptic input while directly assessing infant attention to objects and people in their immediate environment are needed. Future work could also examine whether the addition of haptic cues helps infants to align information from other modalities during naturalistic social interactions. This is an area ripe for further investigation.

**ACKNOWLEDGEMENTS**

We would like to thank all of the parent and infants who participated in this study as well as the many student research assistants who helped to collect and code the data. We would also like to thank the Grand Valley State University (GVSU) Psychology Department and the GVSU Research and Development committee for funding this work.
REFERENCES


G. L. Dueker, S. Portko, M. Zelinsky


THE BODY REGION CORRELATES OF CONCRETE AND ABSTRACT VERBS IN EARLY CHILD LANGUAGE

Josita MAOUENE1*, Nitya SETHURAMAN2, Aarre LAAKSO2, Mounir MAOUENE3

1Psychology Department, Grand Valley State University, USA
2Department of Behavioral Sciences, University of Michigan-Dearborn, USA
3Department of Computer Science, Ecole Nationale des Sciences Appliquees, Tangier, Morocco

ABSTRACT

Literature on verb acquisition has mainly focused on lightness and concreteness for verb acquisition whereas the analysis presented here points to an embodied perspective on word learning, examining whether early-learned verbs are associated with distinct parts of the body, and whether early-learned verbs might be more or less associated with parts of the body depending on whether they are abstract or concrete. Three dimensions of abstractness-concreteness were considered: semantic lightness, imageability, and number of associated objects. These factors were also examined in light of the age of acquisition and frequency of the verbs. This study presents evidence that embodiment is an important factor to consider in children’s acquisition of verbs, with two major findings: 1) those verbs that are more highly associated with a single region of the body, according to adult judgments, are among the first acquired in vocabulary development; 2) associations with a specific region of the body may be a better predictor of whether a verb is learned early than the concreteness or abstractness of the verb. Taken together, these results suggest that we should further investigate the role of bodily experiences in verb meaning acquisition.

KEYWORDS: body regions, verbs, concrete, abstract, corpus study

* Corresponding author:
E-mail: maouenej@gvsu.edu
INTRODUCTION

Recently, embodiment, the notion that the body and its morphology are partners in higher cognitive processes, has been examined in the study of verbs, both in adult usage and in children’s acquisition. The importance of body parts in learning and processing verbs has been shown by Maouene, Hidaka & Smith (2008), who find that adults systematically and coherently associate 100 early-learned verbs (Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994), a list that includes both abstract and concrete verbs, with five body regions: eye, ear, mouth, leg and hand. In the current study, we follow up on these results by asking whether body regions are more strongly associated with abstract verbs or concrete verbs.

The organization of the introduction is as follows. First, we review the literature supporting the idea that body parts are important in verb learning and processing. Second, in two sections, we review the literature arguing that 1) abstract cues and 2) concrete cues are important for learning and understanding verbs. Finally, we return to the specific questions of the current study, which asks whether body parts are equally important for both abstract and concrete verbs by examining abstractness-concreteness using three definitions: light versus non-light; high imageability vs. low imageability; and associated with many object types vs. associated with few object types.

Body parts in verb learning

Body parts have very interesting concrete and abstract characteristics in the two types of environments that matter for language acquisition, the observable world and the linguistic world. In the observed socio-cultural and spatio-temporal world of events, body parts transport us, create art, manipulate tools, possess objects of different sizes, feed us, sense danger, and connect us to others. In many ways, body parts are “causal” and are central to the “intentionality” of many events. In the linguistic world, body parts are represented by nouns and are explicit or implicit semantic components of verbs’ relational meanings (Bowerman, 2005; Bowerman & Brown, 2008). As nouns, body parts can play many syntactic roles, not only subjects and direct or indirect objects but also obliques (such as instrumental and locative). Even as nouns, the number of terms referring to body parts is quite restricted, particularly in young children, making them close to a closed class of words. The presence of body parts is covert in proper nouns (Tom, Sheila), in labels for humans (girl, boy), and in animals, all categories that are very interesting to young children. Finally, body part terms are also apparent in prepositions (back, front) and in objects (the leg of the table, the bottom of the page).

We suspect that the importance of body parts might be scaffolded by the fact that children initially begin learning through a physically engaged body (Piaget,
1953; Thelen & Smith, 1994), from the child’s own perspective (Huttenlocher, et al., 1983), and because developmentally, intelligence resides to a large extent in our own sensory-motor system (Piaget, 1953). Body parts may also be an important early bootstrap for general learning due to many factors, including that (a) for a child, her body is a constant frame of reference (Varela, Thompson & Rosch, 1983), long before productive language kicks in (Tincoff & Jusczyk, 2011); (b) body parts include parts and whole (see for example the importance of part-whole structure in ontologies and perception) and include a ubiquitous spatial organizing principle (de Leon, 1994); and (c) the child’s body is excellent at capturing regularities or patterns, such as auditory patterns in newborns (Gervain, Macagno, Cogoi, Peña, and Mehler, 2008) and infants (Saffran, Aslin, & Newport, 1996) and eye-movement patterns in infants (Smith & Yu, 2010). It would therefore seem very plausible that, early on, the regular way regions of the body get used in activities would contribute to the scaffold of early understanding and use of relational meaning, both abstract and concrete.

Recently, embodiment, the idea that the body may shape knowledge because it stands between the world and the mind (for reviews on embodiment, see, for example, Barsalou, 2008; Glenberg, 2010; Wilson, 2002; Ziemke, 2001; and for its mechanistic challenges, see Pezzulo, Barsalou, Cangelosi, Fischer, McRae & Spivey, 2011), has attracted interest in the study of verbs. In adults, event-related potential studies and fMRI studies have shown that reading a verb (e.g., kick) activates the cortical motor areas relevant to moving the appropriate body part (e.g., leg and foot) (Hauk, Johnsrude, & Pulvermüller, 2004; Pulvermüller, Lutzenberger, & Preissl, 1999; also see Boulenger, Roy, Paulignan, Deprez, Jeannerod, & Nazir, 2006). Behavioral studies also suggest a connection between verb processing and movements made by particular parts of the body. For example, moving the arm away from the body slows judgment about the sentence “Open the drawer”, an action involving the movement of the arm toward the body (Glenberg & Kaschak, 2002). Such results support the idea that the processing of verb meanings may involve or interact with some of the same processes that generate bodily action (Barsalou, 1999, 2003).

In children, James & Maouene (2009) find that merely hearing verbs but not hearing neutral adjectives significantly recruits regions of the premotor and primary motor cortices in four-to-six-year-olds. The results also suggest that the activation might be effector-specific: that is, verbs associated with hand movements activate different regions of the frontal cortex than verbs associated with leg movements, although some overlap is also possible. Further, James and Swain (2011) have shown that, in children, hearing novel verbs used while actively exploring objects but not while passively watching exploration of objects creates sensory-motor systems in the developing brain. Similarly, in a behavioral study, young children were shown to comprehend individual verbs in the context of their own actions but not in the context of the actions of others (Huttenlocher, Smiley, &
Chaney, 1983; see also Piaget, 1953; but, see Rizolatti and Craighero, 2004; Childers & Tomasello, 2006). These results argue for an early connection between verbs and motor regions and suggest that effectors are important in language learning.

The current study follows from Maouene, et al. (2008), in which it was argued that specific body region associations are important in the early development of verbs. The current study goes further to ask whether body part associations are more important for some kinds of verbs than others. Specifically, we ask whether body part associations are stronger for abstract verbs (and particularly a subgroup called light verbs or general all purpose verbs) versus concrete verbs (and particularly, highly imageable verbs and verbs associated with few object types or heavy verbs), and what relationship body parts bear with abstraction and concreteness, two fundamental concepts in verb learning. Hereafter we review the theoretical and empirical findings on the essential distinctions of abstract verbs and concrete verbs and potential routes for their connections with body region specificity.

Abstract cues in verb learning

Prior research has influentially argued that for children to learn the relational meaning of verbs, concrete, observational cues are not helpful. Instead, children rely heavily on abstract cues, particularly linguistic cues (Gillette, Gleitman, Gleitman & Lederer, 1999; Gleitman, 1990; Snedecker & Gleitman 2004). Further, it has been argued that infants toward the end of their first year can abstract the concepts found in the relational meanings of events under the form of schemas of containment and support, source, path, goal, causation, agency, etc. (see Hirsh Pasek & Michnik Golinkoff, 2006; see Mandler, 1992, for a review of experimental studies), supporting the idea that children start the process of verb learning at a general and conceptual level.

One particular class of abstract verbs that has garnered a lot of attention in the literature is the set of verbs called “light” verbs or general all purpose verbs (GAP). As there is no general agreement on what the list should encompass, we incorporated in our list those verbs that researchers have talked about as GAP verbs (e.g., Clark, 1978; Pinker, 1989; Theakston, Lieven, Pine, & Rowland, 2004) that overlap with the verbs that appear much later in development in a construction known as the light verb construction. In this particular construction, verbs which are suggested by some to be "semantically light" appear with eventive nouns such as have a rest, a read, a cry, a think, take a sneak, a drive, a walk, a plunge, give a sigh, a shout, a shiver, a pull, a ring, make a call, waste, a plan, put the shot, the question,
blame, bring snow, the charge, comment, etc. In English, light verbs – and GAP verbs† – have been argued to have a special status by theorists of English verb acquisition (Clark, 1978; Goldberg, et al., 2004; Pinker, 1989; Theakston, Lieven, Pine, & Rowland, 2004). Because they are highly frequent, semantically general, and among the earliest produced English verbs, it has been proposed that light verbs serve an important role in verb and grammatical learning. Goldberg et al. (2004) found that several light verbs were the most frequent verbs used in particular syntactic patterns. By paying attention to this statistical regularity, children learn to associate the light verb with the syntactic pattern and learn an abstract schema that then facilitates the acquisition of many verbs that encode the same underlying causal and argument structure. However, see Ninio (1999a, 199b) for data showing that children use a mixture of general and specific verbs in the simple transitive and simple intransitive constructions right from the beginning, which suggests that light verbs may not have a special status in very early language development.

Clark (1978) calls light verbs general purpose verbs in opposition to specific action verbs and notes that their use comes after children’s even earlier use of also highly abstract particles like up, away, and off. Clark suggests that children first produce light verbs which are very general, and later replace them with more specific verbs; for example, do may be replaced by build, cut, unwind, and go by run, drive, walk. Similarly, Pinker (1989) suggests that the relational meanings of light verbs make them the core meanings of other heavier verbs to which other more specific meaning elements are added. According to Pinker, the relational structures of the light verbs reflect primitive and innate semantic elements. The implication would seem to be that light verbs are early precisely because they are light, general, and frequent.

In this paper, we particularly examine whether light verbs have a special status with respect to body part associations. In the next section, we examine how concreteness is also an important factor for verb learning and examine the connection between concreteness and body part associations.

Concrete cues in verb learning

Although abstract cues are important for verb learning, concreteness has been proposed as an explanation for why some verbs are learned before others. Here we examine two specific aspects of concreteness, imageability and association with fewer object types.

Imageability refers to ‘the ease with which a word gives rise to a mental image’ (Paivio, Yuille & Madigan, 1968) – cited in McDonough, Song, Hirsh-† Hereafter we will refer to them as light verbs although we mean an overlap between verbs appearing in the light verb construction and general all purpose verbs.
Pasek, Golinkoff, & Lannon, 2011. Imageability has been correlated with concreteness (Bird, Franklin & Howard, 2001) and has been related to a semantic notion like boundedness (Langacker, 1987): e.g., the bounded verbs jumping, running, and eating have clear beginning and end points, but the verb believing does not. It has been suggested that highly imageable words that label “perceptually accessible” concepts (a term borrowed from Gentner, 1982, 2006) are produced before those that require additional support from social or linguistic cues (Maguire, Hirsh-Pasek, & Michnick Golinkoff, 2006; Michnick Golinkoff & Hirsh-Pasek, 2008). We examine here whether verbs that are highly imageable are more strongly associated with body regions.

A second aspect of concreteness in verbs that we examine is the number of associated object types. Some authors have suggested that there is a transition from an early-restricted to a more widespread use of verbs (Akhtar & Tomasello, 1997; Chan, Meints, Lieven, & Tomasello, 2010; & Risley, 1995; Tomasello, 2003), assisted by very narrow, context-dependent (concrete) meanings. In particular, Maouene et al. (2011) have proposed to use one metric by which the ‘lightness’ or ‘heaviness’ of early-learned verbs might be measured: the number of objects with which they are associated (in adult judgment) or co-occur (in speech to and by children). The results suggest that early-learned light verbs and heavy verbs differ in the breadth of the objects they are associated with: light verbs have weak associations with specific objects, whereas heavy verbs are strongly associated with specific objects. Hereafter we examine whether verbs associated with fewer object types are more strongly associated with body regions than verbs associated with more object types.

THE CURRENT STUDY

A prerequisite to a claim that body parts are important in connecting world knowledge and linguistic knowledge would be to examine whether the associations with particular body regions that were collected by Maouene et al. (2008) are equally strong on average for the majority of early-learned verbs—both concrete and abstract—or whether these associations are stronger on average for concrete verbs versus weaker for abstract verbs, or whether a more differentiated landscape emerges where the associations with body regions are strong for different groups of concrete and abstract verbs.

We compare verb/body region associative strength across three dimensions of concreteness-abstractness: light or non-light verb meanings (listed in Theakston, Lieven, Pine, J. & Rowland, 2004), less imageable and more imageable verb meanings (adult ratings given in Cortese and Fugget, 2004), and verbs used with many object types or fewer object types (listed in Maouene, Laakso & Smith, 2011). We also examine how these associative strengths are modulated by verb
frequency (obtained from CHILDES) and normative age of acquisition (from the Bates-MacArthur CDI: Fenson et al., 1994).

We predict that, children’s acquisition of verbs is influenced by their tie to the body such that those verbs that are more highly associated with a single region of the body (according to adults), whether concrete or abstract, are among the first acquired in vocabulary development. Further, association with a specific part of the body may be a better predictor of whether a verb is learned early than lightness or concreteness.

**METHOD**

Two analyses compare the strength of the association with the main body region for early-learned abstract verbs and concrete verbs. The metrics used for both analyses were the percentage of agreement among adults on a main body region with 100 early-learned verbs as given in Maouene, et al. (2008). These were obtained by asking 50 adults in an associative task to name the single body part that came to mind when they thought of each verb. For example, 100% of the participants agreed that the verbs *hear* and *listen* connect with the ear region, and 98% of the participants agreed that the verb *bite* connects with the mouth region (comprising teeth, tongue, mouth). Thus, *hear*, *listen* and *bite* have strong associations with a single main body region. By contrast, 66% of the participants agreed that *drive* connects with the hand region (comprising finger, hand, arm). Thus, *drive* has a relatively weak association with a main body region, meaning that it was associated with more than one body region across participant responses. Hence, we refer to this metric as “body region specificity” for the remainder of the paper. The regions had been determined by a dimensionality reduction technique (correspondence analysis). These are the percentages of agreement we have used here for both analyses. In our analyses, we also used the frequency of use of these 100 verbs in the productive vocabulary of children from the CHILDES database (http://childfreq.sumsar.net/, Rasmus Bååth, 2010) and the normative Age of Acquisition from the Bates-MacArthur CDI (Fenson et al., 1994) for these 100 verbs.

The first analysis looks particularly at the average body region specificity for the class of verbs named light (a subset of abstract verbs) and whether it differs for eight groups of non-light verbs, both concrete and abstract, using production frequency as the criterion to define the different groups of verbs. The second analysis examines the average body region specificity for two additional dimensions found to correlate with concreteness: imageability and number of object types. The goal was to see whether highly imageable (concrete) verbs differ from non-highly imageable (abstract) verbs in their body region specificity and whether, similarly, verbs with a low number of syntactic object types (concrete) differed
from verbs with a high number of object types (abstract) on the same measure. Further, the data on body region specificity and the data on number of object types were rescaled so that they would fit on a scale from 1 to 7 (precisely the scale used for imageability judgments in Cortese and Fuggett, 2004) so as to be able to compare the distributions of the verbs between the three dimensions along concrete to abstract.

**Verb list**

The subset of verbs used were those listed in the Bates-MacArthur Communicative Development Inventory for American English (Fenson et al., 1994). This inventory (built from a normative study of over 1800 children) includes a list of 102 verbs that occur normatively in the productive vocabulary of at least 50% of children learning American English by 30 months of age. This list does not include the verbs such as *do*, *want*, *can*, *need*, *must*, which were categorized as “helping” verbs under another section of the MCDI. For Analysis I, all the verbs were used except two (*stay* and *tear*, verbs for which there was no data on body regions from Maouene et al., 2008). For Analysis II, 90 verbs from the MCDI were used for which there were overlapping ratings in imageability, body region specificity and number of object types (listed in the Appendix).

**Body parts-verbs associations list**

The body part-verb associations that underlie the body region specificity metric come from 50 undergraduate students from Indiana University whose first language was American English. They were asked to name the first body part that came to their mind in association with each of the 102 common early-learned verbs mentioned above. The list with the percentage of participants who agreed on a particular body region with a particular verb is published as an appendix in Maouene, et al. (2008). For Analysis I, we used the percentage of agreement for five regions (eye region, mouth region, ear region, leg region and hand region) as a measure of body region specificity. The associative strength of the verbs *swing* (60%), *ride* (66%) and *sit* (98%) were increased in the leg region, because we incorporated *bottom* and *butt* since it can be argued that these parts are physiologically situated in the lower part of our anatomy. For verbs that did not fit in one of those five body regions, we used the maximum agreement for one body part: for example, *fit*: 26% whole body, *love*: 26% heart, *think*: 40% mind, etc. (people have offered brain and mind as body parts and we have respected that, since we can argue that they are effectors too).
List of concrete versus abstract verbs

Two lists were built to analyze the concreteness versus abstractness of verbs: (1) A first list composed of 8 light verbs and 9 non-light verbs from the definition given by Jespersen (1954) and the list reported in Theakston, et al., 2004 (taken from Clark, 1978 and Pinker, 1989) that overlap with the list of 102 verbs from the Bates-MacArthur CDI, for which body region specificity, age of acquisition and frequencies for infants younger than 24 months in CHILDES were available (Table 1). (2) A second list includes the 90 verbs for which both imageability ratings on a 7-point-scale (Cortese and Fugget, 2004) and number of object types (Maouene et al., 2011) were available that overlap with the list of verbs from the Bates-MacArthur CDI for which body region specificity, age of acquisition and frequencies for infants younger than 24 months in CHILDES existed (see Appendix). We used Cortese & Fugget (2004) rather than Masterson & Druks (1998) because Cortese & Fugget’s list contains more than twice the number of verbs (90 verbs against 44 verbs that overlap with the MCDI); see, for example, the analysis in Ma, Michnik Golinkoff, Hirsh-Pasek, McDonough, & Tardif (2008) on the imageability of 44 English verbs.

Verb frequency

The frequency of occurrence of the verbs in productive child language from the CHILDES database was used with the help of an online tool called ChildFreq (created by Rasmus Bååth, 2010). ChildFreq searches the American and British parts of CHILDES, which consist of approximately 5,000 transcriptions, in total ~3,500,000 words. We used the total number of occurrences of the verbs per 1,000,000 words from 12 to 23 months as our criterion (this projection helps compare the frequencies at different ages, because the number of transcripts varies from one age group to another). For our particular age group, it is based on 807 transcripts and 298,299 uttered words.

Age of acquisition norms

We used the norms of the children’s productive vocabulary from the Bates-MacArthur Communicative Development Inventory: Toddler version (Bates-MCDI; Fenson et al., 1994). We took the age of acquisition for a verb to be the first month in which the verb is produced by more than 50% of the children.
RESULTS

This section is organized into two major subsections—one for Analysis I and one for Analysis II. Each of these two analyses was really a set of related analysis, so the subsections are divided as well. The subsection on Analysis I includes sub-subsections for Analysis I-A – Analysis I-H. Similarly, the subsection on Analysis II includes sub-subsections for Analysis II-A – Analysis II-H. The description of each of these individual analyses is divided into four parts. In the first part (Question), we elaborate on the theoretical question at stake in that particular analysis. In the second part (Method), we describe any unique features of the method for that particular analysis not already covered in the overall Method section above. In the third part (Results), we report the results of the particular analysis. In the last part (Implications), we briefly discuss the meaning and significance of the results. There are also Discussion sub-sections at the end of Analysis I and at the end of Analysis II.

Analysis I: Body Region Specificity and Lightness

The first set of analyses (collectively: Analysis I) examines whether body region specificity for the class of light verbs (a subset of abstract verbs) differs from the body region specificity of various groups of non-light verbs (both concrete and abstract). In these analyses, production frequency is used as the criterion to define the different groups of verbs. The first analysis (Analysis I-A) compared the mean body region specificity of the eight light verbs to that of eight non-light verbs where frequency was held constant. The remaining analyses (Analysis I-B – Analysis I-H) compared the mean body region specificity of the eight light verbs to the mean body region specificity of eight different groups of verbs defined by frequencies of occurrence in CHILDES for 12- to 23-month-old infants. In these analyses, we also checked whether each group of verbs by frequency differed significantly from the light verb group in terms of age of acquisition based on parental report (Fenson, et al., 1994). In the following analyses, none of the 96 verbs present in the speech of 11-to 23 month-olds were analyzed twice.

Analysis I-A: Light verbs compared to most frequent non-light verbs

Question. Our overall question in Analysis I was whether the body region specificity of light verbs differs from that of non-light verbs. Thus, the first analysis simply compares the mean body region specificity for the eight light (abstract) verbs to the mean body region specificity for the eight most frequent concrete early-learned verbs.
Method. We used the total number of occurrences per 1,000,000 words from 12 to 23 months as our criterion of frequency. These occurrences appear in brackets for the 16 verbs listed below. For the light verbs, the list includes make (838), put (2903), take (1334), have (1890), go (5547), give (606), get (2701) and bring (150). The most frequent verbs that are both non-light and non-abstract did not include like, help or play because these are semantically general. The list includes: close (1024), drink (811), eat (1521), look (2933), open (1813), read (1441), see (3121) and sit (1884).

Results. A t-test indicates that the mean frequency of occurrence of the light verb group, $M=1996$, $SD=1734.5$ did not differ from the mean of the concrete verbs with the highest frequencies, $M=1818.5$, $SD=829.8$, $t(14)=0.26$, $p=.79$. Furthermore, the mean body region specificity for the eight light verbs, $M=89.3$, $SD=9.8$, does not differ significantly from the mean body region specificity for the eight non-light verbs, $M=90.0$, $SD=10.5$, $t(14)=0.14$, $p=.88$. However, the two types of verbs differ significantly in terms of their age of acquisition: light verbs, $M=24.12$, $SD=2.6$, and non-light verbs, $M=21.5$, $SD=2.0$, $t(14)=2.24$, $p=.042$.

Implications. The most frequent concrete verbs are approximately as frequent as the light verbs. Moreover, the body region specificity of the concrete verbs is approximately the same as that of the light verbs. However, the eight light verbs appear later than the eight most frequent concrete verbs in the productive vocabulary of this sample of young children. In other words, eight highly concrete verbs with high frequency are acquired earlier in CHILDES (downloaded June 2011), than the eight light verbs of equally high frequency usually proposed as quite abstract verbs (Pinker, 1989; Clark, 1978; Theakston et al., 2004, etc.). Interestingly, these two groups did not vary significantly in their average body region specificity.

Analysis I-B: Light verbs compared with low-frequency verbs

Question. In Analysis I-B – Analysis 1-H, we examined the relationship between frequency, body part specificity and age of acquisition. In Analysis I-B in particular, we wanted to know whether the eight light verbs differed from low-frequency non-light verbs in terms of body region specificity or age of acquisition.

Method. For Analysis I-B – Analysis 1-H, we began by grouping the non-light verbs by frequency such that each group has a similar number of verbs. We divided the non-light verbs into seven groups based on production frequencies (occurrences per 1,000,000 words) at 12 months. Two groups contained verbs that had fewer than 100 occurrences per 1,000,000 words. These two groups differed in the pattern of frequency of their verbs at later ages. One of them contained verbs that remained infrequent, whereas the second contained verbs that—although initially infrequent—rose in frequency over time. A third group had verbs that occurred
between 100 and 200 times per 1,000,000 words. The fourth group had verbs that occurred between 200 and 300 times per 1,000,000 words. The fifth group had verbs that occurred between 300 and 400 times per 1,000,000 words. The sixth group had verbs that occurred between 400 and 600 times per 1,000,000 words. Finally, the seventh group contained the remaining verbs with more than 600 occurrences per 1,000,000 words. For all these, we also computed the relation of each group to age of acquisition.

Analysis I-B specifically compared the eight light verbs and the 12 verbs that have the lowest frequency (fewer than 100 occurrences per 1,000,000 produced words) between 12 and 23 months and that stay below a hundred occurrences in child speech until 60 months. These verbs are: clap (53), feed (77), lick (30), rip (0), shake (67), share (13), skate (0), chase (100), smile (16), spill (53), splash (73) and sweep (67).

Results. The mean body region specificity for the eight light verbs, \( M=89.3, SD=9.8 \), does not differ significantly from the mean body region specificity for the twelve concrete and later-learned verbs (from CHILDES), \( M=86.12, SD=12.35 \), \( t(18)=0.62, p=.54 \). And again, the two groups differ significantly in terms of age of acquisition, but this time, the light verbs are earlier acquired, \( M=24.12, SD=2.6 \), whereas the verbs that are less frequent at subsequent ages are acquired significantly later, \( M=27.33, SD=2.1 \), \( t(18)=-3.02, p=.007 \).

Implications. Light verbs like make, put, go, etc., with the highest frequencies (range between 800 and 5447 occurrences per 1,000,000 words) are acquired significantly earlier than a group of concrete verbs with very low frequencies (below 100 occurrences). This result supports the idea that frequency plays a major role in verb acquisition (Goodman, Dale & Ping, 2008).

Analysis I-C: Light verbs compared with verbs with increasing frequency

Question. We wanted to know whether the comparison to light verbs fared any differently among low-frequency verbs with increasing frequency at later ages compared to low-frequency verbs that remain low at later ages.

Method. The third comparison compared the eight light verbs to the ten verbs that have the lowest frequency (fewer than 100 occurrences per 1,000,000 words) at 12 to 23 months and which occur more and more frequently at 24 months and 36 months. These verbs are: break (54), cook (92), finish (77), hate (0), hit (67), hear (13), hurry (13), pour (11), pretend (97) and wish (26).

Results. Here, interestingly enough, the results indicate that the body region specificity for the eight light verbs, \( M=89.3, SD=9.8 \), differs significantly from the body region specificity for the 11 later-learned verbs, \( M=68.8, SD=28.73 \), \( t(11.49)=2.10, p=.04 \) (two-tailed, equal variance not assumed, Levene’s test is significant, \( F=27.29, p < .01 \)). And here, the two groups do not differ significantly in
terms of age of acquisition, but there is a marginal tendency for the light verbs to be acquired earlier, $M=24.12$, $SD=2.6$, over the infrequent verbs, some concrete some abstract, $M=26.7$, $SD=2.1$, $t(16)=-1.91$, $p=.074$.

**Implications.** Among the groups of verbs with the lowest frequencies, the subsample of 10 verbs with low frequencies that are later-learned (a subgroup of the “new comers” in CHILDES) was not well learned. Moreover, these verbs differ significantly from the light verbs in that they had, on average, lower body region specificity, that is, their meaning related to more than one main body region.

### Analyses I-D – I-H: Light verbs compared to more frequent verbs

**Question.** Having examined the relationship between light verbs and two groups of low-frequency verbs in Analysis I-B and Analysis I-C, we were also interested in comparing light verbs with the remaining five groups of relatively high-frequency verbs in terms of body part specificity and age of acquisition.

**Method.** The method was the same as used in Analysis I-B and I-C, but these analyses each used one of the remaining five groups of verbs by frequency.

**Results.** The next four groups did not differ significantly in terms of mean body region specificity or age of acquisition. However, the last group, which has the highest frequencies, *draw* (660), *help* (693), *push* (737), *fix* (750), *fall* (848), *find* (848), *like* (1317) and *play* (1880) ($M=966.6$, $SD=423.2$) per 1,000,000 words showed a marginally significant difference when compared with the light verbs. The mean body region specificity for the eight light verbs, $M=89.3$, $SD=9.8$, differed marginally from the mean body region specificity for the eight verbs with the highest frequencies: $M=69.00$, $SD=27.8$, $t(14)=1.93$, $p=.074$. Table 1 summarizes the tests performed in Analysis I-A – Analysis I-H.

**Implications.** There is a (marginally significant) tendency for the group of verbs with frequencies in the 600 and up per 1,000,000 to differ from the light verbs in terms of body region specificity. This result seems to confirm that a unique and strong association between the verb and a single body region is helpful for toddlers, and that frequency may not explain all of the acquisition data, because those verbs were very frequent too.
Table 1. A summary table of the eight comparisons in Analysis I. Each of the predominately white rows shows data from one of the analyses reported above. The first white row shows data from Analysis I-A, the second white row shows data from Analysis I-B, and so on. The last white row shows data from Analysis I-H.

<table>
<thead>
<tr>
<th>Mean body part percentages</th>
<th>SD</th>
<th>t-value</th>
<th>P-value for t-test</th>
<th>Mean AOA</th>
<th>Mean for t-test</th>
<th>Verbs</th>
<th>In brackets is the project number of occurrences for 1,000,000 words between 12 and 24 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 light verbs</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>make (888), put (2903), take (1334), have (1990), go (5547), give (606), get (2701), bring (150)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 most frequent and concrete verbs</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>close (1024), drink (811), eat (1521), look (2993), open (1013), read (1441), see (3123), sit (1896)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 verbs below 100/1,000,000 occurrences and no increase</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>clap, feed, lock, rip, shake, share, skate, chase, smile, spill, splash, sweep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 verbs below 100/1,000,000 and increasing</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>break, cook, finish, hate, hit, hear, hurry, pour, pretend, wish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 verbs with frequencies in the 100 range</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>blow, bring, build, buy, carry, climb, cover, dance, drive, dry, hug, kick, knock, listen, love, run, show, sing, stand, tickle, touch, wait, write</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 verbs with frequencies in the 200 range</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>bits, cry, drop, dump, hold, pick, stop, throw, wake, work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 verbs with frequencies in the 300 range</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>bump, catch, kiss, paint, ride, swing, walk, wash, talk, think, watch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 verbs with frequencies in the 400 to 500 ranges</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>cut, fit, pull, say, sleep, write, clean, jump</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 verbs with frequencies above the 600 range</td>
<td>0.12</td>
<td>6.5</td>
<td>2.41</td>
<td>2.6</td>
<td>draw (660), help (693), fix (750), push (737), fall (846), find (845), like (1317), play (1880)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* significant (2-tailed), p < .05 and ** significant (2-tailed), p < .01

**Discussion.** In Analysis I, the series of independent t-tests examining the frequencies in CHILDES (at 12 to 23 months) of the group of eight light verbs compared to the eight groups of verb frequencies in CHILDES indicates that seven groups of verbs out of eight were not significantly different in terms of average body region specificity. The only group of verbs that was significantly different from the light verbs was the one with a frequency below 100 occurrences per 1,000,000 words with a growth trend at later ages. This group had an average lower body region specificity, meaning these verbs were related to multiple body regions. In fact, these verbs happened to be acquired significantly later than the light verbs, which in turn were acquired significantly later than the highly frequent and concrete verbs. Following the results in Analysis I, we can make new differentiations: eight highly concrete verbs with high frequency are acquired earlier in CHILDES (downloaded June 2011), than the eight light verbs of equally high frequency usually proposed as quite abstract verbs (Clark, 1978; Pinker, 1989; Theakston et al., 2004, etc.). Interestingly, these two groups did not vary significantly in their average body region specificity. The next result indicates that light verbs like make, put, go, etc., with the highest frequencies (range between 800 and 5447 occurrences per 1,000,000 words) have a significant difference in body part associations from a
group of concrete verbs with very low frequencies (below 100 occurrences) that are not increasing much in frequency at later age; this later group of verbs have a marginal significance in age of acquisition with the light verbs, which supports the idea that frequency combined with body part associations can help differentiate between subtypes of concrete verbs. Finally, there is a tendency for the group of verbs with frequencies in the 600 and up per 1,000,000 to differ from the light verbs in terms of body region specificity (marginal significance). This result seems to confirm that a unique and strong association between the verb and a single body region is helpful for toddlers, and that frequency may not explain all of the development (Goodman, Dale & Ping, 2008), because those verbs were very frequent too. Finally, among the groups of verbs with the lowest frequencies, a group was not well-learned. It corresponds to a subsample of 10 verbs, the verbs with low frequencies that are later-learned, a subgroup of the “new comers” in CHILDES, and these verbs differ significantly from the light verbs in that they had, on average, lower body region specificity, that is, their meaning related to more than one main body region.

Analysis II: Body Region Specificity and Concreteness

Analysis II-A: Correlations with body region specificity

Question. To explore further how body region specificity relates to concreteness and abstraction in early verbs, we used imageability, a criterion previously found to help English children in verb acquisition (McDonough et al., 2011), and object type associations (Maouene et al., 2011), a criterion that allows characterization of the verbs as more concrete or more abstract along a continuum and aligns with the distinction between light and non-light verbs.

Method. We used Cortese & Fugget (2004)'s list of verbs with imageability ratings from adults and Maouene et al.'s list of verbs (2011) with number of object type associations from adults for which we had body region specificities from Maouene et al. (2008). We took all the verbs from the CDI list that were present in the three above mentioned studies, a total of 90 verbs. We then checked whether these measures correlated with the body region specificity of each verb and whether imageability and number of object types correlated with each other.

Results. Body region specificity did not correlate significantly with any of the variables tested: neither with imageability nor with number of object types, frequencies of production of the verbs by toddlers in CHILDES or age of acquisition from the Bates-MacArthur CDI across the 90 verbs. This result is expected since we argue that most concrete and abstract verbs have quite strong body region specificity. Interestingly, the two other measures of concreteness correlated significantly and negatively with each other: \( r(88) = -0.59, p < .05 \), such
that the more imageable a verb is rated, the fewer object types this verb is associated with. The correlations are reported below in Table 2.

Table 2. Correlations between imageability, number of object types, agreement on a body region, age of acquisition and frequencies in CHILDES

<table>
<thead>
<tr>
<th></th>
<th>image</th>
<th>object types</th>
<th>body regions</th>
<th>AoA</th>
<th>freqchildes</th>
</tr>
</thead>
<tbody>
<tr>
<td>image</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>object types</td>
<td>-0.586*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>body regions</td>
<td>0.09</td>
<td>-0.071</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AoA</td>
<td>-0.148</td>
<td>0.238*</td>
<td>-0.175</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>freqchildes</td>
<td>-0.432*</td>
<td>0.235*</td>
<td>0.117</td>
<td>-0.319**</td>
<td>1</td>
</tr>
</tbody>
</table>

Implications. The results indicate that there is no simple linear correlation between body region specificity on one hand and imageability and number of object type associates on the other hand.

Analysis II-B: Calculating difference scores

Question. We know from Maouene et al. (2008) that there is a wide range of body region specificities. Agreement among participants that a verb is associated with a main body region ranges from 26% to 100%. We wanted to be able to compare the distribution of body region specificity to the distributions of imageability and number of object types. We also wanted to compare particular segments of the distribution to come up with a more precise description of different groups of concrete and abstract verbs. The overall goal was to specify the landscape in terms of concreteness and abstraction. To that effect, we rescaled the data on body region specificity together with the data on number of object types so that each would fit on a scale from 1 to 7 (precisely the scale used for imageability judgments by Cortese & Fuggett, 2004).

Method. We proceeded as follows: for body region specificity, we took 100% of agreement on a body region for a particular verb to be equivalent to 7, the maximum score on the scale, so that 1 corresponds to no agreement on a single body region for a particular verb and 7 means everybody agreed on a single body region for that verb. For number of object types, we took the highest number of object types associated with a verb, 141 (that is 141 object types were associated with the verb take), and made that the maximum of the scale, or 7.
Consider, for example, the verb *bite*. It had 98% of participants agreeing (on the mouth region), 66 types of objects associated, and an imageability rating of 5.2 on a seven-point scale. After the rescaling, it has a body region specificity index of 6.86 and a diversity of object types index of 3.28, on the same seven-point scale. All the new values obtained from the rescaling are reported in the Appendix. We then summed the three indices of concreteness and averaged them into a unique composite index of concreteness ranging from 1 to 7 for the 90 verbs.

We reasoned that comparing the body region specificity index (B) of each verb to its composite concreteness index (C) by subtracting the composite concreteness index from the body region specificity index to obtain a comparative difference score (D=B−C) would allow us to classify the verbs. In that case, a negative difference score would mean the verb has low body region specificity relative to its concreteness, while a positive difference score would indicate that the verb has high body region specificity relative to its concreteness. For example, the verb *hate* has a body region specificity index of 2.33 (which corresponds to 40% agreement on mind, rescaled on the seven-point scale) and a composite concreteness index of 3.93 (the average between imageability, number of object types and body region specificity), so the difference score is −1.6, which places *hate* among the most abstract verbs in terms of body region specificity.

**Results.** We found that some verbs had body region specificity indices that were lower than their composite concreteness indices (negative difference scores), that some verbs had body region specificity indices at the same level as their composite concreteness indices (negligible difference scores), and that most verbs had body region specificity indices that were greater than their composite concreteness indices (positive difference scores). We then graphed the curve, with the difference score on the y-axis and the verbs ordered by this value on the x-axis. The graph obtained by this method is shown in Figure 1.
Figure 1.
The distribution of the 90 early-learned verbs (all 90 are plotted, but only a subsample of them are labeled here) ordered by the difference, for each verb, between its body region specificity index and its composite concreteness index, composed of imageability, object types and body region specificity.

The results indicate that most verbs, 80 verbs out of 90, have positive difference scores. That is, each of these 80 verbs has a body region specificity index greater than its composite concreteness index.

Implications. This result confirms our hypothesis that most verbs in this sample have strong associations with a main specific body region. The results further show that only 10 verbs had negative difference scores, that is, had low body region specificity indices relative to their composite concreteness indices. Among those, we have abstract verbs such as hate, play, think, hide, show and love—where the body region is hidden (heart, mind, brain) and where multiple body regions come into play, particularly the eye-brain, eye-heart, and eye-hand regions. We also find concrete verbs such as fit, bump, break and wake, verbs that can be used reflexively with many body parts (you can break your leg, your toes, your arm, etc.) or with the whole body; however, as can be seen, none of the light verbs belong to this list of ten abstract verbs. These results confirm the results of our first analysis—that the light verbs have high body region specificity even though they are considered abstract. These results further suggest that different groups of abstract verbs co-exist: those with higher body region specificity and those with lower body region specificity, along a continuum, and likewise for concrete verbs.
Analysis II-C: Grouping by difference scores

**Question.** Having determined through Analysis II-B that it might be possible to use difference scores to subdivide both abstract and concrete verbs into groups with high body region specificity and low body region specificity, we wanted to be able to compare the resulting groups.

**Method.** We divided the verbs into four categories based on the difference scores: (1) all the verbs with negative difference scores, a group of 10 verbs; (2) the verbs whose difference scores fall between 0.01 and 1.0, a group of 15 verbs; (3) the verbs whose difference scores fall between 1.01 and 2.0, a group of 41 verbs; and (4) the verbs with difference scores greater than 2.0, a group of 24 verbs.

**Results.** The independent samples Kruskal-Wallis test on the difference scores indicates that the mean rank (MR) for the negative range is $MR=7.25$, for the range from 0.1 to 1.0 is $MR=17.73$, for the range from 1.01 to 2 is $MR=49.91$ and for the range above 2.1 is $MR=71.25$, and that the distribution of the difference score across the four categories of ranges varies significantly, such that $H=7.89$, 3 df, $p<.05$.

**Implications.** We can order the verbs along an axis of concreteness where the ‘negative pole’ includes verbs with low body region specificity, low imageability and high object type diversity, and where the ‘positive pole’ includes verbs with high body region specificity, high imageability and low object type diversity. This last result suggests that, if we do so, then two groups of abstract verbs and two groups of concrete verbs can be discerned, and they are all different from each other in terms of their difference scores. This structure can only be unveiled when using groups with different numbers of verbs because the vast majority of the verbs lie in the middle of the continuum.

Analysis II-D: Age of acquisition and difference scores

**Question.** In Analysis II-D – Analysis II-G, we take a closer look at subgroups of verbs along this continuum of difference scores and examine how they relate to age of acquisition, frequency, imageability and diversity of object types. We begin in Analysis II-D with age of acquisition.

**Method.** We took the number of new verbs produced at each of the twelve ages from 19 months to 30 months, by 50% of the children, according to parental report (CDI, Fenson et al., 1940), and we computed the mean difference score at each age. As can be seen in Table 3, the early verbs (ages 19 and 20 months) look different from the middle verbs (22 to 28 months) and from the late verbs (29-30 months).

**Results.** The independent samples Mann-Whitney test indicates that the mean rank, $MR=11.22$, for the 9 verbs learned between 19 and 20 months, differs marginally from the mean rank, $MR=6.05$, for the 8 verbs learned at 21 months, with $SE=10.39,$
such that $U=16$, $z=-1.92$, $p=.059$, two-sided. The trend is significant when we compare the 9 earliest verbs with a mean rank, $MR=14.67$, and the 12 latest verb distributions at 29 and 30 months, where the mean rank is $MR=8.25$, with $SE=14.07$, and $U=21$, $z=-2.34$, $p=.018$, two-sided. 

**Implications.** Taken together, these results shed light on why we did not find a correlation between age of acquisition and body region specificity: the vast majority of the verbs have similar distributions except at the very beginning and at the very end of the age of acquisition curve. In sum, these results suggest that there is a trend for body region specificity to weigh more relative to the composite concreteness index for very early acquired verbs (between 19 and 21 months) compared to the verbs acquired at 22 months and a significant difference between those very early learned verbs and the latest acquired verbs. All the light verbs are situated between the 2nd rank (go at 19 months) and the 55th rank (have at 26 months) on the difference score curve, so that none of them are late-acquired verbs. Further, they have a mean difference score of $M=1.87$, $SD=0.45$, which places them close to the highest weight in body region specificity. This also replicates the findings of Analysis I, where this particular group of abstract verbs was found to have high body region specificity. Table 3 presents the mean difference scores ordered by monthly age of acquisition from 19 to 30 months and the standard deviations.

Table 3. 
*Average difference scores (mean distance) ordered by monthly age of acquisition from 19 to 30 months (the verbs and their normative age of acquisition are from Fenson et al., 1994)*

<table>
<thead>
<tr>
<th>Age categories</th>
<th>Nineteen</th>
<th>Twenty</th>
<th>Twenty-one</th>
<th>Twenty-two</th>
<th>Twenty-three</th>
<th>Twenty-four</th>
<th>Twenty-five</th>
<th>Twenty-six</th>
<th>Twenty-seven</th>
<th>Twenty-eight</th>
<th>Twenty-nine</th>
<th>Thirty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr of verbs</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>8</td>
<td>19</td>
<td>3</td>
<td>12</td>
<td>15</td>
<td>12</td>
<td>0</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Mean Distance</td>
<td>2.34</td>
<td>0</td>
<td>1.91</td>
<td>1.11</td>
<td>1.31</td>
<td>1.77</td>
<td>1.16</td>
<td>1.45</td>
<td>1.58</td>
<td>0</td>
<td>0.95</td>
<td>0.62</td>
</tr>
<tr>
<td>SD</td>
<td>0.3</td>
<td>0.37</td>
<td>0.94</td>
<td>0.96</td>
<td>1.15</td>
<td>1.13</td>
<td>0.73</td>
<td>0.84</td>
<td>0</td>
<td>1.36</td>
<td>1.25</td>
<td></td>
</tr>
</tbody>
</table>

**Analysis II-E: Frequency and difference scores**

**Question.** Having examined age of acquisition in Analysis II-D, we turn to frequency in Analysis II-E. We want to know whether there is a systematic relation between difference scores and frequency.

**Method.** We took the number of verbs occurring at the same frequencies for seven different frequencies, so as to have groups with roughly the same number of verbs per group and similar groups to the ones under Analysis I. Some groups were formed in increments of 100 occurrences per 1,000,000 utterances (0 to 99, 100 to 199, 200 to 299, and 300 to 399). The other three groups comprised verbs with 400 to 599, 600 to 999 and 1000 or more occurrences.

**Results.** As can be seen in Table 4, the most frequent verbs have a tendency to have larger difference scores than the least frequent, but this is only a trend. We would...
need more new verbs lately acquired to see if this trend would persist, get reinforced or disappear. A Mann-Whitney test comparing the difference scores of verbs with frequencies greater than 600 (= the highest frequencies, 22 verbs) to those of verbs with frequencies below 100 (= the lowest frequencies, 19 verbs) was not significant.

**Implications.** These results seem to confirm the results of Analysis I whereby frequency relative to the other dimensions of concreteness does not by itself discriminate abstract from concrete verbs, while they help with learning (see Table 2). Table 4 below shows the different means and standard deviations for each group of frequencies in terms of the difference score between body region specificity and composite concreteness.

Table 4.
*Average difference scores (mean distance) ordered by groups of productive frequencies in childes (toddlers from 12-to 23-months)*

<table>
<thead>
<tr>
<th>occurrences</th>
<th>&gt;1000</th>
<th>600-900</th>
<th>400-599</th>
<th>300-399</th>
<th>200-299</th>
<th>100-199</th>
<th>&lt;100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr of verbs</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>11</td>
<td>10</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>Mean Distance</td>
<td>1.54</td>
<td>1.64</td>
<td>1.37</td>
<td>1.19</td>
<td>1.38</td>
<td>1.34</td>
<td>1.2</td>
</tr>
<tr>
<td>SD</td>
<td>1.06</td>
<td>0.68</td>
<td>1.24</td>
<td>1.14</td>
<td>0.73</td>
<td>0.79</td>
<td>1.15</td>
</tr>
</tbody>
</table>

**Analysis II-F: Imageability and difference scores**

**Question.** To further our description of the difference between subgroups of verbs in terms of other correlated dimensions of concreteness, we want to know whether there is a systematic relation between difference scores and imageability.

**Method.** We checked the difference score when we order the verbs from most imageable (7) to least imageable (1). We took the verbs occurring at each interval of 1 on the scale from 1 to 7 and averaged their difference scores.

**Results.** As can be seen in Table 5, a U-shaped curve seems to emerge from that distribution, where the sections of the distribution ordered by the most imageable verbs and the least imageable verbs do not differ. Among the least imageable verbs, according to adult judgments (Cortese and Fugget, 2004), are all the light verbs. These fall on the interval from 1 to 2.9, confirming that these abstract verbs have as strong body region specificity as the most imageable and thus concrete verbs. They have a mean difference score of $M=1.87$, $SD=0.45$ whereas the other, least imageable and abstract verbs, such as *help, say, like, wait, think* and *find*, in those same intervals, have a mean difference score of $M=1.02$, $SD=1.02$, confirming two types of abstract verbs. For the concrete, most imageable, verbs, we find that 60 of them populate the intervals going from 4 to 4.9, 5 to 5.9 and 6 to 7, and they do not
differ in terms of their difference scores. However, they differ significantly from the


group of 13 verbs that are on the interval from 3 to 3.9 in terms of their mean
difference score. These verbs might be seen as overall less imageable from an adult


perspective, or transitioning towards abstraction. They are work, look, build, dry,


pick, close, hide, buy, hear, share, fit, fix, see, hate and wish.

**Implications.** In brief, a main group of 60 imageable verbs and 30 less imageable

verbs populate these intervals of imageability. Two different groups of “abstract”

verbs emerge in terms of imageability: the light verbs with strong body region

specificity and the abstract non-light verbs with weaker body region specificity.

Table 5 presents the mean difference scores ordered by an interval of one increment

on the scale from 1 to 7 and the standard deviations.

Table 5.

*Average difference scores (mean distance) ordered by imageability intervals*

<table>
<thead>
<tr>
<th>Imageability ranges</th>
<th>6 to 7</th>
<th>5 to 5.9</th>
<th>4 to 4.9</th>
<th>3 to 3.9</th>
<th>2 to 2.9</th>
<th>1 to 1.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr of verbs</td>
<td>3</td>
<td>24</td>
<td>33</td>
<td>16</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Mean Distance</td>
<td>1.5</td>
<td>1.5</td>
<td>1.34</td>
<td>0.99</td>
<td>1.54</td>
<td>1.64</td>
</tr>
<tr>
<td>SD</td>
<td>0.44</td>
<td>0.66</td>
<td>1.02</td>
<td>1.3</td>
<td>0.96</td>
<td>0.28</td>
</tr>
</tbody>
</table>

**Analysis II-G: Diversity of object types**

**Question.** To describe further the difference between subgroups of verbs in terms of

another correlated dimension of concreteness, number of object types (Maouene et al., 2011), we want to know whether there is a systematic relation between
difference scores and number of object types.

**Method.** We checked the difference scores when we order the verbs from many

object types (7), or abstract verbs, to few object types (1), or concrete verbs. We

took the verbs occurring at each interval of 1 on the scale from 1 to 7 and averaged

their difference scores.

**Results.** As can be seen in Table 6, a trend seems to emerge from that distribution

where the average difference score for the verbs with fewer object types differs

from the average difference score for the verbs with more object types. An

independent samples Mann-Whitney test indicates that the mean rank for the 12

verbs that take the most object types (abstract verbs on the intervals 6 to 7 and 4 to

4.9), $MR=14.33$, and the mean rank for the 24 verbs that take the fewest object

types (concrete verbs on the interval 1 to 1.9), $MR=20.58$, with $SE=29.79$, differ

marginally such that $U=94$, $z=-1.68$, $p=.091$, two-sided. The trend is significant

when we compare the 12 verbs that take the most object types (intervals 5 to 6 and

4 to 4.9), with a mean rank $MR=15.04$ with the 29 verbs with the fewest number of

verbs on the other side of the scale.
object types (interval 2 to 2.9), where the mean rank is $MR = 23.47$, with $SE = 34.89$, and $U = 102.5$, $z = -2.05$, $p = .039$, two-sided.

**Implications.** It is thus the case that a small group of 12 abstract verbs, those verbs that are associated with the highest number of object types by adults, have smaller difference scores than a group of 29 concrete verbs. In fact, the distribution tends to bi-modality: a group of 37 more abstract verbs associated with many different objects and a group of 53 more concrete verbs associated with fewer objects tend to differ in their mean difference scores. Here the light verbs fall on four intervals: from 6 to 7 (1 verb), 4 to 4.9 (4 verbs), 3 to 3.9 (2 verbs) and 2 to 2.9 (1 verb); this dimension seems to separate between them more. However, within the group of 12 abstract verbs, the light verbs have larger difference scores, $M = 1.77$, $SD = 0.45$, than the other verbs in this group, $M = 0.26$, $SD = 1.1$, such as break, help, work, show, pick, find and hate.

**Analysis II-H: Transitivity**

**Question.** Since this dimension of concreteness is about objects, we looked at the proportion of verbs that are mostly transitive and mostly intransitive in those four groups of verbs.

**Method.** We used data from the Merriam-Webster on-line dictionary at http://www.merriamwebster.com/dictionary/ because this dictionary makes a clear separation between the transitive and intransitive uses of each verb. Our criterion was whether the verb was first listed with its transitive or intransitive meanings. For example, under a verb like bite, the dictionary first listed all the transitive entries and second the intransitive ones, whereas for a verb like climb, the order was reversed.

**Results.** Interestingly, within the group of 37 verbs that ask for many objects, the group of 12 verbs (interval 6 to 7, no verb interval 5, and interval 4 to 4.9), the mean proportion of intransitive verbs is lower, $M = 0.08$, than for the next group of verbs, $M = 0.36$. This observation is also valid for the group of 53 object-narrow verbs, where the group of 24 verbs with the fewest object type associates have a higher proportion of intransitives, $M = 0.33$ (interval 1 to 1.9) than the other group of 29 object-wide verbs, $M = 0.10$.

**Implications.** In brief, we find interesting subdivisions in terms of percentages of intransitives within the two main groups of 53 object-restricted verbs and the two groups of 27 object-non-restricted verbs that populate these intervals of object types. Two different groups of “abstract” verbs emerge in terms of objects: the light verbs with higher difference scores, which also have a lower proportion of transitive verbs, and the abstract non-light verbs with lower difference scores and a higher proportion of intransitive verbs. Similarly, two different groups of “concrete verbs” can be described: they have similar difference scores, but the group of verbs with
the least number of object associates has a higher percentage of intransitive verbs (interval 1 to 1.9) than the second one (interval 2 to 2.9). Table 6 presents the mean difference scores ordered by the number of object type associates on an interval of one increment on the scale from 1 to 7 and the standard deviations.

Table 6.
Average difference scores ordered by object type intervals

<table>
<thead>
<tr>
<th>object types</th>
<th>6 to 7</th>
<th>5 to 5.9</th>
<th>4 to 4.9</th>
<th>3 to 3.9</th>
<th>2 to 2.9</th>
<th>1 to 1.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr of verbs</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>25</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Mean Distance</td>
<td>-0.1</td>
<td>0</td>
<td>1.09</td>
<td>1.05</td>
<td>1.64</td>
<td>1.55</td>
</tr>
<tr>
<td>SD</td>
<td>2.02</td>
<td>0</td>
<td>0.96</td>
<td>1.09</td>
<td>0.76</td>
<td>0.79</td>
</tr>
</tbody>
</table>

DISCUSSION

In sum, even in the analyses with the larger set of 90 verbs, the verbs picked out as abstract (including the light verbs) versus concrete did not differ reliably in terms of body region specificity. There were no reliable differences for any of the examined measures. Neither were there any reliable differences based on whether the verbs are concrete or abstract, by three different indicators of abstractness: children’s productive frequency between 12 to 23 months in CHILDES, adults’ imageability ratings, and number of types of associated objects. Thus, it appears that early light verbs are as strongly associated with specific body regions as early non-light verbs; that low-imageable verbs are as strongly associated with specific body regions as high-imageable verbs; and that verbs with many object type associations are as strongly associated with specific body regions as those with fewer object type associations.

However, a more differentiated picture emerges when subgroups of concrete and abstract verbs are analyzed separately. Frequency does not separate well between abstract and non-abstract verbs, but it helps with learnability (Analysis I and correlations in Analysis II). Small groups of very early-learned verbs and very late verbs have different distributions in terms of body region specificity (analyses I and II). Imageability and number of object types correlate negatively and significantly with each other, increasing our understanding of what it means for a verb to be concrete or abstract. The distribution of the average difference scores ordered by the imageability intervals forms more of a U shape, whereas the distribution of the average difference scores ordered by object type intervals tends to look more like a bi-modal distribution.

Within these distributions, the difference score varies for some subgroups of concrete and abstract verbs. For imageability, very imageable (concrete) and very
non imageable (abstract) verbs do not vary; only a subgroup of 19 verbs situated on the interval of 3 to 3.9 tends to do so. For number of object types, a small group of 12 verbs with the highest number of object type associates varies significantly from a larger group of concrete verbs that have a low number of object associates. Within the concrete verbs and the abstract verbs, we find also interesting differences, such as the difference between the light abstract verbs and the non-light abstract verbs (imageability) and the proportion of transitive verbs (number of object types) that differ both within concrete verbs and within abstract verbs.

**General Discussion**

In developmental cognitive science there has been increasing discussion as to whether the developing language may be better understood in its relation to the physicality of the body, in real-time, and in a physical world (Adolph, 2010; Iverson, 2010; Smith, 2010; see Samuleson, this issue). This idea is sometimes referred to as the embodiment hypothesis. Similarly, in the research with pre-linguistic infants, a shift in focus is perceptible where researchers study more and more how language emerges in human infants from its network of connections with the whole event rather then just from parental linguistic input (see Dueker, Portko & Zelinsky, this issue). Thus, from the embodiment perspective, although the language stimulus is arguably poor in the beginning (Pinker, 1984), the connections including—but not restricted to—the infant’s and the caregiver’s body and body parts, are extremely rich and are claimed to provide a decisive scaffolding for language acquisition (Bahrick, Flom & Lickliter, 2002; Baldwin, Markman, Bill, Desjardins & Irwin, 1996; Butterworth & Itakura, 2000; Campbell & Nam, 2003; Hoff, 2006; Kelly, 2001; Lewkowicz, 1998a,b; Pereira, Smith & Yu, 2008; Tomasello, 1999; Trautman & Rollins, 2006).

In our particular domain concerned with the early acquisition of verbs, the literature has mainly focused on lightness (a type of abstraction commonly operationalized by frequency) and concreteness (operationalized by imageability ratings and number of associated object types) as the principle drivers of acquisition, whereas the results presented here point to a more embodied perspective of word learning. As such, this study provides two main contributions: 1) The verbs that are more highly associated with a single region of the body, according to adult judgments, are among the first acquired in vocabulary development; 2) Associations with a specific region of the body may be a better predictor of whether a verb is learned early than the concreteness or abstractness of the verb.

The present finding that the verbs that are more highly associated with a single region of the body, according to adult judgments, are among the first acquired in vocabulary development is quite puzzling. Why would strength in
associations on one main body region—according to adult judgments—help acquisition?

One possible argument is that adults use their knowledge of verbs in their speech to children, so children hear regularities and associations that are structured by adult knowledge. This study cannot claim more than that, but our ongoing work seems to point in that direction: Examining in 36 corpora of CHILDES the occurrences of body parts terms immediately before and immediately after the same early-learned verbs from the MCDI, our results indicate that respectively 21% (before) and 26.8% (after) of the verbs co-occur at least once with a body part in the speech of 24 month-olds and these percentages are respectively 39% (before) and 26.8% (after) in the speech of parents to their 24-month-old children (Maouene, Laakso & Smith, in preparation). These preliminary results seem to indicate that using the adult metrics of body parts associations to verbs—because they share the same structure of a body that the children do—is a meaningful and stable enough set of constraints. Further, although nothing tells us in this data that it is learning about the child, we can argue that there is something powerful about the effectors in terms of their role of ubiquitous references. When asking children, “What body part do you use when you bite”, there is little chance that children will answer leg, ear or eye. Preliminary results indicate that indeed for 20 children interviewed, 64% answered mouth and 36% teeth, a result that is as strong as adult associations (Maouene & Maouene, in preparation).

Further, children could be helped because the strength in associations reflect the probability of a specific effector to occur in a particular action the verb refers to and thus emerge as a solid predictive cue. For example, when we compare the nine verbs that are learned the earliest according to the MCDI (bite, drink, eat, kiss, go, sit, watch, hug) with the ten verbs that have a negative difference to the composite index (hate, play, think, hide, show, love, fit, bump, break, wake), we find that that the actions that have a more distributed combination of effectors (fall: leg-whole body, break: leg-arm-back, think: mind-eye-head, pretend: eye-hand-mind, etc.) are learned later. We also have indications that “visualizability” of the effectors might matter: brain, mind, heart are hidden and thus could explain partly why the cognitive verbs and emotional verbs are learned later (like, love, hate, think). Additionally, among the verbs with a negative distance are actions that allow for a self-directed use of many body parts or regions like break, bump (you can bump and break a lot of different body parts), or where the whole body seems to be involved (fit, wake) and speaks to the lack of a strong predictive cue. However, a common objection could be that associations are a product of learning rather then driving learning. The present data cannot address this issue in and of itself, but experimental studies with fetuses and computer simulations might help shed light on this long-standing question.

Finally, one could object to the embodiment claim presented here, that although the present results tell us that mature speakers strongly agree in their
associations with one main body region for most concrete and abstract verbs, they
do not tell us how, or whether, that knowledge is used in meaning comprehension.
We certainly need more experimental work, but different imagery data in adults
(Hauk, Johnsrude, & Pulvermüller, 2004) and in children (James & Maouene, 2009)
indicate that reading verbs or hearing them activate regions of the associative motor
cortex selectively for leg verbs, hand verbs and mouth verbs. Further, with novel
verbs and novel actions, only manipulation (action) but not observation creates
connections in the brain between the auditory cortex and the sensori-motor cortex
(James & Swain, 2011). In brief, the first analyses presented here suggest that there
is enough ground to start considering body part correlations as a possible variable in
verb acquisition in children.

The second finding—that associations with a specific region of the body
may be a better predictor of whether a verb is learned early than the concreteness or
abstractness of the verb—is quite challenging too. The challenge of reframing our
thinking patterns to incorporate embodiment is quite daunting because, if we
believe Heidegger cited in Dourish (2001), we treat our body as a “transparent
equipment”, that is, our body is always here with us, but we only notice it when
something goes wrong. Interestingly, the idea at the origin of the phenomenological
revolution in thinking that Husserl brought about was concerned with how to bring
back the body to mathematics (Dourish, 2001)! All proportions kept, we propose to
bring the construct of embodiment and its operationalization by the correlation with
strength in associations to a main body region to a literature that has considered so
far abstractness and concreteness, operationalized respectively by frequency and
imageability, as the two main correlations driving word acquisition in general and
verb acquisition in particular.

When we consider the arguments on abstraction and concreteness in the
literature, we are left with some unanswered questions and we propose that the
results presented here can potentially answer those particular questions. Basically,
on one hand, the argument on the side of abstraction is that a set of highly frequent
and semantically general verbs, light verbs, (such as do, make, put, get, have, take,
go, give, bring, etc.) are thought to help children in acquiring verbs’ relational
meanings and promote the learning of argument structures (for a review, see
Theakston et al., 2004) because (1) they are early-learned (Clark, 1978); (2) they
are very frequent and thus highly salient (Clark, 1978); (3) they are associated with
the syntactic pattern in which they are most frequently used (Goldberg et al., 2004);
and (4) they consist of relational structures that reflect primitive and innate
semantic elements (Pinker, 1989). However, although these abstract verbs are
particularly early, particularly frequent and semantically general, some other verbs
are also particularly early and particularly frequent and are not abstract at all, such
as eat, kiss, drink, bite, sit, walk, watch, hug, which are the eight first verbs learned
in the MCDI list together with one abstract verb: go (Fenson et al., 1994). Our
results indicate that those verbs are learned significantly earlier than the eight light
verbs, a result that semantic generality cannot explain. So the need for another variable is felt. It is also felt, because, although frequency can explain the difference in acquisition between early-learned abstract verbs and late-learned abstract verbs and can explain why infrequent concrete verbs are acquired later, it cannot easily explain why among the very low frequency verbs some are acquired before others, a result of the present analysis (I-B and I-C).

Let’s consider the other argument, or concreteness and imageability. Researchers have proposed that highly imageable verbs are learned earlier than less imageable verbs (imageability hypothesis, Maguire et al., 2008). Supporting this idea, McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon (2011), correlating adult imageability ratings (taken from Masterson & Druks, 1998) with the MacArthur-Bates CDI age of acquisition English data (Fenson et al., 1994) reported a significant relationship between age of acquisition reported by parents and imageability for 44 English verbs, such that between 14 and 30 months, more imageable verbs were acquired earlier than less imageable verbs. They also found that imageability contributed to the prediction of age of acquisition beyond form class, for an additional 11% of the variance, suggesting that imageability might be a driving factor in predicting age of acquisition in addition to frequency in the input. These results, using imageability alone, can potentially explain why highly imageable verbs such as eat, kiss, drink, bite, sit, walk, watch, hug, are learned before highly frequent non imageable verbs (all the abstract verbs: light and non-light verbs). However they need frequency, to explain why some concrete verbs are learned before some other concrete with the same degree of imageability. Now, if we keep frequency and imageability constant, they cannot explain why some concrete verbs are learned earlier than some abstract verbs or some abstract verbs are learned earlier than some concrete verbs. Further, adult’s rating of imageability may not correspond to children’s ratings of imageability. Typically, look is considered a low-imageable verb in both Masterson and Druks’ and Cortese and Fugget’s reported lists of judgments, whereas we could certainly argue that it is much more concrete for a child as the tenants of the perspective on restricted word use have claimed (Akhtar & Tomasello, 1997; Hart & Risley, 1995; Tomasello, 2003). Recent results by Maouene et al., 2011, support his claim, and indicate that some verbs start more concrete than generally expected if we use verb-object co-occurrences in CHILDES (parent and children speech collapsed) as an operationalized criterion of the distinction between light/ heavy verbs. For example, push co-occurs with buttons 84% of the time in 36 corpora of CHILDES, and starts heavier whereas in adult judgments, it is more of a light verb. Finally, increasing the number of verbs examined, from 44 (McDonough et al., 2011) to 90 (Cortese and Fugget, 2004), did not allow us to replicate the results that overall high-imageable verbs are learned significantly earlier than low-imageable verbs. In that regard, we could argue then that the set of constraints underlying the meaning of verbs, imposed by the interaction between our action and our body morphology in parts
and captured by adult associations, constitutes a more stable set of constraints than the constraints underlying imageability or lightness and heaviness (number of object type associates) because adult experiences are more influenced by abstraction than young children’s.

In conclusion, we propose that these data provide a first step—and a pathway—to understanding how the body we have, and particularly its parts, may help children understand the relational meaning of early verbs, together with correlations with abstract linguistic experiences and other concrete world experiences. The current research helps us understand whether different kinds of verbs are learned in different ways, provides different lessons about early verb semantics, and sheds new light on the role of embodiment in concrete and abstract verb meaning development.

REFERENCES


Maouene, J. & Maouene, M. Body parts and early-learned verbs, children’s version (manuscript in preparation).

Maouene, J., Laakso, A., & Smith, L.B. A corpus analysis of body parts terms occurrences before and after a hundred early-learned verbs (manuscript in preparation).


Appendix
The eight different measures used to analyze a hundred concrete and abstract verbs and their relation to percentages of agreement on a main body region. *t means transitive and i means intransitive.

<table>
<thead>
<tr>
<th>Verbs</th>
<th>Body Regions</th>
<th>Frequencies Childes 12-23 mo.</th>
<th>AoA MCDI</th>
<th>image ratings</th>
<th>Object types rescaled</th>
<th>% body regions rescaled</th>
<th>Distance index in BR to</th>
<th>index</th>
</tr>
</thead>
<tbody>
<tr>
<td>bite t*</td>
<td>mouth</td>
<td>264</td>
<td>21</td>
<td>5.2</td>
<td>3.28</td>
<td>6.86</td>
<td>5.11</td>
<td>1.75</td>
</tr>
<tr>
<td>blow t</td>
<td>mouth</td>
<td>140</td>
<td>23</td>
<td>4.2</td>
<td>2.88</td>
<td>5.6</td>
<td>4.23</td>
<td>1.37</td>
</tr>
<tr>
<td>break t</td>
<td>hand</td>
<td>83</td>
<td>23</td>
<td>4.5</td>
<td>4.02</td>
<td>3.78</td>
<td>4.1</td>
<td>-0.32</td>
</tr>
<tr>
<td>bring t</td>
<td>hand</td>
<td>150</td>
<td>25</td>
<td>2.6</td>
<td>4.47</td>
<td>6.86</td>
<td>4.64</td>
<td>2.22</td>
</tr>
<tr>
<td>build t</td>
<td>hand</td>
<td>100</td>
<td>27</td>
<td>3.8</td>
<td>2.38</td>
<td>6.44</td>
<td>4.21</td>
<td>2.23</td>
</tr>
<tr>
<td>bump i</td>
<td>head</td>
<td>315</td>
<td>25</td>
<td>4.9</td>
<td>3.57</td>
<td>2.52</td>
<td>3.67</td>
<td>-1.14</td>
</tr>
<tr>
<td>buy t</td>
<td>hand</td>
<td>154</td>
<td>27</td>
<td>3.6</td>
<td>2.63</td>
<td>6.02</td>
<td>4.08</td>
<td>1.94</td>
</tr>
<tr>
<td>catch t</td>
<td>arm</td>
<td>331</td>
<td>24</td>
<td>4.4</td>
<td>1.49</td>
<td>6.86</td>
<td>4.25</td>
<td>2.61</td>
</tr>
<tr>
<td>chase t</td>
<td>leg</td>
<td>100</td>
<td>30</td>
<td>4.4</td>
<td>3.62</td>
<td>6.86</td>
<td>4.96</td>
<td>1.9</td>
</tr>
<tr>
<td>clap i</td>
<td>hand</td>
<td>53</td>
<td>23</td>
<td>5.6</td>
<td>1.54</td>
<td>7</td>
<td>4.71</td>
<td>2.29</td>
</tr>
<tr>
<td>clean t</td>
<td>arm</td>
<td>522</td>
<td>23</td>
<td>4.5</td>
<td>3.48</td>
<td>6.16</td>
<td>4.71</td>
<td>1.45</td>
</tr>
<tr>
<td>climb i</td>
<td>arm</td>
<td>177</td>
<td>25</td>
<td>5.4</td>
<td>0.94</td>
<td>3.78</td>
<td>3.37</td>
<td>0.41</td>
</tr>
<tr>
<td>close t</td>
<td>hand</td>
<td>1042</td>
<td>25</td>
<td>3.7</td>
<td>1.99</td>
<td>4.9</td>
<td>3.53</td>
<td>1.37</td>
</tr>
<tr>
<td>cook t</td>
<td>hand</td>
<td>87</td>
<td>23</td>
<td>5.4</td>
<td>2.63</td>
<td>6.44</td>
<td>4.82</td>
<td>1.62</td>
</tr>
<tr>
<td>cry i</td>
<td>eye</td>
<td>217</td>
<td>22</td>
<td>5.3</td>
<td>1.89</td>
<td>6.86</td>
<td>4.68</td>
<td>2.18</td>
</tr>
<tr>
<td>cut t</td>
<td>hand</td>
<td>419</td>
<td>26</td>
<td>5.1</td>
<td>2.33</td>
<td>6.02</td>
<td>4.49</td>
<td>1.54</td>
</tr>
<tr>
<td>dance i</td>
<td>leg</td>
<td>174</td>
<td>22</td>
<td>5.6</td>
<td>3.82</td>
<td>6.44</td>
<td>5.29</td>
<td>1.15</td>
</tr>
<tr>
<td>draw t</td>
<td>hand</td>
<td>660</td>
<td>25</td>
<td>5</td>
<td>2.13</td>
<td>6.86</td>
<td>4.67</td>
<td>2.19</td>
</tr>
<tr>
<td>drink t</td>
<td>mouth</td>
<td>811</td>
<td>21</td>
<td>5.7</td>
<td>1.59</td>
<td>5.74</td>
<td>4.34</td>
<td>1.4</td>
</tr>
<tr>
<td>drive t</td>
<td>hand</td>
<td>177</td>
<td>23</td>
<td>5.3</td>
<td>1.09</td>
<td>4.62</td>
<td>3.67</td>
<td>0.95</td>
</tr>
<tr>
<td>drop t</td>
<td>hand</td>
<td>264</td>
<td>26</td>
<td>4.3</td>
<td>3.77</td>
<td>6.16</td>
<td>4.75</td>
<td>1.42</td>
</tr>
<tr>
<td>dry t</td>
<td>hand</td>
<td>137</td>
<td>27</td>
<td>3.8</td>
<td>3.23</td>
<td>4.62</td>
<td>3.88</td>
<td>0.74</td>
</tr>
<tr>
<td>dump t</td>
<td>hand</td>
<td>207</td>
<td>30</td>
<td>4.7</td>
<td>2.18</td>
<td>4.62</td>
<td>3.84</td>
<td>0.79</td>
</tr>
<tr>
<td>eat t</td>
<td>mouth</td>
<td>1521</td>
<td>19</td>
<td>4.8</td>
<td>2.04</td>
<td>6.44</td>
<td>4.43</td>
<td>2.02</td>
</tr>
<tr>
<td>fall i</td>
<td>leg</td>
<td>848</td>
<td>22</td>
<td>4.8</td>
<td>1.99</td>
<td>3.5</td>
<td>3.43</td>
<td>0.07</td>
</tr>
<tr>
<td>feed t</td>
<td>hand</td>
<td>77</td>
<td>26</td>
<td>4.4</td>
<td>2.53</td>
<td>4.2</td>
<td>3.71</td>
<td>0.49</td>
</tr>
<tr>
<td>find t</td>
<td>eye</td>
<td>848</td>
<td>25</td>
<td>2.5</td>
<td>4.82</td>
<td>6.16</td>
<td>4.49</td>
<td>1.67</td>
</tr>
<tr>
<td>Action</td>
<td>Part of Body</td>
<td>N</td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
<td>T4</td>
<td>T5</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>fit</td>
<td>whole body</td>
<td>492</td>
<td>29</td>
<td>3.1</td>
<td>3.43</td>
<td>1.82</td>
<td>2.78</td>
<td></td>
</tr>
<tr>
<td>fix t</td>
<td>hand</td>
<td>750</td>
<td>29</td>
<td>3.2</td>
<td>3.48</td>
<td>6.16</td>
<td>4.28</td>
<td></td>
</tr>
<tr>
<td>get t</td>
<td>hand</td>
<td>2701</td>
<td>23</td>
<td>1.7</td>
<td>3.72</td>
<td>5.18</td>
<td>3.54</td>
<td></td>
</tr>
<tr>
<td>give t</td>
<td>hand</td>
<td>606</td>
<td>22</td>
<td>2.8</td>
<td>2.63</td>
<td>6.3</td>
<td>3.91</td>
<td></td>
</tr>
<tr>
<td>go i</td>
<td>leg</td>
<td>5457</td>
<td>19</td>
<td>2.2</td>
<td>3.87</td>
<td>6.58</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td>hate t</td>
<td>mind</td>
<td>0</td>
<td>30</td>
<td>3</td>
<td>6.4</td>
<td>2.38</td>
<td>3.93</td>
<td></td>
</tr>
<tr>
<td>have t</td>
<td>hand</td>
<td>1890</td>
<td>26</td>
<td>1.9</td>
<td>4.72</td>
<td>5.18</td>
<td>3.93</td>
<td></td>
</tr>
<tr>
<td>hear t</td>
<td>ear</td>
<td>13</td>
<td>26</td>
<td>3.3</td>
<td>2.33</td>
<td>7</td>
<td>4.21</td>
<td></td>
</tr>
<tr>
<td>help t</td>
<td>hand</td>
<td>693</td>
<td>23</td>
<td>2.8</td>
<td>4.22</td>
<td>5.6</td>
<td>4.21</td>
<td></td>
</tr>
<tr>
<td>hide t</td>
<td>eye</td>
<td>90</td>
<td>25</td>
<td>3.7</td>
<td>3.82</td>
<td>2.8</td>
<td>3.44</td>
<td></td>
</tr>
<tr>
<td>hit t</td>
<td>hand</td>
<td>67</td>
<td>23</td>
<td>4.4</td>
<td>2.18</td>
<td>5.74</td>
<td>4.11</td>
<td></td>
</tr>
<tr>
<td>hold t</td>
<td>arm</td>
<td>298</td>
<td>23</td>
<td>4</td>
<td>3.67</td>
<td>6.86</td>
<td>4.85</td>
<td></td>
</tr>
<tr>
<td>hug t</td>
<td>arm</td>
<td>167</td>
<td>21</td>
<td>5.9</td>
<td>2.18</td>
<td>6.72</td>
<td>4.94</td>
<td></td>
</tr>
<tr>
<td>jump i</td>
<td>leg</td>
<td>573</td>
<td>26</td>
<td>4.8</td>
<td>3.38</td>
<td>7</td>
<td>5.06</td>
<td></td>
</tr>
<tr>
<td>kick t</td>
<td>leg</td>
<td>197</td>
<td>23</td>
<td>5.1</td>
<td>1.19</td>
<td>6.72</td>
<td>4.34</td>
<td></td>
</tr>
<tr>
<td>kiss t</td>
<td>mouth</td>
<td>358</td>
<td>23</td>
<td>6.1</td>
<td>1.94</td>
<td>6.86</td>
<td>4.97</td>
<td></td>
</tr>
<tr>
<td>knock t</td>
<td>hand</td>
<td>124</td>
<td>21</td>
<td>4.9</td>
<td>0.99</td>
<td>5.46</td>
<td>3.79</td>
<td></td>
</tr>
<tr>
<td>lick t</td>
<td>mouth</td>
<td>30</td>
<td>25</td>
<td>5.3</td>
<td>2.63</td>
<td>6.72</td>
<td>4.88</td>
<td></td>
</tr>
<tr>
<td>like t</td>
<td>head</td>
<td>1317</td>
<td>26</td>
<td>2.7</td>
<td>3.87</td>
<td>3.36</td>
<td>3.31</td>
<td></td>
</tr>
<tr>
<td>look i</td>
<td>eye</td>
<td>2933</td>
<td>27</td>
<td>3.9</td>
<td>2.98</td>
<td>6.86</td>
<td>4.58</td>
<td></td>
</tr>
<tr>
<td>love i</td>
<td>heart</td>
<td>144</td>
<td>23</td>
<td>4</td>
<td>3.92</td>
<td>3.78</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>make t</td>
<td>hand</td>
<td>838</td>
<td>23</td>
<td>2.8</td>
<td>4.32</td>
<td>6.58</td>
<td>4.57</td>
<td></td>
</tr>
<tr>
<td>paint t</td>
<td>hand</td>
<td>311</td>
<td>26</td>
<td>6.3</td>
<td>2.38</td>
<td>5.88</td>
<td>4.86</td>
<td></td>
</tr>
<tr>
<td>pick t</td>
<td>hand</td>
<td>298</td>
<td>29</td>
<td>3.8</td>
<td>4.72</td>
<td>5.18</td>
<td>4.57</td>
<td></td>
</tr>
<tr>
<td>play i</td>
<td>hand</td>
<td>1880</td>
<td>23</td>
<td>4.5</td>
<td>3.18</td>
<td>2.38</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td>pull t</td>
<td>hand</td>
<td>455</td>
<td>26</td>
<td>4.1</td>
<td>2.68</td>
<td>6.3</td>
<td>4.36</td>
<td></td>
</tr>
<tr>
<td>push t</td>
<td>hand</td>
<td>737</td>
<td>24</td>
<td>4.1</td>
<td>3.92</td>
<td>6.16</td>
<td>4.73</td>
<td></td>
</tr>
<tr>
<td>put t</td>
<td>hand</td>
<td>2903</td>
<td>25</td>
<td>2.2</td>
<td>4.57</td>
<td>6.44</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>ride i</td>
<td>leg</td>
<td>301</td>
<td>22</td>
<td>4.6</td>
<td>1.29</td>
<td>4.48</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>rip t</td>
<td>hand</td>
<td>0</td>
<td>30</td>
<td>4.7</td>
<td>2.83</td>
<td>6.02</td>
<td>4.52</td>
<td></td>
</tr>
<tr>
<td>run i</td>
<td>leg</td>
<td>174</td>
<td>23</td>
<td>4.6</td>
<td>3.38</td>
<td>6.58</td>
<td>4.85</td>
<td></td>
</tr>
<tr>
<td>say t</td>
<td>mouth</td>
<td>449</td>
<td>27</td>
<td>2.8</td>
<td>2.38</td>
<td>6.72</td>
<td>3.97</td>
<td></td>
</tr>
<tr>
<td>see t</td>
<td>eye</td>
<td>3121</td>
<td>21</td>
<td>3</td>
<td>3.57</td>
<td>7</td>
<td>4.53</td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>Location</td>
<td>Frequency</td>
<td>Freq.</td>
<td>Semantic</td>
<td>Freq.</td>
<td>Semantic</td>
<td>Freq.</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------</td>
<td>----------</td>
<td>-------</td>
<td>----------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>shake</td>
<td>hand</td>
<td>67</td>
<td>29</td>
<td>4.4</td>
<td>3.23</td>
<td>4.76</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td>share</td>
<td>hand</td>
<td>13</td>
<td>27</td>
<td>3.2</td>
<td>3.97</td>
<td>5.6</td>
<td>4.26</td>
<td></td>
</tr>
<tr>
<td>show</td>
<td>hand</td>
<td>127</td>
<td>27</td>
<td>4</td>
<td>4.72</td>
<td>3.92</td>
<td>4.21</td>
<td></td>
</tr>
<tr>
<td>sing</td>
<td>mouth</td>
<td>164</td>
<td>25</td>
<td>4.9</td>
<td>1.34</td>
<td>6.02</td>
<td>4.09</td>
<td></td>
</tr>
<tr>
<td>sit</td>
<td>leg</td>
<td>1884</td>
<td>19</td>
<td>4.7</td>
<td>1.14</td>
<td>6.86</td>
<td>4.23</td>
<td></td>
</tr>
<tr>
<td>skate</td>
<td>leg</td>
<td>0</td>
<td>30</td>
<td>5.9</td>
<td>2.13</td>
<td>6.72</td>
<td>4.92</td>
<td></td>
</tr>
<tr>
<td>sleep</td>
<td>eye</td>
<td>439</td>
<td>22</td>
<td>5.5</td>
<td>1.19</td>
<td>3.36</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td>smile</td>
<td>mouth</td>
<td>164</td>
<td>27</td>
<td>6.5</td>
<td>1.94</td>
<td>6.58</td>
<td>5.01</td>
<td></td>
</tr>
<tr>
<td>spill</td>
<td>hand</td>
<td>53</td>
<td>26</td>
<td>4.2</td>
<td>2.04</td>
<td>6.16</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td>splash</td>
<td>hand</td>
<td>73</td>
<td>26</td>
<td>5.1</td>
<td>0.94</td>
<td>5.6</td>
<td>3.88</td>
<td></td>
</tr>
<tr>
<td>stand</td>
<td>leg</td>
<td>137</td>
<td>26</td>
<td>5</td>
<td>3.33</td>
<td>7</td>
<td>5.11</td>
<td></td>
</tr>
<tr>
<td>stop</td>
<td>leg</td>
<td>258</td>
<td>24</td>
<td>4.5</td>
<td>0.94</td>
<td>4.62</td>
<td>3.36</td>
<td></td>
</tr>
<tr>
<td>sweep</td>
<td>hand</td>
<td>67</td>
<td>27</td>
<td>5</td>
<td>1.34</td>
<td>6.02</td>
<td>4.12</td>
<td></td>
</tr>
<tr>
<td>swim</td>
<td>hand</td>
<td>73</td>
<td>23</td>
<td>5.6</td>
<td>1.69</td>
<td>5.32</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>swing</td>
<td>leg</td>
<td>395</td>
<td>22</td>
<td>5.8</td>
<td>1.29</td>
<td>3.92</td>
<td>3.67</td>
<td></td>
</tr>
<tr>
<td>take</td>
<td>hand</td>
<td>1334</td>
<td>27</td>
<td>2.8</td>
<td>7</td>
<td>6.86</td>
<td>5.55</td>
<td></td>
</tr>
<tr>
<td>talk</td>
<td>mouth</td>
<td>305</td>
<td>26</td>
<td>4.7</td>
<td>2.23</td>
<td>6.72</td>
<td>4.55</td>
<td></td>
</tr>
<tr>
<td>taste</td>
<td>mouth</td>
<td>90</td>
<td>29</td>
<td>4</td>
<td>2.28</td>
<td>7</td>
<td>4.43</td>
<td></td>
</tr>
<tr>
<td>think</td>
<td>mind</td>
<td>301</td>
<td>30</td>
<td>2.6</td>
<td>2.53</td>
<td>2.8</td>
<td>2.64</td>
<td></td>
</tr>
<tr>
<td>throw</td>
<td>hand</td>
<td>268</td>
<td>23</td>
<td>4.3</td>
<td>1.59</td>
<td>6.72</td>
<td>4.20</td>
<td></td>
</tr>
<tr>
<td>touch</td>
<td>hand</td>
<td>150</td>
<td>26</td>
<td>4.2</td>
<td>3.67</td>
<td>6.72</td>
<td>4.86</td>
<td></td>
</tr>
<tr>
<td>wait</td>
<td>leg</td>
<td>110</td>
<td>26</td>
<td>2.6</td>
<td>3.08</td>
<td>3.5</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>wake</td>
<td>eye</td>
<td>201</td>
<td>27</td>
<td>3.4</td>
<td>2.23</td>
<td>4.2</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>walk</td>
<td>leg</td>
<td>368</td>
<td>21</td>
<td>5.2</td>
<td>2.18</td>
<td>6.86</td>
<td>4.75</td>
<td></td>
</tr>
<tr>
<td>wash</td>
<td>hand</td>
<td>358</td>
<td>22</td>
<td>5.1</td>
<td>1.89</td>
<td>6.16</td>
<td>4.38</td>
<td></td>
</tr>
<tr>
<td>watch</td>
<td>eye</td>
<td>311</td>
<td>25</td>
<td>5.9</td>
<td>2.04</td>
<td>6.3</td>
<td>4.75</td>
<td></td>
</tr>
<tr>
<td>wipe</td>
<td>hand</td>
<td>107</td>
<td>25</td>
<td>5</td>
<td>2.88</td>
<td>4.62</td>
<td>4.17</td>
<td></td>
</tr>
<tr>
<td>wish</td>
<td>mind</td>
<td>26</td>
<td>30</td>
<td>3</td>
<td>2.63</td>
<td>2.94</td>
<td>2.86</td>
<td></td>
</tr>
<tr>
<td>work</td>
<td>hand</td>
<td>211</td>
<td>23</td>
<td>3.9</td>
<td>4.52</td>
<td>4.62</td>
<td>4.35</td>
<td></td>
</tr>
<tr>
<td>write</td>
<td>hand</td>
<td>479</td>
<td>27</td>
<td>4.8</td>
<td>1.44</td>
<td>6.58</td>
<td>4.27</td>
<td></td>
</tr>
</tbody>
</table>
SELF-GENERATED ACTIONS DURING LEARNING OBJECTS AND SOUNDS CREATE SENSORI-MOTOR SYSTEMS IN THE DEVELOPING BRAIN.

Karin Harman JAMES1,2,3*, Paroma BOSE1

1Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA
2Program in Cognitive Science, Indiana University, Bloomington, IN, USA
3Program in Neuroscience, Indiana University, Bloomington, IN, USA

ABSTRACT

Previous research shows that sensory and motor systems interact during verb perception, and that these interactions are formed through self-generated actions that refer to verb labels during development. Here we expand on these findings by investigating whether self-generated actions lead to sensori-motor interaction during sound perception and visual perception. The current research exposes young children to novel sounds that are produced by object movement through either a) actively exploring the objects and producing the sounds or b) by seeing and hearing an experimenter interact with the objects. Results demonstrate that the motor system was recruited during auditory perception only after learning involved self-generated interactions with objects. Interestingly, visual association regions were also active during both sound perception and visual perception after active exploratory learning, but not after passive observation. Therefore, in the developing brain, associations are built upon real-world interactions of body and environment, leading to sensori-motor representations of both objects and sounds.

KEYWORDS: fMRI, embodiment, brain development

Because individuals are situated in an environment that they actively explore, vision and motor experience are associated during learning about objects. Recent research indicates that motor systems are active during visual object perception if the object is associated with a history of action (e.g., Chao & Martin, 2000; Grezes & Decety, 2002; Longcamp, Anton, Roth & Velay, 2005; James & Gauthier, 2006; Weisberg, 2008).
van Turennout & Martin, 2007). Thus, embodied experiences result in the co-activation of sensory and motor systems in the brain. Although controversial, mounting research suggests that observation of action in humans is very different from actual self-generated action in terms of brain responses in children (James & Swain, 2011) and in adults (Butler & James, 2011, Turella, Pierno, Tubaldi, & Castiello, 2009; Dinstein, Hasson, Rubin, & Heeger, 2007; see also Dinstein, Gardner, Jazayeri, & Heeger, 2008), although others show an overlap of responses during observed and executed actions (Iacoboni, 2005; Rizzolatti & Craighero, 2004). Motor systems are also active in the processing of some verbs, another kind of stimulus that is associated with action (e.g. Pulvermüller, Harle, & Hummel, 2001; Hauk, Johnsrude & Pulvermüller, 2004; James & Maouene, 2010). Other evidence indicates that it is not just verbs that activate the motor system, but any word that has a history of association with an action – including nouns (Saccuman et al., 2006; Arevalo et al., 2007). These findings are changing contemporary understanding of the multi-modal and sensori-motor nature of the processes that underlie perception and cognition (Barsalou, et al., 2003).

In the developing system, it stands to reason that action is crucial for learning. This is not a new idea, being proposed by such leaders in our field as Piaget (1977) and Gibson (1969), and more recently has been elaborated upon by several developmental psychologists (Thelan & Smith, 1994; Bertenthal, & Campos, 1987; Bushnell & Boudreau, 1993; Needham et al., 2002). However, the mechanisms that support this important interaction are not well understood. By studying neural responses during learning with and without action, we can begin to understand why action is important to learning in many domains. Thus far, we have documented that action changes brain processing in children during cognitive operations such as verb processing (James & Maouene, 2009) object perception (James & Swain, 2011), and letter perception (James, 2010). However, only one study to date has directly compared active learning directly with passive observation of action in young children using fMRI (James & Swain, 2011). In that study, children were trained on learning novel verbs that were associated with actions on objects. The actions that were associated with the verbs were either learned actively, through the child producing the action and naming it, or passively, by observing an experimenter produce the action and name it. In both situations, the children learned the verbs and the actions, but when functional imaging was performed, only the active condition resulted in sensori-motor networks responding to the verb names. Passively learning the verbs while watching others did not recruit a motor network, whereas active learning did (James & Swain, 2011). From an embodied framework, it makes sense that self-generated actions that produced certain percepts - be they sound or image percepts – would result in sensory-motor associations in the brain. When learning about objects, simply watching another act on an object to produce images and sounds did not appear to result in motor re-activation in the brain (James & Swain, 2011), lending support to the idea that our
own actions on things – our embodiment – is important for learning. The question remains however, whether this sensory-motor association is specific to word processing, or if any sound will produce such co-activation of these systems after self-generated learning.

As our primary form of communication, the understanding and production of speech is a behavior that occurs rapidly and seemingly effortlessly within a typical learning environment. Because spoken language is learned in young children in every culture and under almost all circumstances, it is thought by some to be a unique human behavior (e.g. Chomsky & Ronat, 1998). In addition, language perception has special status in the brain. That is, humans have specialized systems that process language – presumably these systems have evolved to allow us to learn and understand language at an astonishing rate (e.g. MacWhinney, 1999). This is most apparent when young children learn their mother tongue, but continues through childhood, up to the point when learning language becomes more effortful (e.g. Bialystok, 1991). It is possible that learning words that are associated with actions leads to motor system re-activation during subsequent perception, but that other sounds may not. This would support the idea that learning language is a special form of perception, crucial for human development. However, an alternate explanation is that systems in the brain become associated through any self-generated experience – language is not special in this regard. This latter hypothesis would predict that any sensory-motor association formed during learning would result in motor system re-activation during subsequent perception - be it visual, auditory, haptic, etc.

The current study addresses these competing hypotheses by requiring children to learn associations among actions and sounds that are not words. If language is special in producing auditory-motor associations in the brain, then motor re-activation should not be seen after learning in this case. However, if motor re-activation does emerge after this learning, then it is the association that results in re-activation and not language specifically. Results of our research suggest the latter – that any association during learning will result in sensory-motor associations forming in the neural systems that support these functions. Further, we test visual perception of actions after active vs passive learning for two reasons: a) to compare our findings here with our previous work on visual perception after active learning (James, 2010; Butler & James, 2011; James & Swain, 2011), and b) to see if viewing videos are different from viewing static images after learning. However, these finding are not stressed here, as our primary goal was to investigate sound perception.
METHODS

Participants

Participants were 18 children between the ages of 5 years 1 months and 7 years 1 month (9 males). All children had normal or corrected to normal vision and normal hearing as reported by parents. Participants were not on any medications and had no history of neurological compromise. All were delivered at term without a record of birth trauma. Parents reported right-hand dominance for 16 participants and left-hand dominance for 2 participants. Functional imaging data from four participants was not gathered due to discomfort in the MRI environment. Thus, data from 14 individuals is reported here, 7 males, ranging from 5 years 4 months to 7 years 1 month.

General procedure

Participants underwent both a training session and a test session. The training session was performed outside of the fMRI environment. There were two within-participant conditions: one in which the participants performed an action on an object themselves (active condition), and another in which the experimenter performed the action and the participant observed her (passive condition). The actions were performed on novel, 3-D objects that produced a sound when a specific action was performed on them (see Figure 1). This resulted in learning an association among an unlabelled action, a novel object, and a sound. Stimulus exposure in the active and passive conditions was equated for each participant. Subsequent to this training, an fMRI session was performed to test 1) whether or not the learned actions would activate the motor cortex when the novel sounds were heard (auditory perception), when the novel objects were seen (visual perception), and/or when the novel objects were heard and seen (auditory + visual perception); and 2) in each perception condition, whether motor system activation would occur after self-generated action (active condition) and/or after observing the action of another (passive condition). As control conditions, novel sounds and novel objects that were not experienced in the training session were also presented to the participants.
Training Stimuli

The objects used had particular actions and sounds associated with them. The actions themselves could be performed at different time scales (faster, slower), and therefore the sound produced and the actions were somewhat reliant of the actor. However, we tried to equate the actions as much as possible (and therefore the resultant sounds they made). The actions and their associated sounds are presented in Table 1. Participants learned actions/sounds for 10 novel objects, 5 through active interaction with objects and 5 through passive observation of actions. The objects acted upon were novel, three-dimensional plastic objects, painted with monochromatic primary colors. The objects were approximately 12 X 8 X 6 cm and weighed approximately 115 g. Each object was constructed from 2-3 primary shapes (see Figure 1). When objects were acted upon, their shape changed, and each action was unique to each object. For example, one object required pulling out a retractable cord from its center. When the objects were not acted upon, the action was not afforded by their appearance alone – that is, it was not obvious how to interact with the object upon visual perception. Each object made an associated sound when they were acted upon, and it was the sound produced that was of specific interest in this study.

Figure a) Examples of 3-D plastic objects used during training and testing. Images are taken from video clips from the MRI testing stimuli.
Table 1.
Object label, action description and sound description of stimuli used in the training sessions.

<table>
<thead>
<tr>
<th>Object Label (not given to subject)</th>
<th>Action description</th>
<th>Sound description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Yocking</td>
<td>Pulling out part</td>
<td>Paper cutter</td>
</tr>
<tr>
<td>Green ratchting</td>
<td>Putting part onto object</td>
<td>Clinking</td>
</tr>
<tr>
<td>Purple sprocking</td>
<td>Pushing down part</td>
<td>Squeeking</td>
</tr>
<tr>
<td>Red tilfing</td>
<td>Pulling apart</td>
<td>Scraping plastic</td>
</tr>
<tr>
<td>Yellow wilping</td>
<td>Pulling up part</td>
<td>Squishing</td>
</tr>
<tr>
<td>Blue quaning</td>
<td>Rotating part</td>
<td>Stirring ice</td>
</tr>
<tr>
<td>Red luings</td>
<td>Pushing button</td>
<td>Glass breaking</td>
</tr>
<tr>
<td>Purple manuing</td>
<td>Pulling apart</td>
<td>Radio white noise</td>
</tr>
<tr>
<td>Red nooping</td>
<td>Pushing two parts together</td>
<td>Metal scraping</td>
</tr>
<tr>
<td>Yellow panking</td>
<td>Pushing part into hole</td>
<td>Toy rattle</td>
</tr>
<tr>
<td>Purple pancing</td>
<td>Closing/opening</td>
<td>Buzzing</td>
</tr>
<tr>
<td>Blue tooping</td>
<td>Pushing in part</td>
<td>Swirling water</td>
</tr>
<tr>
<td>Green laking</td>
<td>Pushing down top</td>
<td>Scissors cutting</td>
</tr>
<tr>
<td>Red patchoing</td>
<td>Shaking object</td>
<td>Bong bong sound</td>
</tr>
<tr>
<td>Yellow sapping</td>
<td>Bouncing object</td>
<td>Rattle snake</td>
</tr>
<tr>
<td>Purple lewing</td>
<td>Putting two parts into holes</td>
<td>Clicking</td>
</tr>
<tr>
<td>Blue dewing</td>
<td>Pushing object up</td>
<td>Screeching</td>
</tr>
<tr>
<td>Red toppzing</td>
<td>Stirring stick on object</td>
<td>Tweeting</td>
</tr>
<tr>
<td>Purple hanning</td>
<td>Rotating object</td>
<td>Tires screeching</td>
</tr>
<tr>
<td>Yellow silking</td>
<td>Pulling up parts</td>
<td>Twigs breaking</td>
</tr>
<tr>
<td>Green halking</td>
<td>Pushing down part</td>
<td>Barking</td>
</tr>
<tr>
<td>Blue razzing</td>
<td>Closing two sides together</td>
<td>Chairs scraping on floor</td>
</tr>
<tr>
<td>Red bulking</td>
<td>Pushing one then another button</td>
<td>Whistling</td>
</tr>
<tr>
<td>Purple gupping</td>
<td>Twirling piece on object</td>
<td>Siren</td>
</tr>
</tbody>
</table>

Training Procedure

The participant and an experimenter sat across from each other at a table. Each had an object set (five objects) randomly lying in a straight horizontal line directly in front of them. All objects were in full view of the participant. Object sets were counterbalanced across participants. The training procedure was structured like a game to engage the participants, resulting in all participants completing the training. A second experimenter acted as the referee and directed the session. The participant was told that they would be playing with some toys that have specific actions. The referee demonstrated the action associated with each of the participant’s objects; the experimenter then demonstrated the action associated with each of her objects. Five cards, colored as the monochromatic colors of each object, were placed in a bag. The child was instructed to draw a card, find the toy whose
color matched that of the card, and perform the appropriate action on the object 5
times. If required, the referee aided the participant in choosing the correct object
and performing the correct action. During learning, the experimenter would say
“Look at this, it does this (perform the action)” but did not draw attention to the
sound the object made. Actions were not labeled in this study. After drawing all of
the cards, the child was rewarded with a sticker. The experimenter drew cards and
performed actions on her objects just as the participant had done. During this
passive observation condition, the experimenter said “Now I am going to pick this
object…Look at what it does!” The participant and experimenter alternated turns of
the game. Thus, the children participated in both ‘active’ (self generated) and
‘passive’ (observing) conditions. This process was repeated 5 times total, resulting
in the experimenter and participant interacting with each of their objects 25 times
(5x5 active objects + 5x5 passive objects = 50 total exposures). After completing
the game, the referee suggested cleaning up the toys before continuing on to the
fMRI session. The participant was instructed to demonstrate the action on each
object at least one more time before handing it over to the referee to be put away.
The experimenter did the same with the ‘passive’ set of objects. Duration of session
was approximately 20-30 minutes.

fMRI Test Stimuli

Auditory stimuli consisted of sounds the 10 learned objects made during the
training session along with sounds of 5 unlearned objects. These sounds were pre-
recorded and presented to the participants through headphones in the fMRI facility.
They were the natural sounds that each action made. Visual stimuli were videos of
the actions being performed on learned objects as well as 5 videos of similar,
unlearned objects. The videos depicted the objects from a variety of planar (axis of
elongation was 0 degrees from the observer) and ¾ (axis of elongation was 45
degrees from the observer) viewpoints. The main shapes of the object could be seen
in every video. Auditory and visual stimuli were also combined at times to where a
visual video was shown in conjunction with the corresponding auditory sound.

Testing Procedure

After screening and informed consent given by the parent, all participants were
acclimated to the MRI environment by watching a cartoon in a MRI ‘simulator’.
The simulator is the same dimension as the actual MRI, and the sound of the actual
MRI environment is played in the simulator environment. This allowed the children
to become comfortable in the environmental set-up before entering the actual MRI
environment (see James, 2010). After the participant felt comfortable in this
environment, and if the parent was comfortable with the participant continuing, they
were introduced to the actual MRI environment.
Following instructions, a preliminary high-resolution anatomical scan was administered while the participant watched a cartoon. Following this scan, the functional scanning occurred. One to three runs were administered depending on the comfort of the participant. During functional scanning, auditory and visual stimuli were either presented separately or combined - where the appropriate sounds and object videos were linked together. Each run consisted of 6 blocks of stimuli (2 actively learned, 2 passively learned, and 2 novel objects). Blocks were 18-20 seconds long with 10-second intervals between blocks. Each run began with a 10-second rest period and ended with a 20-second rest period. This resulted in runs that were just under 3.5 minutes long. Each presentation of individual stimuli lasted for approximately 2 seconds. During auditory presentation runs, the participant was required to passively listen to the stimuli – no visual stimuli were presented. The participants were told to listen to the sounds that they hear. During the visual perception runs, the participant was required to passively view the videos – no auditory stimuli were presented, participants were instructed to simply watch the videos. During combined auditory and visual runs, the participant viewed the videos while the sound associated with the object played in tandem. Again, participants were instructed to watch the videos-in all three run types, no reference to the study session was made in the instructions. Neural activation, measured by the BOLD (Blood Oxygen Level Dependant) signal in the entire brain was then recorded during exposure to the stimuli. Imaging sessions took approximately 15 minutes in total.

After the functional scans, the participant was removed from the environment, debriefed, and rewarded for their time.

fMRI Acquisition

Imaging was performed using a 3-T Siemens Magnetom Trio whole body MRI system and a phased array twelve channel head coil, located at the Indiana University Psychological and Brain Sciences department. Whole Brain axial images were acquired using an echo-planar technique (TE = 30 ms TR = 2000 ms, flip angle = 90°) for BOLD based imaging. The field of view was 22 x 22 x 9.9 cm, with an in plane resolution of 64 x 64 pixels and 33 slices per volume that were 4mm thick with a 0 mm gap among them. The resulting voxel size was 3.0 mm x 3.0 mm x 4.0 mm. Functional data underwent slice time correction, 3D motion correction, linear trend removal, and Gaussian spatial blurring (FWHM 6 mm) using the analysis tools in Brain Voyager ™. Individual functional volumes were co-registered to anatomical volumes with an intensity-matching, rigid-body transformation algorithm. Voxel size of the functional volumes was standardized at 1 mm x 1 mm x 1 mm using trilinear interpolation. High-resolution T1-weighted anatomical volumes were acquired prior to functional imaging using a 3D Turboflash acquisition (resolution: 1.25 mm X 0.62 X 0.62, 128 volumes).
Data analysis procedures

Whole-brain Group contrasts were performed on the resultant data. The functional data were analyzed with a general linear model (GLM) using Brain Voyager’s™ multi-subject GLM procedure. The GLM analysis allows for the correlation of predictor variables or functions with the recorded activation data (criterion variables) across scans. The predictor functions were based on the blocked stimulus presentation paradigm of the particular run being analyzed and represent an estimate of the predicted hemodynamic response during that run. In addition, the movement made by the participants along 3 axes was also included as predictors in the analysis. Any functional data that exceeded 5mm of motion on any axis were excluded prior to analyses. This criterion resulted in excluding 2 blocks of data from one participant and 1 block of data from 3 participants. The attrition rate for this study was 4 individuals who were unable to commence the fMRI portion of the experiment. Exclusion of these data does not significantly alter the power of the present analyses. Data were transformed into a common stereotactic space (eg. Talairach & Tournoux, 1988) for group-based statistical analyses. Direct contrasts of BOLD activation were performed on the group between active sounds and passive sounds that were learned (new sounds were used as a baseline – see below for more detail). In addition, contrasts between activation during perception of objects that were learned actively vs passively were analyzed as well as the combined seeing + hearing the video presentations.

Contrasts in the group statistical parametric maps (SPMs) were considered above threshold if they met the following criteria in our random effects analysis: (1) significant at \( p < .05 \), corrected with the False Discovery Rate method (Benjamini, 1995), with a cluster threshold of 270 contiguous 1 mm isometric voxels. (2) Peak statistical probability within a cluster at least \( p < .001 \), corrected.
Recruitment of brain regions during sound perception that are significantly more active after learning sounds with active interaction than by passive experience. Left hemisphere is on the right. Coordinates reported for all regions are according to the conventions of Talairach & Tourneaux, 1988). 2A) Middle frontal gyri (-23,-9,60; 19,-9,60) Precentral gyrus (-30,-18,60) and Medial frontal gyrus (-5,-21,60). 2B) Middle frontal gyri-anterior (-43,22,40; 39,29,40), Cingulate cortex (-7,22,43; 5, 22,43) and (-13,-10,42), and the Precuneous (-11,-52,42). 2C) Activation in the Insula (35,6,6; -39,12,6) and the Superior Temporal Gyrus (52,-23,8). 2D) Activation in the Fusiform gyrus (-49,-51,-21). No regions were significantly more active after passive learning.
RESULTS

Auditory sound perception:

To localize brain regions that were recruited more after active learning than passive learning, a contrast was performed between actively learned sounds and sounds learned through passive observation. Novel sounds were used as a comparison, such that each contrast was structured as follows: (Active sound-novel sound) – (passive learned sound-novel sound). Hearing actively learned sounds resulted in significantly greater activation than passively learned sounds in the middle frontal gyrus bilaterally (Figure 2a), the left precentral gyrus and medial frontal gyrus (Figure 2a), the cingulate gyrus and precuneus (Figure 2b), the anterior middle frontal gyrus (Figure 2b) the insula bilaterally (Figure 2c), the right superior temporal gyrus (Figure 2c) and the bilateral fusiform gyri (Figure 2d).

![Figure 3](image)

Figure 3.
Regions that are significantly more active when viewing videos of objects and actions after active exploration during learning vs. passive observation during learning. 3A) Superior Frontal gyrus (29,-11,51) and the Inferior parietal lobule (-37,-50,51). 3B) Fusiform gyrus (-44,-43,-12). No regions were more active after passive observation in this contrast.
Visual object perception:

When contrasting actively learned with passively learned objects while participants viewed videos of the objects and actions (active learned video-novel action) – (passive learned action-novel action), three brain regions were recruited significantly more during active than passive learning: The left inferior parietal lobule (Figure 3a), the right superior frontal gyrus (Figure 3a), and the left fusiform (Figure 3b).

![Fusiform gyrus](image)

**Figure 4:**
Significantly greater activation in the see and hear video conditions after active exploration vs. passive observation during learning. The only regions more active here are the fusiform gyri (-19,-88,-17 to -32,-80,-17) and 29,-61,-17).

Seeing and hearing learned objects:

To determine activation to these conditions, we performed a contrast between (active see+hear-novel see+hear) – (passive see+hear-novel see+hear). In the see + hear condition, only the fusiform gyrus was more active, bilaterally, for the actively learned objects than the passively learned objects (Figure 4). There were no regions where passive observation resulted in greater activation than active exploration during learning.


**DISCUSSION**

As organisms that are situated in a given environment and interact with the world through our bodies, it stands to reason that we learn most efficiently when we are able to interact manually with objects. Mounting evidence supports the idea that active interaction leads to better learning in both adults (Harman et al., 1999; James et al., 2002; James, Humphrey & Goodale, 2001; Engelkamp & Zimmer, 1994; Masumoto et al., 2006, Butler & James, 2011) and children (James, 2010, James & Swain, 2011; Longcamp et al., 2005). However, the neural mechanisms that underlie active learning are not well understood. In adults, we know that a history of actions with objects results in activation of the motor system during visual perception – the very system that is used during learning. This ‘re-activation’ is thought to underlie the advantageous affects of active learning. That is, object representations incorporate the motor movements that are used during initial learning, and these motor programs, or plans, are active again during visual perception of the learned objects. Here we demonstrate four important findings that contribute to this body of knowledge. First, in children, learning about objects actively results in very different brain system recruitment than observing others act. Second, sounds learned actively result in motor, auditory, and visual system recruitment more than passive observation upon subsequent auditory presentation. Third, videos of objects that are learned actively result in motor and visual system recruitment more than passively learned objects, and fourth, when given all learned information about objects through multi-sensory presentation, active learning results in greater visual system recruitment than passive observation. We will discuss each of these results in turn.

**Sound perception:**

Perception of learned sounds is different depending on whether or not those sounds are learned through active engagement with an object or through passively observing another acting on an object. Specifically, the parts of the motor system, including the middle frontal gyrus (premotor areas), the precentral gyrus (primary motor strip), and the cingulate cortex are recruited after active learning. These regions have previously been shown to be recruited during auditory perception after active learning of verbs in children (James & Swain, 2011). As such, these results provide evidence that not only does language recruit motor regions after active learning, rather, any association formed between and auditory sound and an action will result in motor system re-activation upon subsequent auditory only presentation. In addition, the insula, an area regarded as involved with sensori-motor processing, was also engaged during auditory presentation after active learning, but not after passive observation. Previous work has shown that the activation of the insula is related to the experience of actions being performed by
oneself Farrer & Frith (2002). The activation of this region in the current study suggests reactivation of sensori-motor processes during auditory perception. This mirrors previous work showing that newly learned sound-action associations led to activation of the insula in adults during subsequent auditory perception (Mutschler et al., 2007). This, however, is the first study to show that this region reactivates in children. Being that the right superior temporal gyrus was recruited as a result of active learning suggests that the auditory association regions (STG) were also involved in reactivation more after active learning. Thus, a reactivation of the sensory system involved during encoding occurred— a result that has not yet been shown using the auditory modality.

Interestingly, and most surprisingly, the visual cortex was also recruited during auditory perception after active learning. In other studies, this region of the fusiform has been shown to be engaged after active learning more than passive learning, but only using visual perception during test (James, 2010). However, in the previous research testing auditory verb processing, this region was not active for active learning more than passive observation (James & Swain, 2011), nor was it active during verb perception in children (James & Maouene, 2009). Thus, it seems that sound perception may link to visual object representations more than auditory verb perception. Perhaps the sounds are linked more to the visual information than are verbs, but understanding why this is the case requires further research. We can speculate that perhaps active learning helps to create a stronger association between visual and auditory information. Indeed, we have recently shown that audiovisual associative recognition can be enhanced by active learning (Butler & James, 2011).

**Visual processing**

During visual presentation of learned objects, active learning resulted in a recruitment of the IPL more than did passive learning. This is a region associated with grasping in humans and monkeys (eg. Culham & Valyear, 2006) but also and perhaps more relevantly, with tool use (Johnson-Frey, 2004). If our novel objects are represented as known tools after active interaction, then activation in this area is not surprising. Possibly the role of the inferior parietal lobule is not involved in tool representation as much as it is involved in action representation – being located close to parietal regions associated with grasping. Videos associated with actions may recruit regions that are associated with the actual actions, whereas hearing sounds may not activate the actual action patterns associated with interaction, but rather the sensori-motor representation associated with the frontal cortices. The IPL has been implicated previously in action observation that is specific to actions performed by the observer (Clavo-Merino et al., 2005).
In addition, re-activation of the visual cortices was apparent after active learning more than passive observation. This finding supports and extends previous work showing visual system reactivation after active versus passive learning (James, 2010).

Interestingly, visual perception did not recruit auditory cortices, suggesting that perceiving the videos did not access the auditory sounds that the videos produced during training. This is different from the finding that hearing sounds did activate the studied visual images associated with them (see above). This unidirectional re-activation is a novel finding, and one that requires further study.

**Multimodal presentations**

The only regions more active after active engagement with objects than passive observation when viewing multimodal presentations was a large section of the visual association cortex, namely, the fusiform gyrus. Again, we see that this region is sensitive to active learning, suggesting that the fusiform receives input from motor regions that are used during learning. We have found this previously with young children (James, 2010) and adults (Butler & James, 2011; James & Atwood, 2009). Sensori-motor processing in the fusiform gyrus has been shown to emerge during perception of some forms that we have motor experience with—such as letters (James & Gauthier, 2006; James et al., 2005). However, the exact role of this region for the integration of sensori-motor signals is not known. One study found that the fusiform was recruited during the integration of visual and haptic information (Kim & James, 2010), a result that may reflect the role of the motor system in haptics and its integration with visual processing. We have hypothesized that motor efferent signals are integrated into visual representations after motor experience, and the end point of these efferents may be the fusiform gyrus (James, 2010; James & Gauthier, 2006).

Therefore, self-generated actions were required for the emergence of motor system recruitment during auditory processing in the developing brain. Numerous theories have suggested that action and perception, when coupled, form representations that may be accessed by perception alone (e.g., Prinz, 1997) – that these representations contain the motor programs associated with the percept. In addition, performed actions will activate visual cortices without concurrent visual stimulation just as perception can activate motor systems without concurrent movement (e.g., James & Gauthier, 2006). The frontal system codes information that associates previously performed actions with present perceptions, and is therefore recruited to a significantly lesser extent during perception after action observation. This finding appears, on the surface, to stand at odds with work showing frontal activation during action observation (e.g. Rizzolatti & Craighero, 2004), however, our contrasted conditions both observe actions at test, thus removing any unique affect of this behavior.
It is possible that actively interacting with objects allows for the participants to ‘imagine’ the actions upon subsequent auditory or visual presentation of the objects, resulting in differences between active and passive experience. In other words, covert enactment of the motor patterns associated with the actively learned objects may have recruited the brain regions in this study (i.e. Jeanorrod, 2001). This would contrast somewhat with our interpretation that the sensori-motor information gets associated through the learning experience, resulting in an activation pattern that accesses both sensory and motor information directly. The current work cannot distinguish between these two alternatives, as we cannot ascertain whether activation seen here is due to imagining actions or automatic access to learned actions. Either way, however, active experience changes how the brain processes subsequent auditory and visual information. The level at which this experience affects subsequent perception, be it from directly accessing prior motor activity or through allowing actions to be imagined, is an important question for future work that can distinguish timing information in neural processing – an obvious short coming of fMRI blocked designs (eg. Hauk et al., 2008).

In the adult brain, we know that action words, and in some cases, object perception recruits motor systems. Here we show that in order for this adult-like sensori-motor response to occur, children need to actively interact with objects in the environment. Therefore, we provide initial evidence for the types of interactions that produce adult-like neural responses. In addition, we provide the first evidence that shows that sounds associated with objects also recruit these sensori-motor pathways in the young child. Providing such developmental information is important for understanding the cascading effects that certain experiences have on neural response patterns, and therefore, on human cognition.

Furthermore, this work allows us to come closer to understanding the role of the frontal, parietal and occipital systems for object and sound processing. Based on our present results, we propose that one function of the motor association areas is to associate past experience with present perception, but in a fairly specific manner – associating a history of self-generated actions with perception. At least in the developing brain, perception and action become strongly linked as a result of self-generated action: in general, experience must be sensori-motor and not sensation-of-motor.

ACKNOWLEDGEMENTS

The authors wish to acknowledge all the children and their parents who participated in this study, and also, our MR technicians, Thea Atwood, Rebecca Ward, Ben Pruce, and Chris Chung. Thanks to Linda B. Smith, Susan S. Jones and Thomas W. James for helpful comments on earlier versions of this manuscript. This work was partially supported by the Lilly Endowment Fund MetaCyte initiative and the Indiana University Faculty Support program to KHJ.
REFERENCES


EFFECTS OF SENSORI-MOTOR LEARNING ON MELODY PROCESSING ACROSS DEVELOPMENT

Elizabeth M. WAKEFIELD*, Karin H. JAMES
Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA

ABSTRACT

Actions influence perceptions, but how this occurs may change across the lifespan. Studies have investigated how object-directed actions (e.g., learning about objects through manipulation) affect subsequent perception, but how abstract actions affect perception, and how this may change across development, have not been well studied. In the present study, we address this question, teaching children (4-7 year-olds) and adults sung melodies, with or without an abstract motor component, and using functional Magnetic Resonance Imaging (fMRI) to determine how these melodies are subsequently processed. Results demonstrated developmental change in the motor cortices and Middle Temporal Gyrus. Results have implications for understanding sensori-motor integration in the developing brain, and may provide insight into motor learning use in some music education techniques.

KEYWORDS: fMRI, visual-motor learning, music, MTG

Perception of the world requires interaction between our bodies and the environment. Cognition, being determined by our perceptions, can thus be considered embodied – not only formed by our active interactions with the world, but also augmented by our ongoing interactions. One line of evidence supporting this embodied framework is that motor systems are recruited when organisms perceive their environment: a history of our actions is stored and reactivated upon subsequent perception of learned stimuli (e.g., Arevalo, 2008; Arevalo et al., 2007; Chao & Martin, 2000; Hauk, Johnsrude, & Pulvermüller, 2004; James, 2010; James & Atwood, 2009; James, Butler, & Mueller, 2008; James & Gauthier, 2006; James & Maouene, 2009; James & Swain, 2011; Longcamp, Anton, Roth, & Velay, 2003;
Nyberg et al., 2001; Pulvermüller, Harle, & Hummel, 2001; Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005; Pulvermüller, Lutzenberger, & Preissl, 1999; Saccuman, Cappa, Bates, Arevalo, Della Rosa, Danna, & Perani, 2006). Furthermore, the same perceptual task will recruit different neural systems depending upon whether the task was initially learned with or without motor system involvement (e.g., Hauk et al., 2004; James, 2010; James & Atwood, 2009; James & Maouene, 2009; James & Swain, 2011; Lahav, Saltzman, & Schlaug, 2007; Nyberg et al., 2001).

In adults, this influence of the motor system has been demonstrated during perception of items such as tools (Chao & Martin, 2000), kitchen utensils (Chao & Martin, 2000), letters (Longcamp et al., 2003; James & Gauthier, 2006), action-related words – both verbs (e.g., Hauk et al., 2004; Pulvermüller et al., 2001; Pulvermüller et al., 2005; Pulvermüller et al., 1999) and nouns (Saccuman et al., 2006; Arevalo, 2008; Arevalo et al., 2007), during action commands (Nyberg et al., 2001) and during perception of action in some cases (e.g., Buccino et al., 2004; Decety, Chaminade, Grezes, & Meltzoff, 2002; Grezes, Armony, Rowe, & Passingham, 2003; Iaconobi & Dapretto, 2006; Iacoboni, Woods, Brass, Bekkering, & Mazziotta, 1999; Rizzolatti & Craighero, 2004). Other examples of the motor system influencing perception come from studies on gesture which show that individuals’ gestures when explaining a task affect their representation of the task, and how they will perform it in the future (Beilock & Goldin-Meadow, 2010; Goldin-Meadow & Beilock, 2010), and from studies of musicians who show motor reactivation when listening to a piece they have learned on their instrument (e.g., Bangert et al., 2006; Baumann et al., 2007; Haslinger et al., 2005; Landau & D’Esposito, 2006; Margulis, Mlsna, Uppunda, Parrish, & Wong, 2009). Thus, there are many examples of how adults’ perceptions are shaped by motor system involvement.

Whether or not motor experience also leads to the recruitment of different systems during visual perception in children has been addressed by a handful of studies. In particular, some researchers have investigated the neural substrates of perception in children, comparing ‘active’ learning – learning with actual manual exploration with ‘passive’ learning – learning only through the perceptual system. For example, children showed increased motor system activation during visual and auditory object recognition tasks if they had explored objects actively, rather than passively, prior to a recognition task (James & Swain, 2011). In addition, children learning to recognize letters showed an increase in neural activation in visual association areas when viewing letters if they had been taught letters actively, through printing, as opposed to when they had been taught letters passively, through visual recognition alone (James, 2010). Finally, studies on gesture use have shown that learning a concept with the help of gesture results in learning that is retained across time, compared to learning without gesture (Cook, Mitchell, & Goldin-Meadow, 2008). The findings from these studies suggest that utilizing a motor
system component while learning changes how the brain processes subsequent information in children, not just in adults.

In the majority of the research discussed, the actions performed by children and adults have directly changed something about their perception of the object being acted upon. When a child writes letters, she actively creates figures on a paper. When an adult hammers a nail or strikes a piano key, the movement of the hammer or finger is casual. In the case of gesturing, gestures are not performed on objects, but are often representative of actions. Gestures are thus concretely representative, although they are not directly causal. This leads to an interesting question: how are actions that are not directly causal or concretely representative, but more abstractly representative through learned associations, connected to perceptions, and does this change across development? If a history of learning through motor interaction facilitates learning and/or allows the individual to associate more abstract motor movements with perception, then we would expect to see a difference in performance and neural activation patterns between adults and children. Addressing this question will further our understanding of how and when our actions influence cognition, and whether there is a limit to embodied cognition. In the present study, we address this question through a paradigm driven by a real world use of non-causal, abstract, but representative actions in music education.

Non-causal motor movements are often incorporated into music education techniques to increase the speed of acquisition of musical skill, and the musicality of students (Shiobara, 1994; Sidnell, 1986). Teachers employing the Dalcroze method, developed in the early twentieth century, stress movement to music as a way for children to internalize rhythm, and improve their musicality (i.e., eurhythmics) (Seitz, 2005). Educators working from the Kodály and Orff traditions, which were also developed in the early twentieth century, teach sight singing with motor signs (i.e., Curwen hand signs) to help children learn intervals and maintain good pitch memory. Previous literature indicates that the way we internalize and conceptualize music is integrally related to movement (e.g., Eitan & Granot, 2006; Keller & Koch, 2008), thus it is not surprising that educators would naturally incorporate motor components into music education techniques. Given this strong connection between music and movement, adding a movement component during music learning may be beneficial because it grounds learning in the physical world; however, there are very few empirical studies that have been conducted to evaluate the actual effects of using these techniques, as compared to traditional methods that do not stress a motor component. Of the studies that do exist, some show that motor movements aid learning (Liao & Davidson, 2007), whereas others indicate that a motor component does not appreciably improve music-learning ability (Cassidy, 1993). These inconclusive results may stem from the type of movement being used. Liao and Davidson instructed children to create movements that they felt represented the music, whereas Cassidy (1993) utilized Curwen hand signs: a system in which hand signs are paired with single steps of the musical scale.
Although some hand signs are related to the scale step they represent (i.e., the leading tone is paired with an upward pointing of the finger, as if to point to the tonic), most are only meaningfully connected to the pitches they represent after being learned, and none are intuitively connected outside of the context of the scale. The movements used in Liao and Davidson’s study, then, are similar to the causal movements that are known to affect perception. The abstract movements, used by Cassidy (1993) may not affect perception in the same way, although anecdotal evidence of their benefit in music education suggests that the movements are helpful in some way.

In the present study, we investigated how non-causal, but associable movements can change neural processing during the learning of sung melodies, and affect subsequent recognition of melodies by adults and children. Importantly, we compared learning with and without a visuo-motor component within both populations; then compared this difference between the groups. Based on previous literature, we hypothesized that (a) we would see reactivation of the motor system to melodies learned in the visuo-motor condition, as opposed to other conditions that did not include a motor component if associable motor movements are processed in a similar way to more causal movements, (b) if motor movements must be concretely representative of the to-be-learned stimuli, then we may not see any facilitation in this study, due to the abstract nature of our motor movements and their arbitrary (although consistent) connection with the to-be-learned melodies, but if the brain processes abstract and more direct, causal movements similarly, we may see an improvement in learning with the motor system, as has been seen in the literature (e.g., gesture studies).

**EXPERIMENT 1**

**Method**

**Participants**

Fifteen undergraduate and graduate students (7 male and 8 female), at Indiana University participated in the study. All participants were right-handed, native English speakers who reported no history of neurological or psychiatric disorders and reported normal or corrected-to-normal visual acuity. Participants were between the ages of 19-27 years ($M = 24$ years). Informed consent was obtained from 17 university students, according to the guidelines of the Indiana University Internal Review Board. Twelve of the 15 participants had learned to read musical notation, but only 9 remembered how to read music at the time of the experiment. Adults had various amounts of instrumental training. The data from 2 participants...
were excluded from the analysis, based on a failure to successfully complete a required training session.

**Stimuli**

*Melodies.* Twenty-eight melodies were composed for the study and 12 of the 28 melodies were taught to participants during a training session. All 28 melodies were recorded for use in an fMRI session. Melodies consisted of four short sections (i.e., measures), and were composed in the style of simple children’s songs, with simple rhythmic patterns (See Figure 1a for melody example). Melodies were recorded by a female vocalist as two separate stimuli, the first two sections and the second two sections, at 96 beats per minute (bpm) using an Edirol R-09 24-bit WAVE/MP3 Recorder (Roland Corporation, Nakagawa, Japan). Each of the 56 resulting half melodies was approximately five seconds. Melodies were edited on a Macbook using Audacity 1.2.6. Typicality ratings were not acquired, however, none of the participants commented that melodies were familiar or similar to other songs, and melodies were specifically composed *not* to resemble known songs.

*Hand signs.* Eight hand signs were used during training. These hand signs were adapted from preexisting hand signs created by John Curwen. Each hand sign corresponded with a discrete pitch, and a discrete modified solfège syllable (hand signs are depicted in Figure 1c). Hand signs were created so as not to be intuitively representative of particular pitches, thus are not directly related to the melodies – the association must be learned through the training.

*Figure 1.*
(a) An example of a melody learned by adult participants (b) An example of a visual representation of a melody learned by adults (c) Hand signs taught to adults during the motor learning condition.

![Figure 1](image-url)
Procedure

Training session. Participants underwent one training session in which they learned 12 melodies under three conditions from an experimenter; four melodies per condition. Conditions were presented in a randomized order across participants. In an auditory condition, participants heard melodies that were sung using nonsense syllables analogous to solfège syllables (e.g., Fo, Ga, Di instead of Do, Re, Mi), and repeated them. In a visuo-motor condition, in addition to hearing melodies, participants learned hand signs corresponding to notes in the melody with their right hand (analogous to Curwen hand signs, see Figure 1c), and repeated both the melodies and hand signs. When the experimenter produced the melodies, vertical height of the hand signs corresponded with the pitch being sung (i.e., hand signs paired with high pitches were produced spatially higher than hand signs paired with low pitches). Participants were not specifically told to incorporate the spatial aspect of the hand signs and observation of learning demonstrated that hand signs were not consistently produced at particular heights. In a visual condition, in addition to hearing melodies, the syllables being sung were presented visually (see Figure 1b), and participants repeated the melodies. Therefore, for the visual and motor conditions, there was additional input to the system—in one case visual, in the other, motor.

Melodies were taught using a recall procedure adapted from Racette and Peretz (2007). For each melody, participants listened to the experimenter sing a complete melody once, after which the experimenter taught participants the melody in sections. To learn sections of the melody, participants listened to the experimenter sing a section, then sang the section with the experimenter, then sang the section by themselves. Melodies were made up of four sections (i.e., measures). Using the procedure described, participants learned the first and second sections; then sang the first and second sections together. The third and fourth sections were learned the same way. Finally, participants performed the complete melody. Training for the visual and auditory melodies was complete when participants had sung each melody accurately without help from the experimenter. For melodies learned with a visuo-motor component, participants were required to produce the hand signs correctly in addition to accurately performing the melodies.

fMRI session. Participants underwent a scanning session the day following training, which consisted of six functional imaging runs and a high-resolution structural scan. Functional localization: Participants had two block-design runs to localize functional regions of interest in the brain, including areas involved in auditory, motor, and visual processing. To localize regions for auditory processing, participants heard 3-note sung melodies. To localize motor regions, participants made hand movements of their choice with their right hand, and to localize visual processing areas, they viewed modified solfège syllables that were similar to, but
not the same as those used in the training session. Blocks were 14 seconds in duration, with a 10 second inter-block interval. There were four blocks of each stimulus type, presented randomly within a run. Runs were 5 minutes and 8 seconds, containing 12 blocks total. There were two localizer runs in total. **Experimental sessions:** Following the localizer runs, four additional runs were presented in a fast event-related design, during which participants listened to parts of melodies (the first or second half of each melody). These melody segments were either learned in the training session, or were novel (unlearned) melodies, composed using the same guidelines as the learned melodies. Participants were not presented with visual stimuli, or asked to produce motor movements during these runs; they simply listened to melodies they had learned in the training session, and the novel, unlearned melodies. Participants indicated whether they remembered learning each melody via button-presses on a left-hand paddle. The melodies were presented in a random order and no hand movements were detected during scanning, as confirmed by a camera in the bore of the magnet, except for the button press response (which was present during all melody conditions). Melody stimuli were five seconds long, with inter-stimulus intervals (ISIs) of 8, 10, or 12 seconds. There were 32 presentations of the melodies per run, 24 learned and 8 unlearned and four runs in total. Different unlearned melodies were used in each run. The final scan was a high-resolution anatomical scan, onto which the functional data were overlaid for the purpose of analysis.

Visual stimuli were presented via a Mitsubishi XL30 projector, which projects images onto a screen that participants viewed through a mirror from the rear part of the bore of the scanner. Auditory stimuli were presented via headphones worn by the participants (Siemens). Both types of stimuli were presented via SuperLab 4.0.7b software on a Macintosh MacBook laptop.

**fMRI parameters.** Imaging occurred in a 3-Tesla Siemens Magnetom Trio whole-body MRI system and a phase-arrayed eight-channel head coil in the Imaging Research Facility located in the Indiana University Psychological and Brain Sciences department. The field of view was 220 mm, 100%, with an in-plane resolution of 64x64 pixels and 33 slices per volume (3.4 mm thick, no gap). Images were acquired using an echo-planar technique (echo time, TE = 30 ms, time of repetition, TR = 2,000 ms; flip angle = 90°) for brain oxygen level-dependent (BOLD) imaging. Although sparse sampling is often used when auditory stimuli are presented, it was determined that the possible effect of scanner noise was equated for all conditions, and thus subtracted out of any comparisons among conditions. Analysis was conducted using Brain Voyager. Images were 3D motion-corrected, linear trend was removed, and Gaussian spatial blurring (FWHM 6 mm) was applied. Individual functional volumes were aligned with anatomical volumes using an intensity-matching, rigid-body transformation algorithm. Individual anatomical
volumes were normalized into Talairach space using the eight-parameter affine transformation (Talairach & Tournoux, 1988).

**fMRI data analysis procedures.** To localize regions of interest (ROIs), a whole-brain analysis of the data from the functional localizers was performed. In each individual, the contrasts of interest were performed: Motor (hand movements) versus rest, visual (perception of syllables) versus rest, and auditory (melody perception) versus rest. Based on these contrasts, five ROIs were identified: one motor region in the left somatosensory cortex (resulting from motor vs. rest; peak average Talairach coordinates x(-38), y(-26), z(53)), one visual region in the occipital cortex of each hemisphere (resulting from visual vs. rest; peak average Talairach coordinates x(-28), y(-81), z(-1) and x(34), y(-79), z(-1)), and one region involved in auditory processing in each temporal lobe (resulting from auditory vs. rest; peak average Talairach coordinates x(-50), y(-19), z(8) and x(58), y(-16), z(3)). Within these individually defined ROIs, we extracted the percent signal change of the BOLD response for the experimental runs in which the learned and novel melodies were presented. A deconvolution procedure was used to separate the BOLD signal of the melodies from the ISIs due to the short duration of the ISIs relative to the stimulus events. The deconvolution parameters in Brain Voyager were used to achieve the deconvolved data. Within each deconvolved time course for an event, the peak percent signal change relative to baseline was used as a dependent measure of BOLD response. Data from each ROI were then averaged across participants. A repeated measures analysis of variance (ANOVA) was conducted on the data from each ROI, followed by planned comparisons in each ROI between conditions: Motor versus Auditory, Motor versus Visual, and Visual versus Auditory.

**RESULTS**

**fMRI Behavioral Performance**

During the scanning session, participants were asked to listen to melodies and indicate whether they learned the melodies in the training session on the previous day, or whether the melodies were new (i.e., novel). A d prime analysis was conducted to determine a percent correct value that accounts for response bias, correcting for hit and false alarm rates. Based on this analysis, participants correctly identified 95.4% ($SD = 0.05$) of the melodies from the motor training condition, 96.4% ($SD = 0.06$) from the visual training condition, and 98.2% ($SD = 0.02$) from the auditory condition. To determine whether participants were making correct identifications of learned melodies based on recognition of melodies instead of random guessing, $t$-tests were conducted between the percentage of correct
identifications for each condition and chance ($p = 0.25$). The percentages of correct identifications made by participants were significantly above chance, $t(14) = 14.83$, $p < .001$; $t(14) = 15.19$, $p < .001$; and $t(14) = 52.95$, $p < .001$, for motor, visual, and auditory melodies, respectively. A repeated measures ANOVA was conducted to determine whether participants were more proficient at learning melodies in a particular condition. There was no significant difference between percentages of correct identifications between melodies in the visuo-motor, visual, and auditory conditions, $F(2, 14) = 1.98$, ns., indicating the melodies in the three conditions were learned equally well.

**fMRI ROI Analyses**

_Auditory ROIs._ Repeated measures ANOVAs were conducted on the percent BOLD signal change in both the right and left auditory ROIs for melodies in the three learning conditions (visuo-motor, visual, and auditory). A significant effect of learning condition was found in both the left auditory ROI, $F(2, 14) = 15.872$, $p < .001$, and right auditory ROI, $F(2, 14) = 8.375$, $p < .001$, (see Figure 2). Planned comparison within-group $t$-tests were conducted in both the left and right auditory ROIs. The main effect of learning condition shown bilaterally in the auditory ROIs was driven by significantly greater activation to the melodies learned in the auditory condition than to those learned in the visual condition, ($t(14) = 4.73$, $p < .01$, and $t(14) = 3.53$, $p < .01$ in the left and right ROIs, respectively), and by greater activation to melodies learned in the auditory condition than to those learned in the visuo-motor condition, ($t(14) = 3.92$, $p < .01$, and $t(14) = 2.92$, $p < .01$ in the left and right ROIs, respectively). No other significant differences were found between percent BOLD signal change to melodies learned in different conditions. These results indicate that learning melodies through auditory perception resulted in a re-activation of auditory regions. Re-activation of auditory cortex did not result after learning melodies with visual stimuli (in addition to auditory) or after learning with hand signs (in addition to auditory).
Figure 2.
The bar graph displays the average of the peak BOLD activation for each individual to each melody type (visual, motor, and auditory) taken from the left and right auditory ROIs. The specific talairach coordinate from which the peak BOLD activation to the melody types was extracted varied among individuals. Individual talairach coordinates were determined by performing a contrast between the auditory localizer and rest, and using the peak activation to extract data from the combined runs in which participants listened to melodies from the learning conditions.

Visual ROIs. Two repeated measures ANOVAs were run on the left and right visual ROIs, again to determine whether the condition in which melodies were learned would affect percent BOLD signal change. No significant effect was found in either the left, $F(2,14) = 1.974, p = 0.16$, or right visual ROIs, $F(2,14) = 1.125, p = 0.34$.

Motor ROI. Finally, a repeated measures ANOVA was conducted to determine whether significant differences in activation to melodies learned in the three conditions occurred in the motor ROI. No significant difference was found between learning conditions, $F(2, 14) = 1.39, p = 0.27$. This lack of effect may be driven by the method by which the ROI was defined. The area of the brain involved in producing motor actions, which was found using the functional localizer, may be different than the area of the brain activated when remembering motor movement. Thus, additional direct contrasts were performed to further investigate those data.
fMRI Direct Contrasts

Three whole-brain contrasts were performed: one contrasting BOLD activation to melodies learned in the visuo-motor condition to those learned in the visual condition, a second comparing visuo-motor to auditory conditions, and a third comparing visual to auditory conditions. There were two areas that showed significantly greater activation to melodies learned in the visuo-motor condition, as compared to those learned in the visual condition: a region of the left precentral gyrus (PrCG) and a region of the left middle temporal gyrus (MTG) (see Figure 3). There were no regions that showed a significant difference in activation to melodies learned in the visuo-motor condition, as compared to those learned in the auditory condition, nor were there significant differences in activation to melodies learned in the visual or auditory conditions.

Figure 3.
Whole-brain contrasts between melodies learned in the visuo-motor condition and melodies learned in the visual condition. When participants were auditorally presented with melodies learned in the visuo-motor condition, a region of the lPrCG showed greater activation than when participants were presented with melodies learned in the visual condition (Talairach coordinates x(-60), y(-4), z(30)). When participants were auditorally presented with melodies learned in the motor condition, a region of the lMTG showed greater activation than when participants were presented with melodies learned in the visual condition (Talairach coordinates x(-60), y(-49), z(2)).
DISCUSSION

Our ROI analyses demonstrated that learning melodies by simply listening to, and repeating them, resulted in greater auditory cortex re-activation bilaterally than learning melodies with either a visuo-motor or visual component. However, there was no significant re-activation in the motor, visual, and auditory ROIs as a result of learning melodies with an auditory and visual component, or as a result of learning melodies with an auditory and motor component. In contrast with the ROI findings, our direct whole-brain contrasts of the experimental runs demonstrated that learning melodies with motoric hand signs activated a part of the cortical motor system and the MTG upon subsequent auditory presentation of the learned melodies.

Interestingly, there were no re-activations found in the visual cortex upon subsequent auditory presentation of melodies, and, there was no behavioral effect of the different learning conditions: all melodies were learned equally well. The significance of these findings is discussed below.

1. Reactivation of auditory cortex to auditory melodies. The auditory cortex was reactivated bilaterally upon subsequent presentation of melodies learned with the auditory system alone to a greater extent than to those learned with the auditory and motor or visual systems. This finding suggests that a trace of the learning experience is stored in the auditory system, but only when auditory information is not paired with another modality during learning. This activation seems to encompass both primary and secondary auditory processing areas.

One possible explanation for this finding is that participants were dividing their attention between listening to the melodies and looking at the visual representation of the melodies, or making motoric hand signs in the visual and visuo-motor conditions, respectively. Specifically, in the visual condition, using the visual representation of the melodies may have allowed the participants to attend less closely to what the experimenter was singing, as they did not have to listen to the experimenter to know what syllables to sing. If this is true, the use of the visual representation may have led to less effortful auditory processing. In the visuo-motor condition, as previously discussed, the hand signs were not intuitively representative of the pitches they were associated with, thus participants may have had to concentrate on learning the motor movements, perhaps attending less to auditory processing. Therefore, although multisensory processing is usually beneficial, in this case, the addition of either the visual or visuo-motor component to the melody learning may have caused less intentional processing of the auditory information, resulting in no auditory re-activation in these cases. Alternatively, the findings could have resulted from competitive inhibition of auditory processing by the motor and visual cues that were learned in the visuo-motor and visual conditions, respectively. Forming associations between the visual cues and the
melodies, or between the motor hand signs and the melodies may have interfered with deeper auditory processing in these conditions, whereas auditory processing would not have been inhibited when participants learned melodies in the auditory alone condition. In support of this idea, research has shown that when individuals simultaneously attend to stimuli in more than one modality while only making responses based on one of the stimuli sets, as opposed to only attending to the set of stimuli being responded to, performance deficits are seen (Dell’acque & Lolicoeur, 2000; Fernandes & Moscovitch, 2000; Jolicœur, 1999). For example, Dell’acque and Lolicoeur (2000) showed that when individuals were asked to pay attention to the pitch of a tone that accompanied stimuli in a visual discrimination task, participants were less accurate than when they were told they could ignore the tone. This demonstrates that auditory information can interfere with the encoding of visual information. In other words, whereas participants were being tested on the visual information, having to attend to and remember tones interfered with learning this visual information. In the present study, the opposite phenomenon may have been occurring – the main task was to learn set of auditory stimuli (i.e., the melodies), but the additional visual and visuo-motor stimuli may have been taxing attentional resources that would otherwise have been allocated to auditory processing. Moreover, in a second study, participants were asked to listen to, and encode a word while paying attention to various visual stimuli sets (Fernandes & Moscovitch, 2000). The authors found that when the visual and auditory sets were lexical in nature, encoding interference occurred. In our study, participants saw syllables visually while hearing and singing these syllables in the melodies; thus, the visual representation of the syllables may have interfered with the auditory encoding.

2. Reactivation of PrCG to motor melodies. Associating melodies with hand signs during learning resulted in the recruitment of the left PrCG upon hearing the learned melodies again. This result is consistent with several findings suggesting that when stimuli are learned actively, that is, with motor system involvement, the motor system is re-activated upon subsequent perceptual encounters with the stimuli (e.g., D’Ausilio, Altenmüller, Belardinelli, & Lotze, 2006; Lahav et al., 2007; Margulis, et al., 2009). Note that this activation was not due to button presses, as these occurred to all test stimuli, those learned with motor signs, and with visual stimuli.

One suggestion of the mechanisms underlying reactivation is the common coding hypothesis, which posits that there is a common representation in the brain for perceived stimuli and planned actions (e.g., Prinz, 1997). In the case of the present study, the participants would therefore have coded the sung melodies and the hand signs that accompanied these melodies together, such that upon later presentation of the melodies, the stored motor movements were reactivated. This finding suggests that for adults, a motor component that is not directly causing the pitches or syllables produced was processed in a similar way, at least in the
premotor system, as motor movements more directly connected to a representation. Even abstract motor representations can thus cause changes in cognitive processing.

3. Activation of MTG to melodies learned with a motor component. The whole-brain contrast that was performed showed that a region of the left MTG was recruited to a greater extent when individuals listened to melodies learned with motoric hand signs as compared to those they learned with the aid of visual representations. This result can be taken as evidence that the hand signs were stored as part of the representation of the melodies: Previous literature suggests that the MTG stores semantic representations – information about stimuli in our environment that we acquire through learning (e.g., Beauchamp & Martin, 2007; Binder Desai, Graves, & Conant, 2009; Blumenfeld et al., 2006; Martin & Chao, 2001). According to the sensory-motor model of semantic representations (Martin, 1998; Martin et al., 2000), the MTG responds to the aspects of stimuli that are movement or motion related, thereby constructing part of our semantic knowledge for the stimuli. As an example, previous literature shows that MTG activation occurs when individuals see tools they have actively manipulated in use, the movement trajectories of these tools, and even the sounds the tools produce (e.g., Beauchamp & Martin 2007; Chao & Martin, 2000; Martin, Wiggs, Ungerleider, & Haxby, 1996; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008). This suggests that the MTG was activated in conjunction with the presentation of melodies learned with a visuo-motor component in the present study because the auditory stimuli of specific notes sung on specific syllables elicited a representation of the action that had been learned with them.

4. No effect of different learning conditions on activation of visual cortex. Unlike motor and auditory learning, learning visual images that corresponded to the melodies did not result in significant activation in the visual cortex upon subsequent presentation of the learned melodies. This finding suggests that either the visual images were not stored, or that the images were not associated with the auditory information strongly enough to be reactivated. Previous literature suggests that when music notation is read, musicians show excitation of the vocal cords as they covertly perform the music (Brodsky, Kessler, Rubinstein, Ginsborg, & Henik, 2008), but to our knowledge, there are no studies that show that musicians have activation of the visual system when hearing music they learned with the aid of music notation. This suggests that, indeed, visual images of a representation of music are either not stored with the auditory information, or are not associated to the extent of being reactivated by listening to melodies learned with a notational system of any sort. Additionally, if integration is occurring, it may be that alterations in functionality occurred in a multisensory area, instead of within unisensory visual and auditory areas that were focused on in our ROI analysis. For instance, one major area of audio-visual integration is the superior temporal sulcus.
which we did not include in our analyses.

5. No effect of different learning conditions on melody recognition. As previously discussed, it is clear that participants were able to learn the melodies: recognition of melodies learned in each of the conditions was significantly above chance during the fMRI session, thereby indicating that training was successful. Whereas the motor system and MTG activation suggest that the hand signs are being processed and stored in a similar manner to more causal movements, this does not seem to be affecting behavior. This is puzzling, as literature on multisensory and sensori-motor learning suggests that incorporating a modality in a learning paradigm will improve an individual’s ability to learn, even if this modality is not utilized during recall (e.g. Seitz, Kim, & Shams, 2006). It may be, however, that the lack of difference in the percent of recognized melodies could be due to a low sample size or a ceiling effect, as all recognition percentages were near perfect, or that differences may have arisen if we have conducted testing at a different time. This finding will be discussed more fully in our General Discussion.

The ability to process abstract motor movements in a meaningful way may be driven by previous experience using the motor system. In the embodied framework, what we do across the lifespan influences our cognitive abilities, thus, the way an adult processes abstract hand signs may be quite different from the way a child would. In Experiment 2, we address this possibility. Using the same basic experimental design with child participants, we are able to explore how children learn melodies with an abstract motor component.

EXPERIMENT 2

In Experiment 1, we investigated how adults process melodies that they learned with or without a motor component. We found that adults showed reactivation in the motor cortex, as well as MTG activation, which is indicative that the hand signs were not only stored in a common representation with the melodies, but were also stored as a communicative and meaningful part of the representations. In Experiment 2, we were interested in whether we would find similar results with children, or whether there are developmental differences in how children and adults process abstract motor movements paired with melodies. Children have less experience associating abstract motor movement to sensory input. Thus, the question of the amount of everyday experience making sensori-motor associations that is required for motor system recruitment during sensory processing was addressed here. From previous literature, we can hypothesize that children will show motor cortex reactivation as a result of within-experiment experience associating sensory and motor information (e.g., James, 2010; James & Swain,
2011). It is unclear, however, whether or not children will recruit the MTG during melody perception after learning, as this may be dependent on whether they can confer meaning on abstract hand signs. If children are unable to recall learned melodies and do not show MTG activation, this may provide evidence that the MTG is important for the retention of the melody learning. Additionally, we are interested to determine whether children will show a behavioral benefit from learning with a motor component, as literature would suggest they might, which would differ from our finding with adults.

Method

Participants

Eleven children, 1 male and 10 females, between the ages of 4 years and 4 months, and 7 years and 10 months ($M = 6$ years and 1 month; $SD = 1$ year and 1 month) participated in the study. All participants were right-handed, native English speakers with no reported history of neurological or psychiatric disorders, and normal or corrected-to-normal visual acuity. Of the children, only 1 of the 11 could read music and had studied music for over a year. Ten of the 11 children had some experience with music in a classroom setting. Informed consent was obtained from the parent or guardian of 27 children, according to the guidelines of the Indiana University Internal Review Board. Data from 16 participants were excluded from the analysis, based on a failure to successfully complete a required training session or because of excessive movement during the imaging session.

Stimuli

Melodies and hand signs. Nine melodies were composed for the study and 6 of the 9 melodies were taught to participants during a training session. All nine melodies were recorded for use in an fMRI session. Melodies were similar to those composed for Experiment 1, but consisted of 2 sections (i.e., measures), instead of 4. Each melody was approximately five seconds (see Figure 4a for example melody). Five of the 8 hand signs from Experiment 1 were utilized (see Figure 4c). The shorter melodies were used because the longer melodies were assumed to be too difficult for the children to learn.
Figure 4.
(a) An example of a melody learned by child participants (b) An example of a visual representation of a melody learned by children (note: original visual stimuli were presented in color) (c) Hand signs taught to children during the motor learning condition.

Procedure

Training session. Children learned six melodies during the training session in 2 of the 3 conditions used in Experiment 1: the visual condition in which children learned sung melodies while looking at a visual representation of what they were singing (see Figure 4b for example), and a visuo-motor condition, in which the discrete pitches of the melodies were paired with discrete hand signs (adapted Curwen hand signs), and children were required to repeat both the melodies and the hand signs. In Experiment 1, adults sang melodies on modified solfège syllables in all conditions. Because learning the nonsense syllables was deemed too difficult for children to learn (and depended on reading ability, which can be highly variable for a group of this age), each discrete pitch was paired with the name of a color for Experiment 2. The visual representation used was colored circles in the order they were sung in each melody (see Figure 4b). As in Experiment 1, melodies were
taught using a recall procedure adapted from Racette and Peretz (2007). Training was complete when children had sung each melody accurately without help from the experimenter; thus, melodies were learned to the same criterion by children as by adults in Experiment 1. To assess learning, children listened to nine melodies after the training session, 6 learned and 3 novel, and told the experimenter whether they had learned the melodies or whether the melodies were new.

After training, children were exposed to an fMRI-like environment through experience in a simulator: a model of the fMRI machine in which children can hear the sounds that will be made by the scanner and experience being inside the bore of the MRI machine. This simulation was performed because the fMRI environment is novel, and can be frightening to children. By gradually exposing children to an fMRI-like environment through experience in a simulator, children are more likely to be comfortable during the experimental session (for more on this methodology see James, 2010; James & Maouene, 2009; James & Swain, 2011).

**fMRI session.** After children were comfortable in the simulator they underwent a scanning session, which consisted of three functional imaging runs and a high-resolution structural scan, similar to those in Experiment 1. Functional localization: Children underwent a 4-minute block-design run to localize functional regions of interest (ROIs) in the brain, including areas involved in auditory, motor, and visual processing. Auditory and motor localizers were the same as those used in Experiment 1. For the visual localizer, children were shown blocks of colored circles. Children were asked to remain still during the visual and auditory tasks, but performed hand movements during the motor task. Experimental sessions: Children underwent two 3 min 30 sec event-related runs, during which they listened to melodies that they had either learned in the training session, or novel (unlearned) melodies, composed using the same guidelines as the learned melodies. Children were not presented with visual stimuli, or asked to produce learned motor movements during these runs; they simply listened to melodies they had learned in the training session, and the novel, unlearned melodies. Children indicated whether they remembered learning each melody via button-presses with their index and middle fingers on a left hand paddle. The final scan was a high-resolution anatomical scan, onto which the functional data were overlaid for the purpose of analysis.

**fMRI data analysis procedures.** As in Experiment 1, we determined ROIs from our localizer. We conducted contrasts of motor versus rest, visual versus rest, and auditory versus rest in all individuals from the localizer scans. Based on these contrasts, three ROIs were identified: one motor region in the left somatosensory cortex (peak average Talairach coordinates x(-33), y(-25), z(55)), and one region involved in auditory processing in each temporal lobe (peak average Talairach coordinates x(-49), y(-22), z(9) and x(51), y(-22), z(12)). The visual localizer was
not usable because the fixation image used to hold the attention of the children turned out to be just as visually complex as the stimuli used for the visual localizer, so visual ROI analysis was based on the adult ROIs from Experiment 1 (Talairach coordinates $x(-28)$, $y(-81)$, $z(-1)$ and $x(33)$, $y(-84)$, $z(-2)$). T-tests were conducted on the data from each ROI, comparing neural activation to melodies learned in the motor and visual conditions. Imaging parameters for the fMRI session were the same as those used in Experiment 1.

**Results**

**Behavioral Data**

Children varied in their ability to follow the instructions during the fMRI session, and therefore we did not analyze responses to children’s differentiation between learned and novel melodies while in the scanner. That is, it appeared that children confused the button responses: sometimes not responding at all for a period of time, sometimes using incorrect fingers, and sometimes pressing two buttons at once. In the past, we have used passive viewing/hearing during fMRI with children this young, as responses tend to be too difficult (e.g., James, 2010; James & Maouene, 2009; James & Swain, 2011), here we attempted to collect behavioral data, but as before, to no avail. Therefore, to assess learning, we used the answers given at the end of the training session, when children listened to nine melodies and identified melodies they learned, to determine whether children could correctly differentiate between learned and novel melodies. Based on a $d$ prime analysis, correcting for response bias, on average, children correctly identified 50.6% ($SD = 0.43$) of the melodies from the motor training condition, and 41.2% ($SD = 0.38$) from the visual training condition. The percentage of correct identifications made by participants was not above chance either for motor melodies, $t(10) = 0.05$, $ns$, or for the melodies learned in the visual condition, $t(10) = -0.76$, $ns$. This indicates that although children learned the melody to the same level of performance as adults, they either could not recognize the melodies they learned, or did not understand the question they were being asked. Alternatively, the time frame of testing affected their accuracy: many children were preoccupied, knowing they were about to undergo an fMRI session. Despite their poor performance subsequent to the fMRI session, the imaging results suggest that indeed, they did differentiate the melodies (see below).

**fMRI ROI Analyses**

**Auditory ROIs.** A paired-samples $t$-test was conducted to determine whether a significant difference in BOLD activation existed between melodies learned in the motor and visual conditions within the bilateral auditory ROI. No significant effect was found either in the left, $t(10) = 0.83$, $p = 0.43$, or the right $t(10) = 0.46$, $p = 0.65$.
ROI. This result is similar to that of the adults in Experiment 1, as the significant difference found in the auditory ROIs in the adults was due to a greater neural activation to melodies learned in the auditory condition (which was not included as a condition in Experiment 2), whereas no differences were found between activation to presentations of visuo-motor or visual melodies in the adults.

*Visual ROIs.* Paired-samples *t*-tests were run to determine if there were significant differences in BOLD activation in the bilateral visual ROI to melodies learned in the visuo-motor and visual conditions. No significant differences were found in either the left, \( t(10) = 0.08, p = 0.93 \), or right \( t(10) = 0.91, p = 0.38 \) ROIs. These findings are similar to those in Experiment 1, in which no significant differences were found in activation patterns in the visual ROIs of adult participants.

*Motor ROI.* Finally, a paired-samples *t*-test was conducted to determine whether significant differences in activation to melodies learned in the visual and visuo-motor conditions occurred in the motor ROI. A difference was found in the motor ROI, in which greater neural activation occurred when children heard melodies learned with a motor component, than without a motor component, \( t(10) = 2.16, p = 0.05 \) (see Figure 5). This finding was particularly interesting, as the auditory and visual ROI analyses revealed the same patterns as were found in Experiment 1, but in the present experiment, children showed increased activation to melodies learned with a motor component in the Motor ROI, whereas no difference was found in the Motor ROI for adults.

*Figure 5.* The bar graph displays the average of the peak BOLD activation for each individual to melodies learned in the motor and visual conditions, taken from the motor ROI (peak average Talairach coordinate \( x(-33), y(-25), z(55) \)). The specific talairach coordinate from with the peak BOLD activation to the melody types was extracted varied among individuals. Individual talairach coordinates were determined by performing a contrast between the auditory localizer and rest, and using the peak activation to extract data from the combined runs in which participants listened to melodies from the learning conditions.
fMRI Direct Contrasts

As with the adult participants in Experiment 1, a whole-brain contrast was performed contrasting BOLD activation to melodies learned in the visuo-motor condition, to those learned in the visual condition. This contrast revealed an area in the right postcentral gyrus (PoCG) (Talairach coordinates x(47), y(-27), z(40)), that showed significantly greater activation to melodies learned with a motor component than to those learned without a motor component (see Figure 6). There were no regions that showed more activation when children were presented with visual melodies, as compared to visuo-motor melodies.

Figure 6.
Whole-brain contrasts between melodies learned in the motor condition and melodies learned in the visual condition. When participants were auditorally presented with melodies learned in the motor condition, a region of the rPoCG showed greater activation than when participants were presented with melodies learned in the visual condition (Talairach coordinates x(47), y(-27), z(40)).
DISCUSSION

Our direct contrast and ROI analysis indicated that learning melodies with motoric hand signs activated parts of the cortical motor system upon subsequent presentation of the learned melodies. Our ROI analysis also revealed that activation patterns in the visual and auditory cortices were not affected by learning condition differences, as was found with adult participants in Experiment 1. Interestingly, unlike our adult participants, children were unable to recognize melodies learned with either a motor or visual component, despite differences in activation patterns to these melodies. Neural activation patterns are more sensitive than behavioral measures: if the children did not code the different learning sessions, then the neural activation would be the same for all conditions—but it was not. Therefore, although behavioral performance in recognition was below chance, the brain was coding and registering the different learning sessions. These findings are discussed in turn below.

1. Reactivation of motor cortex to motor melodies. Results from the fMRI session revealed that upon subsequent presentation of melodies learned with a motor component, children show reactivation of the left PrCG (based on the ROI analysis) and the right PoCG (based on the direct contrasts). These results complement previous research suggesting that learning with action will result in motor system recruitment upon later presentations of learned stimuli (e.g., Hauk et al., 2004; James, 2010; James & Atwood, 2009; James & Maouene, 2009; James & Swain, 2011; Lahav et al., 2007; Nyberg et al., 2001), but just as with the adults, our results show something new: The motor components used in previous research are intuitively connected with the stimuli being presented, or in some cases, cause the stimuli (i.e., depressing a key on a piano is a motor movement that causes a note to sound). Our result is novel—our motor component was more abstract than motor components used previously, in that it was not intuitively connected with the melodies that were learned, and it did not cause the melodies. Additionally, we showed activation not only in the left PrCG, an area implicated in the performance of action, but also in the right PoCG, an area implicated in body awareness. This indicates that although hand signs were performed with the right hand, leading to reactivation in the left hemisphere, crosstalk between the hemispheres also occurred, leading to activation in the contralateral PoCG upon subsequent presentations of motor melodies. This finding complements previous findings that children show a change from more inclusive, non-differentiated processing within the brain when first performing a task, to more differentiated, or compartmentalized processing as adults (e.g., Johnson, 2000; Smith & Gasser, 2005). For example, when learning to read, children initially engage a much more diffuse network of brain areas than that of mature readers (Turkeltaub, Gareau, Flowers, Zeffiro, & Eden, 2003), similarly, children initially process words (e.g., Mills, Coffey-Corina,
Similar to the motor cortex reactivation to melodies learned with a motor component by adults in Experiment 1, one suggestion of the reason reactivation occurred is the common coding hypothesis, proposed by Prinz (1997). Based on this hypothesis, children in the present study coded the hand signs and sung melodies together, leading to the reactivation of the motor movements when the melodies were presented.

2. No effect of different learning conditions on activation of visual or auditory cortices. Whereas differences were seen in activation of motor areas to melodies learned with a motor component as compared to melodies learned with a visual component, no significant differences in neural activation were found in visual and auditory areas. This finding is similar to that of Experiment 1, in which no differences were found in activation patterns to different melody types in the visual cortex, and the differences found in activation patterns in the auditory cortex were due to an increased activation to melodies learned in the auditory alone condition; a condition not used in Experiment 2. Although adult and children were taught with different visual components (syllables vs. colors), it can be argued that data obtained from the children could be compared to the adults as neither syllables nor colors are traditionally associated with notes or hand signs. As in our discussion of the results from adults, the lack of difference in activation either suggests that visual images were not stored with the melodies, or that they were not associated strongly enough to be reactivated during subsequent presentations of melodies learned in the visual condition. An additional interpretation is that representations were stored in audio-visual integration areas such as the STS (e.g., Stevenson et al., 2007; Stevenson & James, 2009), but that this area was not included in our ROI analysis. The lack of difference in the auditory cortex indicates that neither motor, nor visual components can strengthen an auditory representation of a melody. Thus, the use of a motor component during learning with children only seems to affect subsequent melody processing in motor areas.

3. Recall at chance for both visual and motor melodies. Results indicated that children were unable to recognize melodies after learning them with a motor or visual component. This finding is unanticipated: like adults, we would have expected the motor component to be beneficial to learning based on multisensory and sensori-motor learning research (e.g., Seitz et al., 2006), and gesture literature (e.g., Cook et al. 2008; Goldin-Meadow et al., 2009). This result will be discussed in relation to the behavioral findings of Experiment 1, the difference in procedure for Experiments 1 and 2, and the neuroimaging results from both experiments.
GENERAL DISCUSSION

In the present study, we investigated how adults (Experiment 1) and children (Experiment 2) process melodies learned with (visuo-motor condition) or without (auditory and visual conditions) a motor component. Our results showed that just as actions that are directly influencing something being perceived (i.e., using a tool, playing the violin, writing letters), lead to reactivation of the motor system upon subsequent perceptual encounters, actions that are only abstractly connected to a perceptual experience (i.e., hand signs learned with pitch-syllable combinations), also lead to subsequent motor system recruitment: Both adults and children showed reactivation of the motor cortex upon subsequent auditory presentation of melodies learned with a visuo-motor component. We also revealed that there were no changes in processing in the visual cortices, based on different types of training, and that there were no differences in the auditory cortices when melodies learned with a visuo-motor or visual component were processed (although we did show that adults, who received training in an auditory only condition, showed greater activation in the auditory cortex bilaterally to melodies learned in this auditory alone condition).

The similar results from Experiments 1 and 2 reveal that some of the effects of learning melodies with a visuo-motor component are consistent across development. We found, however, three differences between children and adults, indicating that the way we learn with an associable, but non-causal motor component, such as Curwen hand signs, changes with development. First, although the motor cortex was reactivated in both age groups when participants were presented with visuo-motor melodies, different areas were reactivated in adults and children. Specifically, children showed more regions of reactivation than adults. Second, adults showed activation of the MTG when presented with melodies learned with a motor component, whereas children did not show MTG activation. Third, whereas adults recognized melodies learned in all conditions equally well, children were unable to recognize the melodies they had been able to produce correctly from memory after the training session, regardless of melody learning condition.

Whereas adults and children both showed left premotor activation when listening to melodies learned with a visuo-motor component, children also showed right somatosensory activation. As previously discussed, this finding complements research demonstrating that across development and with acquisition of experience, individuals shift from a diffuse processing network to a more specialized network (e.g., Johnson, 2000; Smith & Gasser, 2005). What is interesting about the results of the present study is that part of this diffuse motor network in children is the somatosensory cortex, which is involved in body placement, not motor planning. Diffuse networks are often recruited because children need extra areas when learning to process new stimuli – for example, the recruitment of right extrastriate cortex during reading decreases as reading skill increases because as children gain
experience with letters, they are less reliant on a non-lexical form recognition system (Tukeltau et al., 2003). The present study may indicate that children initially rely on their body placement in the world for learning to a much greater extent than adults. Perhaps for children who learn more experientially than adults, a sense of body placement is an important part of the representation of movement, and that children learn in a more ‘embodied’ fashion than adults who have built up other ways of representing the world. The somatosensory activation in children may also be related to the more important difference: their lack of MTG activation.

The most critical difference in our neuroimaging data is the finding that adults showed activation of the MTG when listening to melodies learned with a motor component, but children did not. The MTG has been implicated as an area that stores the motoric aspect of object or concept representations, and is reactivated when an individual perceives a stimulus with which he or she has associated a motor movement (e.g., Chao & Martin, 2000; Beauchamp & Martin, 2007; Martin et al., 1996). The MTG also shows activation when individuals process gesture (e.g., Kircher et al., 2009; Villarreal et al., 2008; Wakefield & James, in preparation). The finding that adults show this activation for non-causal motor hand signs suggests that they are able to build motoric representations related to something perceived, in this case, a melody, even when the motor movement is only abstractly related to the melody. It seems that children, however, are unable to build this type of connection, even though research indicates that this area is activated when children perceive gesture (Wakefield & James, in preparation). This may mean that it is only with enough experience connecting action and perception that abstract motor movements can become meaningful representations like concretely related actions or gestures. If this is the case, it may be that only meaningful movements can facilitate learning, and this is why adults could recognize melodies and children could not. A second possibility is that children simply lacked the motivation to retain knowledge about the hand signs. It may be that the act of producing the movements was enough to cause the motor cortex to store a motor representation connected to the melodies, but that the MTG would only be activated if children encoded the melodies and hand signs more fully. Related to the MTG activation, it may be that the somatosensory activation occurred in children but not adults because children lacked the ability to connect the movements to the melodies in a deeper way, via representation storage in the MTG. Together, these results demonstrate a developmental difference in how children and adults process action in the context of perception.

It could be argued that because there were no behavioral differences in how well adults recognized melodies learned in the visual, auditory, and visuo-motor conditions, there is no evidence that the MTG activation is facilitative. We argue that low sample size or a ceiling effect (recognition in all conditions was near perfect), led to the lack of difference in the percent of recognized melodies. A related finding of interest was that adults correctly identified learned melodies

whereas children did not. This finding may be due to the difference in experimental procedure: adults completed the recognition task the day following learning, whereas children completed the entire experiment in one session. Research has demonstrated that sleep may be important for auditory learning (Gottselig et al., 2004). After learning a set of melodies, in a study conducted by Gottselig and colleagues, participants slept, relaxed in a dark room, or were asked to watch a movie and pay attention to it. This manipulation was followed by a second test of melody learning. Participants who slept or relaxed performed better on the second learning task than those who watched a movie. The authors concluded that sleep may be a key component for solidifying auditory learning, although they did not conclusively show that sleep was better than simply relaxing without added sensory input. In our study, then, it could be that children would have recognized the melodies if they had been tested the day following. This highlights an interesting finding: both children and adults associated motor learning with the melodies, as evident from the neuroimaging data, whereas it was only adults who recognized melodies. This may indicate that motor learning happens immediately, but for auditory learning to be solidified, sleep, or a restful consolidation period, is necessary.

In practice, Curwen hand signs are taught and used over an extended period of time in the context of music education. With time, it is likely that these hand signs become communicative for the individuals who use them, and may be a useful tool in learning. Results from the present study may indicate that it is not only the presence of a visuo-motor component that will influence its interaction or contribution to an individual’s ability to learn, but also the communicative power of the action. To further investigate this idea, it would be beneficial to conduct a study in which participants trained with motoric hand signs over a longer period of time, during which the hand signs would theoretically become analogous to gestures, or to conduct a study with individuals who are already familiar with the Curwen hand sign system, to the point of understanding them in a way that was meaningful and communicative.

In summary, developmental change occurs during melody processing in that children do not encode motor signs as part of semantic representations, but both children and adults form associations among sensori-motor systems, and children actually form a wider network of motor associations. This association leads to motor system recruitment during purely perceptual tasks only when movement is incorporated into the learning episode. Later, in adulthood, sensori-motor associations are made, and motor movement is coded as meaningful to the learning episode. Therefore, as we mature, motor movements that are not causal, but instead, are simply associated with a learned event become more meaningfully connected to the event.
REFERENCES


A NEW PERSPECTIVE ON EMBODIED SOCIAL ATTENTION

Hanako YOSHIDA, Joseph M. BURLING*

Department of Psychology, University of Huston, Texas, USA

ABSTRACT

Over the years observational studies have made great progress in characterizing children’s visual experiences and their sensitivity to social cues and their role in language development. Recent technological advancements have allowed researchers to study these issues from the child’s perspective, leading to a new understanding of the dynamic involvement of bodily events. A number of recent studies have suggested that bodily actions play an important role in perception and that social partners' bodily actions may become synchronized. In the present perspective paper, we will provide a new perspective on how children's own views are generated individually and play a dynamic role in learning. By doing so, we first discuss the role of early social input in language learning as it has been treated in the literature and then introduce recent studies in which typically developing hearing children, deaf children of deaf families, and children with autism were observed in a social context using the new child-centered technology. The hypothesis of a link between sensorimotor experiences and embodied attention — specifically how different bodies produce different kinds of attention — will be discussed. Understanding the role of bodily events (the child’s and the child’s social partners’) in early visual experiences will provide insight into the development of learning mechanisms and the processes involved in learning disabilities.

KEYWORDS: embodied attention, social interaction, child-centered perspective

* Corresponding author:
E-mail: yoshida@uh.edu
Children learn about their world — about objects, actions, other social beings, and language — through social interactions. Social interactions are coordinated by a suite of bodily actions including gestures, head and eye movements, body sway, and posture (Langton, 2000; Shockley, Baker, Richardson, & Fowler, 2007; Shockley, Richardson, & Dale, 2009). A number of researchers consider the body to play a significant role in our experience of the world (Barsalou, Niedenthal, Barbey, & Ruppert, 2003; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). Developmental researchers have long studied the role of social input by careful observations of what is happening in the learning environment (e.g., Bakeman & Adamson, 1984; Bates, 1979; Fogel, Lyra, & Valsiner, 1997; Iverson et al., 1994; Iverson, 2010; Toda & Fogel, 1993; Toda, Fogel, & Kawai, 1990). Scientific progress has also been made through laboratory studies of children’s use of social cues, particularly in the domain of language learning. These studies show that children are highly sensitive to a variety of cues and that these social cues appear instrumental to early word learning (Baldwin, 1991; Bloom, 2000; Tomasello, 1995).

All this previous work considered social interactions from the perspective of an observer who coded what each participant was doing: the signals the child sent, the adult’s responses, and the child’s responses to the adult behavior. Social actions were seen as forming a kind of language of messages that are sent and received. But every behavior — every pointing of a finger, every shift in eye gaze — has two parts: its direct effect on the sensory system, and the social message such a signal might send. For example, shifting one’s direction of gaze changes what one sees and sends a signal about the object being attended to. A gesture by one’s social partner serves as a communicative signal. Such gesture use has a direct effect on the sensory system and input. The aim of this perspective paper is to introduce a new perspective about the roles of children's own views in learning in light of the embodiment idea; specifically, we will emphasize the role of new technologies and recent developmental attempts. By doing so, we consider the direct sensory effects of bodily actions in a social context — focusing specifically on how a child’s actions and the parent’s actions are dynamically linked and directly influence visual input and, thus, potentially learning.

**KNOWING WHERE THE EYES ARE IS NOT ENOUGH**

In the long history of developmental science, the most common view used in observations is a “room view”, a third-person view from a corner or from another room via a one-way mirror. This has certainly been sufficient when a controlled environment is needed in which a solitary child is given a single item to look at; however, real learning environments, where young children learn social components and language, are rarely so simple. Indeed, their everyday learning
environments are very chaotic and often composed of many objects and actions. Also, learning environments are dynamic. For example, children and their social partners constantly change the elements of the environment (i.e., objects and actions) and their own locations within it. This makes the environment even more complex, and yet all these elements are relevant for learning, especially in social contexts. Constantly changing coordinated eye gaze, language use, and body posture function to facilitate communication and mutual understanding (Pickering & Garrod, 2004; Shockley et al., 2009). Thus, to be sensitive to these constant changes in visual experiences, taking a child-centered view is essential. The third-person view typically cannot distinguish which parts of a complex scene a child might be attending.

Another issue affecting the choice of perspective is that adult coders are influenced by where they think the child should be looking. If a child is looking at the wall behind his or her social partner when the partner says “oh, no” it is likely that coders will score the look as if the child had looked at the partner, not at the wall. This is due to our strong tendency to use intuition to describe what should be happening. This type of coding issue and intuition bias can be particularly critical in social domains where many social contacts and complex emotional exchanges might be present to reinforce such overinterpretation. It is a particular problem when measuring social development. Infants look back and forth between their social partner’s face (and gestures) and an object. Indeed, looking between a social partner’s gaze and potential referents has been thought to supply information about the partner’s gaze direction and affect (Butterworth, 1995; Tomasello, 1995). Thus, coding requires precision. Observations containing social contexts benefit most from the correction of these intuition biases. Acquiring an unbiased and more precise coding of what the child is looking at can be essential for this type of measure.

Recent technological developments have made it possible to capture input from the viewer’s perspective using eye-tracking equipment. Yet until very recently, with the advent of head-mounted eye trackers (which have not been used much in infant–parent social settings), eye tracking has typically been used with stationary babies watching a TV screen, and thus researchers were unable to capture the role the whole body plays in social behavior or the relation between the child’s own actions and what the child sees (but see Franchak, Kretch, Soska, & Adolph, in press, for a child-centered perspective in a self-navigation context).

Two recent eye-tracking studies that recorded social scenes from infants’ perspectives successfully documented natural tendencies in visual exploration when the videotaped scenes were shown to different groups of young infants and adults (Aslin, 2009; Frank, Vul, & Johnson, 2009). The studies reported substantial rates of looking at faces as well as age-related differences in fixation. Both studies also noted interesting attention shifts from face to hands when the hands were performing actions, indicating that social attention is sensitive to the content of
actions. These studies successfully captured what children look at when scenes are presented in two dimensions. Observational precision (i.e., how precisely one can detect where an infant is looking at any given moment) will continue to improve as we persist in making technological advancements.

However, although these studies made use of eye-tracking technology, they did not address the idea of embodied attention because they studied infants looking at videos produced by equipment on other infants’ bodies. Embodied attention means that what is attended to in the learning environment is partly influenced by body structure. Body structure in this case refers to the child’s own physical constraints and actions, such as eye gaze, head movements, posture shifts, grabbing or holding, and the same actions performed by another body — a social partner. What this means is that our perceptions (moment-to-moment visual experiences) may not be sufficiently understood in isolation from our own body and movements. Children “select” their visual input both by their own actions and by how their actions influence the behaviors of others. This core idea of there being a meaningful link between bodily experiences and perceptual experiences has received much attention (Barsalou, Pecher, Zeelenberg, Simmons, & Hamann, 2005; Bertenthal, Campos, & Barrett; 1984; Martin & Chao, 2001), and recent behavioral evidence has further suggested the importance of sensorimotor coordination (e.g., seeing, touching, manipulating objects) in children and the dynamic linkage to the learning environment for our understanding of visual selection and social input (Adolph & Berger, 2006; Needham, Barrett, & Peterman, 2002; Pereira, Smith, & Yu, 2008; Thelen & Smith, 1994; see Sheya & Smith, 2010, for a review).

New methods have been developed to explore this dynamic loop between social attention from the child’s point of view and the child’s own moment-to-moment interactions. One study explored early visual behavior in a naturalistic free-play context where 14-month-old infants’ eye gaze was recorded using a head-mounted eye-tracking device (Franchak et al., in press). This study demonstrated that infants’ selection of their own visual field depended heavily on their own actions, and they seldom looked at faces (in this case, their mother’s face). Further, the study revealed a new relation between infants looking at their own hands and their own actions while engaged in manual actions, and a direct impact of the social partner’s movement on the child’s view.

**HEAD CAMERAS AND THE HEAD-CENTERED VIEW**

Efforts have been made to document children’s own natural visual exploration while they participate in moment-to-moment social interactions. These studies have used relatively less complex technology yet are still suitable for addressing embodied social attention. Typically, a small head camera (about the size of a quarter) is placed on the child as the child and parent play with a toy. The camera
captures visual events in which the children themselves are participating. The children are creating their own visual input through social exchange, and in contrast to the eye-tracking methods described above, this is captured from the child’s perspective (often a child-mother play session). This perspective is taken from the forehead (as opposed to tracking eye gaze) and thus is a relatively less complicated form of data collection that does not interfere with natural viewing and their interactions.

One major concern about using the head-mounted camera has been that it is positioned at some distance from the eyes, and thus it may be difficult to identify the focal point of the child’s attention with precision. Still, several researchers have found a strong head–eye coupling in infants (e.g., Daniel & Lee, 1990; von Hofsten & Roseander, 1996; Savelbergh, von Hofstein, & Jonsson, 1997; Smith, Thelen, Titzer, & McLin, 1999), and more recently Yoshida and Smith (2008) documented how the head view can secure sufficient measurement of where children attend to in common social contexts. This study assessed the correspondence between the head camera view and the direction of eye gaze in a calibration task with geometry similar to that of tabletop toy play (e.g., horizontal movement, reaching for objects). The study verified the sufficient overlap between the head camera images and the child’s visual field, suggesting the effectiveness of the head camera in the interactive task context, which simulates a natural learning environment. Moreover, in social settings in which people are continually moving their head and eyes, eyes alone are unlikely to be a good cue.

Although studies have suggested early sensitivity to eye gaze and its relevance to language learning (e.g., Brooks & Meltzoff, 2008), little work has been done on how temporal dynamics of eye-gaze following can emerge in complex social contexts where heads, eyes, and bodies are constantly engaging in actions. Indeed, most experiments on the following of eye gaze, in infants and adults, manipulated eye-gaze direction in a fixed, straight-on face (see Langton, Watt, & Bruce, 2000, for a review). Thus these studies have not accounted for the natural social contexts in which heads and eyes move both together and independently (Einhäuser et al., 2007).

Additional support for head–eye coupling comes from studies showing that gaze and head orientation independently influence decisions concerning social attention. Adults, children, and infants demonstrated difficulty ignoring the direction of the head when judging eye-gaze direction (Doherty & Anderson, 1999, 2001; Doherty, Anderson, & Howieson, 2009; Langton, 2000, Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Moore, 2006; Moore & Corkum, 1998). Accordingly, the head-camera method seems to be sufficient for studying children’s perspective while engaging in social actions, which occurs in most natural learning environments. Recent results support the importance of taking a head-camera view when evaluating children’s own creation of visual input.
A group of researchers (Pereira & Smith, 2009; Pereira et al., 2008; Pereira, Yu, Smith, & Shen, 2009; Smith, Yu, & Pereira, 2011; Yoshida & Smith, 2008) explored children’s proximate visual behavior in the context of naturalistic play with the mother playing and teaching a set of words to her child. The typical procedure used with studies of this kind attaches a small head camera to the child’s forehead with an elastic band while the child engages in toy play with a parent. The goal is to capture visual events — in which the child creates his or her own moment-to-moment social exchange — from the child’s perspective. Each study has contributed to the understanding of what components are most likely to be available to the child and how they might relate to social events, and how that might depend on the individual’s learning experiences.

These studies focused on exploring the visual field of toddlers (ranging from 12 to 24 months old) by using a variant of the head-camera method, documenting several novel and enlightening results concerning the dynamics of toddlers’ first-person view. One study (Yoshida & Smith, 2008) noted that the room view and the child view (head-camera view) are very different (see Figure 1 for synchronized view frames from Yoshida & Smith, 2008). Indeed, it is relatively clear from the head-camera view what is available to the child’s attention at any given moment, but not when observed from parents’ views (also via head cameras). Children’s’ unique behavior of bringing objects closer to their eyes can reduce the information available in their view. This is a potentially significant constraint in early learning.

![Figure 1.](image)

Left: third-person view; right: first-person view.

Furthermore, the closer view of an object is also indicative of the relation between a child’s view and the child’s body structure — a preference for closer views is related to a child’s shorter arms and the location of arms relative to the head. Yoshida and Smith’s (2008) and Smith et al.’s (2011) results also suggest a strong tendency to look more at objects that are in hands — their own and their
parents’ — and to shift attention to new objects as parents’ hands move in that direction. These results using a different methodology appear to contrast with earlier findings of a preference for faces (e.g., Aslin, 2009; Frank et al., 2009), highlighting the potential role of parents’ action in organizing early attention. These studies indicate how early attention is dynamically linked to a child’s own body and actions and is embedded in a social partner’s action.

**DIFFERENT BODIES, DIFFERENT SOCIAL ATTENTION?**

Studies that captured the visual content generated by an individual child have emphasized the importance of considering embodied attention to fully understand children’s visual selection and social input. The seemingly tight loop connecting embodied attention, an individual’s own actions in a social environment, and social partners’ actions suggests that such embodied attention might differ when it is generated by children whose sensorimotor experiences and social interaction are different and uniquely integrated.

Our current attempts to understand embodied attention — specifically feedback from sensorimotor, language, and social experiences — have focused on social contexts that mimic word-learning environments that diverge from those of the typically developing, hearing child (Yoshida, 2010; Yoshida & Kushalnagar, 2009). We have concentrated on three groups in an effort to explore the dynamic structure of embodied attention: typically developing hearing children, typically developing deaf children of deaf families, and hearing children with Autism spectrum disorder (ASD). The sensorimotor, language, and social experiences differ among these children both quantitatively and qualitatively.

In one study (Yoshida & Kushalnagar, 2009), typically developing deaf children aged 4 to 5 years were compared with typically developing hearing children of the same age range in an extension of the parent–child interaction sessions used in earlier work by Yoshida and Smith (2008). Parents were instructed to interact naturally with their child with the goal of teaching a series of words by demonstrating the meanings with toys and actions. Head-mounted cameras were attached to the young participants in order to provide a child-centered perspective of the learning environment. The head-mounted camera (Figure 2) is a Watec (WAT-230A) miniature color camera weighing approximately 30g (36×30×15mm). The camera was sewn into a headband that could be placed on the forehead such that the mounted camera is close to the child’s eyes. The angle of the camera is adjustable. A second camera — the third-person camera — was set up to record a broad view of the task context that included the tabletop, the child, and the parent. The two cameras were synchronized by an Edriol (model V-4) four-channel video mixer. The head camera moves with the head (not with eye movements). However, a prior calibration study directly comparing independent measures of head and eye
direction in typically developing toddlers found a 90% correspondence in direction. After fitting the child with the headband and giving instructions, the experimenters left the testing section and hid behind a curtain to monitor the head-camera view to continue ensuring the camera angle. The entire play session lasted about 12 minutes: parents taught 16 words, each word was taught for 40 seconds.

*Figure 2.*
Head camera

A total of 6 coders coded the video from the head camera (2 deaf coders for American Sign Language transcribing) and 4 coders for nonlinguistic contents frame-by-frame using MacShapa, an application for observational data analysis. They calculated the amount of looking time for coding categories including the individual toys, the parent’s face, the parent’s hands, and the child’s hands to obtain the proportion of looking per session.

This approach — namely, exploring the social interactions among attention, body, and the environment — is used here to consider how different populations of children (e.g., deaf, hearing) partition their visual scenes and to observe the dynamic processes that shape their word-learning experiences. Parent–child interactions are inevitably influenced by the capabilities of both the child and the parent. For instance, the use of either sign language or verbal speech as a means of communication has a direct impact on what is observed from a child’s point of view. Which simultaneously appearing objects are attended to seems to differ between hearing and deaf children, and the proportion of time spent attending to certain objects over others varies, as well. This could be related to differences in bodily communication styles. Communication through speech among hearing individuals frees up the hands of both parent and child so that objects within their immediate environment can be adapted for use in teaching new words. Dyadic interactions involving parents who use sign language to communicate with their children involve more prolonged periods of attention to crucial communicative components, such as the parent’s hands or face, and the default location of hands might be higher and closer to the face (Yoshida, 2010; see Figure 3 for synchronized view frames from Yoshida & Kushalnagar, 2009).
The structure of deaf children’s embodied communicative acts also constrains their views in that their language (hands) dominates their view, and their sensory experiences (mostly through the visual modality) assign more attentional weight to visual components. Because the deaf children in this study were all being raised by parents who are also deaf, their initial encounter with language perception was likely to be through visual experiences, and their language production is also through the visual system. Their visually organized language use relies on meanings and ideas, and the meanings children learn early are often considered to be concrete and visually tangible at an early stage in general (Gentner, 1982). Accordingly, for deaf children who have access to sign language in the beginning of life, language perception, production, and the mapping of language to meaning are all linked through visual experiences, suggesting early dominance of visual processing. The question is how this visual dominance influences attention to social components, such as a partner and her action. Figure 4 illustrates the overall trend in viewing the parent’s hands or face for deaf children and typically developing hearing children (Yoshida & Kushalnagar, 2009). The graph compares the proportion of time spent looking at the parent’s hand or face, averaged across 16 teaching trials. Deaf children on average captured these items more frequently than hearing children, especially at the beginning of a word-learning session.
Figure 4.
Proportion of time spent looking at the parent’s hands or face averaged across trials.

Allocating attentional resources to hands and faces lessens the possibility of other objects — such as the toys in the word-learning session — being the focus of the child’s attention for a prolonged period of time. Attention to the intricacies of hand movements and facial expressions is less likely in hearing children. Hearing children are also less likely to continually switch from hands and faces to other objects, leading to a more local perspective with attention focused on individual objects for prolonged periods of time. Deaf children, on the other hand, are more likely to simultaneously keep track of the dynamics inherent in their visual scene, leading to a much more global perspective (Yoshida & Kushalnagar, 2009).

Despite apparent differences in their language-learning environments, typically developing deaf children often manage to achieve typical language development, and their adaptability may lie in how they carved up their visual environment. This is evident in the results suggesting that the necessary referents for word learning exist for both deaf and hearing populations. However, it is not our intention to proclaim that the developmental trajectories of deaf and hearing individuals are in any way identical. Other factors, such as the timing of the onset of hearing loss, the implantation of electronic cochlear devices, the hearing capabilities of parents with deaf children, or of hearing children with deaf parents, have their own complexities that can give rise to different word-learning
environments. Nevertheless, the unique structure of embodied attention generated by typically developing deaf children influenced by their unique bodily experiences (via an alternative language-learning environment including their ASL use) seems to generate relevant visual experiences. This suggests that selectively attending to relevant aspects within their visual environment leads to outcomes similar to those of children without hearing loss, but with different means of reaching those outcomes.

**The Case of Autism Development**

The dynamic coordination of hand movements, eye gaze, and posture that can be present in individuals is a crucial component of communication (Chartrand & van Baaren, 2009; Shockley et al., 2009). Such coordination seems to differ depending on individuals’ bodily interactions with the environment and their sensorimotor experiences. Yoshida (2010) addressed the role of bodily structured attention by exploring embodied attention in children with autism, which represents one case of atypical development. This is a particularly interesting developmental question. Many cognitive developmental disorders involve attention, and there is considerable evidence of comorbidity of these cognitive disorders with early sensorimotor patterns (Hartman, Houwen, Scherder, & Visscher, 2010; Larroque et al., 2010). For instance, a number of studies have suggested that the function of shared attention (e.g., child and caretaker jointly attending to the same item) develops differently in children with Autism spectrum disorder. Shared attention has often been discussed in terms of its effectiveness in guiding attention in language learning. The development of shared attention thus can be a key element in language development.

In the dynamic social embodiment framework, shared attention can be viewed as an essential developmental process in which children initiate a demonstration of their attention (e.g., clear, sustained attention to a toy), which parents can then follow — parents can adjust their topic to their children’s moment-to-moment attentional shifts (Adamson & Bakeman, 1991). However, sustaining attention to objects has been considered particularly challenging for young children with autism (Pierce, Muller, Ambrose, Allen, & Courchesne, 2001). From the perspective of social embodiment, the body and what the body does both influence sensory inputs and the social setting. Understanding these links in children with atypical development will lead to recognition of potential developmental consequences.

Yoshida (2010) used a head camera set-up to address how a deficiency in sustained attention influences embodied attention in 4- to 5-year-old children with ASD. Preliminary findings suggest distinct differences between ASD and typically developing children in their focus of attention. The child’s eye view revealed that
objects used to teach novel and familiar words, when paired with hand gestures and manipulations, were present in the field of view of children with ASD for very little time. Irrelevant objects within their immediate environment occupied a substantial proportion of their visual field. Extended periods of time spent observing nonreferential items such as the floor, ceiling, or a table meant that there was less opportunity during the word-learning session for relevant objects to be introduced into the child’s visual field.

In addition to these distinct patterns of looking, observations taken from the head camera revealed a tendency to extend the gaze to the parent’s face only, a significant departure compared to typically developing children. This learned behavior may be related to common practice and a form of communicative therapy provided to those with autism to support their development of social interactions. Such learned tendencies to observe the parent’s face may lead to overall looking patterns that differ from those of typically developing children. This propensity for not attending to some types of relevant information during the word-learning process could provide insight into the language development of children with ASD.

**DISCUSSION**

Children’s early learning environments contain a number of elements that are dynamically linked in real time as they and their social partners move and act on these elements. Researchers attempting to carefully document the focus of children’s attention in complex social settings have suggested that the patterns of children’s looking at faces and their social context may be mutually influential, especially when manual actions are involved. However, in recent approaches in social cognition, it has been suggested that it is not possible to fully understand social interactions without addressing how all the bodily interactions (gestures, head and eye movements, body sway, and posture) are coordinated (Langton et al., 2000; Langton, 2000; Shockley, Richardson, & Dale, 2009; Shockley et al., 2007). This promotes the idea that a child’s perspective is part of a social dynamic loop.
We have discussed ideas and empirical work suggesting how different bodily actions (both the child’s and the parent’s) are dynamically linked and directly influence the moment-to-moment generation of visual input and thus potentially matter for learning. This is far more dynamic and complex than it may seem at first glance, given that different bodies generate different embodied attention based on different body structures, experience different sensory inputs, and perceive different social signals that guide attention. The schema is illustrated in Figure 5. In the figure, a child is attending to an aspect of a scene — in this case, a cup. What is being attended to determines the in-the-moment input (a cup). The hands and body structure of the child play a role in guiding the child’s attention, but this tight loop between visual experience and action is dynamically coupled with a mature social partner. This means that the embodied attention is generated through at least two perception–motor loops (here, the child’s and the parent’s) and that each action changes the sensory input of the child and the parent. What the parent does affects the child’s moment-to-moment visual experiences. Thus, it is not sufficient to study social interactions solely from the perspective of an observer, nor to consider a child’s perspective that does not take into consideration his or her own body and actions.
The role of bodily cues in social attention (e.g., hand movement, hands acting on objects, eye gaze) appears to be relatively clear, yet recent research findings suggest that such a role may well be subtle and possibly implicit in nature. For instance, in whole-body cues, and coordination itself such as in global body rhythms generate synchronous effect (Shockley, Santana, & Fowler, 2003; Smith et al., 2011; Spivey, Richardson, & Dale, 2009). One line of studies documented a relation between communication effectiveness and body coordination between social partners through body sway and posture (Langton, 2000; Shockley, Richardson & Dale, 2009; Shockley et al., 2007). Another study suggested a dynamic link between interaction quality — including smoothness of turn-taking behaviors and bodily coupling (e.g., of speakers) — and effective word learning in the context of parent and toddler social events (Shockley et al., 2007). Accordingly, particular body movements that are obvious and meaningful may not independently contribute to the creation of social attention. Rather, synchrony of movements and more global body structures as a whole are more likely to participate in the loop. There might be dynamic relational local bodily elements and larger body cues creating attention and possibly influencing learning (Bangerter, 2004; Kita, 2003; Richardson, Dale, & Kirkham, 2007). This sensitivity to synchronous elements (invariants) is consistent with the early emergence of multimodal correspondences in learning (Gogate, Bahrick, & Watson, 2000) and thus may well be the key to our understanding of human learning and communication.

Traditional views of early social input have been refined by technological advancement, and certainly the field has benefited from our ability to observe from their own point of view what children scan in complex visual scenes. The new technology has also helped eliminate coding biases. Use of the technology has led to new findings about early sensitivity to action and the context-dependent nature of attention, and it also has confirmed previous observations about early attentional preferences for human faces and eye gaze for particular task contexts. This perspective paper focused on the embodied nature of social attention, i.e., how it emerges through interaction between a child and the child’s social partner, each dealing with his or her own physical constraints, sensorimotor input, and cognitive capacity. Within this framework, we can expect to see different types of social attention for different groups of individuals whose everyday experiences — through sensorimotor input and cognitive development — are uniquely constrained. Systematic observations of children’s proximate visual experiences in social contexts have provided insight into the dynamic relations (Yoshida, 2010; Yoshida & Kushalnagar, 2009; Yoshida & Smith, 2008). Studies of deaf children of deaf parents (Yoshida & Kushalnagar, 2009) revealed sustained attention to face and hands — which has a communicative function — and more attentional shifts to capture a global perspective compared to typically developing hearing children. Observations of hearing children with ASD (Yoshida, 2010) revealed that they pay less attention to objects that are the topic of conversation (e.g., held in hands and
acted on) and more attention to nonreferential items. These studies suggest that attentional biases are generated through the dynamic social loops of all participants: perception, action, and body structure. Differences in the dynamics can generate differences in attentional biases and thus in language development.

Taking the child’s perspective of the perception–action–body loop is a crucial step in understanding social structure and early visual experiences and their influence on cognition, from language to problem solving. This may lead to further insight into the learning mechanisms and processes involved in learning disabilities.

ACKNOWLEDGMENTS

The preparation of the manuscript and research introduced in the manuscript were supported by the University of Houston, a National Institutes of Health grant (R01 HD058620), and the Foundation for Child Development.

REFERENCES


*Cognition, Brain, Behavior. An Interdisciplinary Journal*  
15 (2011) 535-552
H. Yoshida, J. M. Burling 551


ABSTRACT THINKING IN SPACE AND TIME: USING GESTURE TO LEARN MATH

Susan Wagner COOK*
Department of Psychology & Delta Center, University of Iowa, Iowa City, IA, USA

ABSTRACT

Mathematical concepts and words for objects are generally thought of as abstract and disembodied. In this theoretical paper, I review a line of research demonstrating that hand gestures are important for mathematical thinking. Gestures are spontaneously produced in conversations about mathematical concepts, these gestures can influence speakers’ processing, and gestures are implicated in changing children’s and adults’ mathematical thinking. This work offers evidence that mathematical representations may not be as abstract and disembodied as they seem, but rather may be supported by embodied representations. This example suggests that the body, and its affordance for action in the world, is important for conceptual development.

KEYWORDS: gesture, mathematics, learning, embodied cognition, cognitive development

EMBODIED THINKING

Conceptual development is typically described as a progression towards increasingly abstract and disembodied representations. Mathematics provides a salient example. Mathematical concepts are generally thought of as abstract and disembodied, which can be seen in their wide-ranging applicability to a variety of situations, and in their inability to be observed in the real world. For example, a single function can represent relationships from a variety of possible real world situations, which can cause learners to have difficulty appropriately mapping
between real world situations and their symbolic mathematical representations (Coquin-Viennot & Moreau, 2007; Mattarella-Micke & S. L. Beilock, 2010).

I have been exploring the role of hand gestures in the development and communication of mathematical thinking. Hand gestures are produced by speakers of all languages during communication. These gestures generally provide a perspective on cognitive processes that is complementary to what is revealed in spoken language. In one line of work, I have been exploring the question of what hand gestures can reveal about mathematical thinking. In this theoretical paper, I review this work with an eye towards exploring how an embodied perspective can inform our interpretation of these findings. I seek to answer the question of whether or not mathematical reasoning can also be described as situated and embodied rather than as abstract and amodal. If gestures reflect embodied thinking, and if gestures are important in mathematical thinking, this would suggest that mathematical thinking is also embodied.

This line of research is influenced by a now substantial body of data showing that language, as well as other cases of putatively abstract reasoning, may be embodied to a much greater extent than previously appreciated. I will first briefly discuss some of the evidence for embodied cognition in general, focusing on language and motor processes as these are most relevant for thinking about gesture as an embodied representation. I will then discuss evidence that hand gestures are functionally involved in mathematical thinking in both children and adults. Finally, I will address some possible criticisms and data that are difficult to explain from this perspective.

Before reviewing data on the role of the body in language and communication, it is useful to make clear what I mean by the terms. Situated cognition reflects the notion that the real-world environment where processing takes place affects ongoing processing, because cognition involves systems supporting perception and action in the world (Wilson, 2002). Embodied cognition has been used to reflect a variety of claims about cognition (Wilson, 2002). Many claims about embodied cognition fit under the definition of situated cognition proposed here. For example, the notions that cognition can be offloaded onto the environment, and that the environment is part of the cognitive system (Wilson, 2002) are claims about cognition that do not directly reference the body in any manner but do reference the particular time and place where cognition is happening as being important for cognition. To my mind, truly embodied cognition requires that something about the affordances of one’s body is affecting cognition. And, since it seems clear that cognition for acting in the real world requires the affordances of the body to be involved, for embodied cognition to be interesting, we need the affordances of the body to be implicated even when the task at hand does not require overt physical action in the environment. Thus, embodied cognition implicates affordances of the body, while situated cognition implicates affordances of the environment. These two claims about cognition are not mutually exclusive.

Like mathematics, human language has classically been seen as the paradigm case of an abstract and disembodied system. However, spoken language is typically not produced in isolation, but instead is often accompanied by gestures of the hands – spontaneous, non-conventionalized movements of the hands that seem related to the information content of the accompanying speech. Hand gestures are universally produced in conjunction with speech, and they develop in conjunction with speech throughout childhood. These gestures have suggested to many individuals that the representations supporting natural language production might be less abstract and disembodied than they at first seem. Furthermore, because knowledge represented in gestures is represented using the body, this knowledge has been claimed to be embodied (e.g. Gallagher, 2005; Gibbs, 2006; McNeill, 2005; Núñez, 2005). Speaking does not require hand movements, so gesture does meet the criteria of the body being involved, but not required, for the task at hand.

Consistent with the idea that language might be supported by embodied cognition, recent developments in cognitive science more generally suggest that the abstractness of human thinking may be somewhat illusory. Instead, it is proposed that human cognition may be grounded in sensorimotor and bodily processes that are situated in a specific time and place. In the domain of language, language comprehension has been shown to involve motor and mental imagery. For example, Glenberg and Kashak (2002) found that processing sentences that imply movement in a particular direction facilitates production of overt movements in that same direction relative to production of movement in the opposite direction, suggesting that the abstract content of the sentence interacts with the physical process of producing movement. Similarly, Stanfield and Zwaan (2001) showed that processing sentences that imply objects in a particular orientation or configuration facilitates recognition of images of these objects with that same orientation or configuration, again suggesting that the abstract content of the sentence interacts with visual processes for recognizing perceived objects.

Motor processes have also been implicated in cognitive processes that do not require overt motion. For example, mental and physical rotation appear to rely on common mechanisms. Wexler, Kosslyn and Berthoz (1998) demonstrated that, during a mental rotation task, physically moving a joystick in a direction compatible with the mental rotation resulted in fewer errors and faster reaction times compared with performing a motor rotation in the opposite direction. Similarly, Schwartz and Holton (2000) demonstrated that performing an action that would result in a physical rotation in the world (pulling a string to unwind a spool) facilitated mental rotation of the object that would be rotated, while performing an action that would result in the opposite physical action interfered with mental rotation, relative to no action at all. Thus, physical actions can influence mental actions. Performing actions affects visual processing as well. Wohlschläger (2000) demonstrated that hand rotations, as well as planned but not executed hand rotations, influenced the perceived direction of ambiguously rotating stimuli.
Hand gestures are actions performed by the hands, but they do not directly influence objects in the world. Hand gestures may reflect situated and/or embodied cognitive processing underlying communication. Gestures are necessarily situated in space and time in a particular real-world environment. Moreover, the structure of the available environment can affect the gestures that are produced in that environment (Özyürek, 2002). When speakers are allowed to gesture, they talk less about non-present and more about perceptually available information, suggesting that the opportunity to gesture may focus attention on the current environment relative to when gesture is restricted (Alibali & Kita, 2010). Gestures are also associated with spatial reasoning (Alibali, 2005; Chu & Kita, 2008; Kita & Özyürek, 2003) and mental imagery (Morsella & Krauss, 2004; Wesp, Hesse, Keutmann, & Wheaton, 2001). Thus, gestures seem to reflect mental processes that support perception and action in the immediate environment and so they meet our criteria for reflecting situated cognition.

Gestures also appear to reflect processing that is embodied. Gestures are strongly associated with action simulations produced by one’s own body acting in the world (Beilock & Goldin-Meadow, 2010; Cook & Tanenhaus, 2009; Hostetter & Alibali, 2010). For example, the gestures produced after solving the Tower of Hanoi problem reflect specific features of the actions necessary to solve the problem, revealing that speakers are accessing the specific motor plans used to solve the task when they are describing the general set of moves used in the solution. After speakers solve the problem with real objects, which requires lifting of the disks over the top of the pegs to move them, they produce two-handed lifting gestures with high trajectories. In comparison, speakers who solve the problem on the computer, where the objects can be dragged off of the pegs without lifting, produce one handed dragging gestures with lower trajectories (Cook & Tanenhaus, 2009). These differences in gesture occur without differences in the accompanying speech, revealing activation of action representations in the context of an apparently less embodied verbal description. Moreover, these differences occur outside of the environment where the Tower of Hanoi has been solved, and so they meet our criteria for embodied cognition rather than situated cognition. Given that hand movements are not necessary when speaking, gestures will typically meet our criteria for embodied thinking, involvement of the body when it is not necessary. However, if gestures cannot be shown to influence cognition, then they do not provide evidence that embodiment matters for ongoing cognition.
EMBODIED MATHEMATICS

The general perspective of embodied cognition has been applied to mathematical language. In an influential book, George Lakoff and Rafael Nunez (2000) argued that mathematical concepts are supported by visual and motor representations. As a specific example, one basic mathematical concept is the notion of equality. Although equality may seem elementary as a mathematical concept, this concept is relatively new in the history of mathematics (Lakoff & Nunez, 2000, 376). Before inventing the notion of equality, mathematicians used the language of combining things to describe equations, with statements like “Two plus two yields four” (Lakoff & Nunez, 2000, 376), suggesting that interactions among mathematical objects in equations were conceptualized according to the same principles by which real objects interact with one another – putting two rocks with two more rocks yields a set of four rocks, and so there is no relation between the quantities but rather operations over them. According to Lakoff and Nunez, the abstract concept of equality as expressing a relation between quantities is still grounded in physical experience combining objects into groups.

Lakoff and Nunez provided a theoretical account for the grounding of mathematical ideas in physical concepts, and based their empirical evidence for this account in the language used to describe mathematical reasoning. From the point of view of cognitive linguistics, metaphors represent an extension of inferences from a more physical source domain, the ground, to a more abstract target domain. For example, on this account, the mathematical notion of infinity is an extension of the real world experience of processes having an end applied to the counting process. According to Lakoff and Nunez, all mathematical concepts have their basis in inferential structure derived from real world experience, and this basis is seen in the metaphoric language used to describe mathematical concepts. Metaphor provides evidence for grounding of cognition in sensori-motor experience, but does not necessarily implicate embodied cognition because specific properties of one’s body are not implicated in reasoning.

Like metaphoric language, hand gestures are pervasive during mathematical reasoning. It is perhaps not surprising that our earliest mathematical behavior, counting, is typically accompanied by pointing gestures indexing the set of objects that are being counted (e.g. Carlson, Avramides, Cary & Strasberg, 2007). Gestures have been reported in a variety of mathematical tasks, including reasoning about fractions (Edwards, 2009), reasoning about functions (Gerofsky, 2010), and converging sequences (Nunez, 2005). Not surprisingly, mathematicians gesture when they are talking about mathematical concepts (McNeill, 1992; Nunez, 2005). Importantly, the gestures produced by mathematicians do not just refer to written mathematical notion, but rather reflect dynamics of the underlying mathematical ideas (Margheritis & Nunez, 2010; Nunez, 2005).
GESTURES AND MATHEMATICS

My work has focused on children and adults reasoning about equivalence relations. Gestures are pervasive in these tasks, although the material is not particularly spatial or otherwise suited for gesture. For example, in one experiment where gesturing was actually discouraged for half of the sample by requiring participants to restrict their gestures before having the opportunity to gesture at will, 70% of children and 88% of adults gestured while explaining their mathematical reasoning (Goldin-Meadow et al., 2001). These high rates of gesture suggest that gestures may be functionally beneficial to speakers explaining their reasoning.

Further, when participants gesture when they are explaining math problems, they make fewer demands on their verbal and their visuospatial working memory system in comparison with when they produce similar explanations without gesture (Goldin-Meadow et al., 2001; Wagner, Nusbaum & Goldin-Meadow, 2004). In these experiments, children and adults were asked to maintain items in verbal or visuospatial working memory while explaining their solutions to math problems. In this paradigm, the number of items recalled can be used to infer the nature of demand on working memory during the explanation task. In our experiment, participants explained problems in conditions which either included gesture or did not include gesture. Participants were better able to recall the unrelated items when they gestured during the explanation task. Moreover, the effect of gesture depended on how the mathematical content expressed in the gesture is related to the concurrent speech – when the mathematical content in the gesture matched that in the accompanying speech, participants were aided by gesture (Wagner, Nusbaum & Goldin-Meadow, 2004). The effect does not depend on whether the presence of gesture was spontaneously generated or was the result of experimental instructions. When speakers are instructed to gesture, they show benefits comparable to those associated with gesturing spontaneously (Cook, Yip & Goldin-Meadow, 2011), and when speakers are instructed not to gesture they show decrements comparable to those associated with failing to gesture spontaneously (Goldin-Meadow et al., 2001; Wagner, Nusbaum & Goldin-Meadow, 2004). In addition, the benefit is not simply the result of moving one’s hands. Instead, the movements produced must express meaning related to that seen in the accompanying speech (Cook, Yip & Goldin-Meadow, 2011). Taken together, these results suggest that the conceptual information encoded in gesture is important for speakers engaged in mathematical reasoning. That is to say, using the body matters, when one is communicating mathematical concepts.

Similar findings are seen in other math tasks. As an example, Alibali and DiRusso (1999) found that pointing gestures improved early counting performance by facilitating keeping track of items as well as by coordinating the number words and items. Similarly, Graham (1999) found that children adhered to one-to-one correspondence in their gesture prior to their speech. The supportive effect of
gesture on counting is not limited to children who are just learning to count. Pointing gestures also facilitate both counting and combining of addends in adults (Carlson, Avraamides, Cary & Strasberg, 2007).

Data from other paradigms provides converging evidence for the claim that important mathematical content can be encoded in hand gestures. Using EEG, researchers compared the brain activation associated with gestures of mathematical concepts and with words for these same concepts by presenting words or gestures at the end of sentences and asking subjects to report whether or not the words or gestures matched the context. The authors found that words and gestures elicited identical N400 effects. Because the N400 is thought to index semantic processing of information and is elicited when semantically anomalous material is encountered, these findings suggest that gestures representing mathematical concepts are processed in the same way as conventional mathematical words (Lim et al., 2009).

GESTURE AND COGNITIVE CHANGE IN MATHEMATICS

The fact that gestures are spontaneously produced in conversations about mathematical concepts, and that these gestures can influence speakers’ processing, offers a first line of evidence that mathematical representations may not be as abstract and disembodied as they seem. A second line of evidence comes from investigations of the factors involved in changing these representations – how do mathematical gestures affect learners? If gestures facilitate learning of mathematical concepts, this would provide supporting evidence for the claim that mathematical concepts are situated and embodied. Input that is aligned with the underlying representational format of the knowledge to be changed and/or the knowledge to be gained should be particularly likely to lead to changes in representation of this knowledge. And, in fact, hand gestures have been shown to be an effective mechanism of change in mathematical thinking. Gestures have been linked to changes in children’s mathematical thinking in a variety of ways, and both the gestures that children produce and the gestures that children observe have reliable effects on subsequent thinking.

First, the link between changes in children’s mathematical thinking and gesture is seen in the fact that children’s own gestures are diagnostic of the state of their thinking. Gesture can be used as an individual difference factor in order to assess the likelihood that a particular child will be receptive to instruction in a concept. In particular, children who produce gestures that encode relevant problem features that are not expressed in the accompanying speech – so-called gesture-speech mismatches – are more likely to benefit from instruction than children who do not gesture, or children who produce only gestures that express information identical to that encoded in the accompanying speech. For example, a child asked to explain the math problem $6 + 5 + 3 = _ + 3$ might put 14 in the blank and justify his answer by saying, “I added the six plus the five plus the three,” and pointing at the
six, the five and the three on the left side of the equal sign. In this case, the child has expressed the same information, adding up the numbers to the equal sign, in both speech and gesture. Another child might say the same thing, while pointing at the six, the five and the three on the left side of the equal sign and then the three on the right side of the equal sign, gestures which suggest a strategy of adding up all the numbers in the problem, not just the numbers on the left side of the equal sign. Production of these sorts of mismatching gestures is associated with subsequent success in learning to solve these problems correctly (Alibali & Goldin-Meadow, 1993; Church & Goldin-Meadow, 1986; Perry, Church & Goldin-Meadow, 1991).

Importantly, representation of multiple pieces of information in speech does not have this same predictive validity (Church et al., 1995), which suggests that gesture may be special in its sensitive to emerging or developing knowledge. Research has also shown that this period of mismatching gesture does not just reflect the fact that change is imminent, but rather appears important for changing children’s thinking because children who acquire the correct concept without proceeding through a state of gesture-speech mismatch are less likely to maintain their new correct knowledge than children who go through the mismatch state (Alibali & Goldin-Meadow, 1993).

Indeed, even children who do not appear to be ready to learn can generate helpful gestures when they are asked to move their hands, and these gestures can support the subsequent acquisition of new knowledge just like spontaneously produced gestures. In a series of studies, my colleagues and I asked some children to be sure to gesture while explaining math problems prior to receiving instruction in these problems (Broaders, Cook, Mitchell & Goldin-Meadow, 2007). Some children who were asked to gesture generated new strategies in gesture, strategies that they did not produce in a baseline period prior to this instruction to gesture. In contrast, children who were asked not to gesture, or who were asked to simply solve additional problems, were very unlikely to generate new strategies in speech; this suggests that it was not simply solving more problems or solving problems in response to instruction that led to the generation of new strategies. Importantly, the new strategies that children generated were not just mindless hand waving, but instead were related to the task at hand. When all children were given instruction in the task, those children who generated new strategies were more successful than children who had not generated new strategies (Broaders, Cook, Mitchell & Goldin-Meadow, 2007). Similar findings have been reported by Pine, Luflkin and Messer (2004), who found that children’s improvement in a balance task was mediated by the gestures that they produced in response to instruction. Alibali (1999) reported that with repeated problem solving, children spontaneously generated and abandoned strategies in gesture more frequently than in speech, suggesting that children may use gesture as an efficient means of experimenting with developing problem-solving ideas. When children gesture, either spontaneously or in response to instruction, they sometimes generate helpful representations.
Children can also learn from helpful gestures when they are given to them. In two studies, my colleagues and I asked children to produce particular gestures and observed the effects on their subsequent learning (Cook, Mitchell & Goldin-Meadow, 2008; Goldin-Meadow, Cook & Mitchell, 2009). The first study found that children who were asked to produce correct gestures were more likely to show sustained learning than comparable children who were asked to produce the correct speech. In this study, we asked children solving mathematical equivalence problems to either 1) express the concept in words by saying, “I want to make one side equal to the other side,” 2) to express the concept in gesture by sweeping one hand under the left side of the problem and then the other hand under the right side of the problem, or 3) to do the speech and gesture together. Children produced the words, gestures or both throughout instruction in the concept. We found that children who produced the gestures alone, or who produced the words and gestures together, were particularly likely to maintain their learning three weeks after instruction (Cook, Mitchell & Goldin-Meadow, 2008). In a second study, we manipulated how helpful the gestures were, and found that children learn more from gestures that are more helpful. In this study, children were again taught the words, “I want to make one side equal to the other side” but with the accompanying gesture of either a V hand under the two numbers on the left side that add together to make the answer, or a V hand under the two numbers on the left side that do not add together to make the answer. Both of these gestures suggest the grouping strategy, but the correctly located gestures do so more strongly. Children were more likely to learn in the condition where the gestures were more helpfully placed, suggesting that it is not simply moving one’s hands that is helpful, but rather, moving one’s hands in an informative manner (Goldin-Meadow, Cook & Mitchell, 2009). Moreover, in this study, the information represented in gesture was never presented in speech to children, yet children were observed to spontaneously express this information verbally, suggesting that they were both extracting the meaning from the gesture and translating into a new format (Goldin-Meadow, Cook & Mitchell, 2009). Related findings have been seen in children’s developing understanding of physical reasoning. For example, Boncoddo, Dixon and Kelly (2010) found that children’s discovery of the alternation rule for gear interaction could be predicted from the number of alternating hand actions that were produced during learning.

Second, the link between changes in children’s mathematical thinking and gesture is seen in the fact that providing gesture in the input that learners receive is an effective mechanism for changing children’s mathematical thinking. Children learn more from instruction that includes gesture than from comparable instruction that does not include gesture (Ping & Goldin-Meadow, 2008; Valenzeno, Alibali & Klatzky, 2003). Moreover, children are particularly likely to learn when they are provided with input that contains information in gesture that is not expressed in the accompanying speech, that is to say, information in gesture that is like the information that would be spontaneously produced in association with transitions in
knowledge (Singer & Goldin-Meadow, 2005). Importantly, in this study, children learned more when multiple pieces of information were simultaneously distributed across speech and gesture rather than presented sequentially in speech. However, these findings cannot clarify whether it was the simultaneity of speech and gesture, or the representation of information in gesture, that led to the gains in learning.

One way that providing gesture in the input leads to learning is by eliciting appropriate gesture from learners. In one study, we varied the input that children received so that some children received gesture and other children did not (Cook & Goldin-Meadow, 2006). We also observed the gesture children produced in response to this input and found that children who mimicked the instructor’s gesture when they were exposed to it, and children who spontaneously generated comparable gestural representations when they were not exposed to it, were particularly likely to learn (Cook & Goldin-Meadow, 2006). This finding suggests that children may catalyze changes in their own thinking by gesturing, a third mechanism by which gesture may change thinking.

**GESTURE AND MATHEMATICS IN ADULTS**

However, the finding that gesture is important in children’s mathematical reasoning still leaves open the possibility that mathematical representations are becoming more abstract over development independent of the influence of gesturing. Children’s mathematical reasoning may be more embodied than adult’s mathematical reasoning, and as children learn, they may be developing more abstract and less embodied representations of mathematical concepts. However, gesture appears to operate as a *continuous* factor in development. The patterns of performance that we see in children are reflected in adults.

First, as I have suggested, gestures are also observed when adults are engaged in mathematical reasoning. In fact, adults gesture even when solving mathematical problems alone in a room and without any plans to subsequently communicate their solution procedures to anyone (Smith, 2007). As described above, adults as well as children benefit from gesturing while explaining mathematical reasoning (Goldin-Meadow et al., 2001; Wagner, Nusbaum & Goldin-Meadow, 2004). Even when overt gestures are not observed, action representations appear to be activated in support of mathematical tasks in adults. Presenting numbers changes activation in hand muscles (Sato, Cattaneo, Rizzolatti, & Gallese, 2007). Digit magnitude interacts with grip aperture during reaching such that higher digits are associated with larger grip aperture (Andres, Ostry, Nicol, & Paus, 2008). And smaller numbers facilitate production of precision grips while larger numbers facilitate production of the power grip (Lindemann, Abolafia, Girardi, & Bekkering, 2007).
Moreover, like the gesture-speech mismatch findings reported in children, gesture appears to provide a privileged picture of adults’ mathematical reasoning and problem-solving. For example, Alibali and her colleagues (1999) found that the strategy college students were likely to use when solving algebra problems could be predicted from the gestures used when explaining the problems. Indeed, in situations where gesture and speech were in conflict, speakers were as likely to use the strategy expressed in gesture as that expressed in speech. In a second example, speakers gesturing about limits and continuity in calculus produce gestures that convey notions of motion, although, of course, the underlying mathematical concepts and structures are not moving in any space or time (Nunez, 2008). Instead, these gestures may reflect the conceptual metaphor underlying these mathematical ideas (Nunez, 2008).

Moreover, consistent with what is demonstrated in children, seeing gestures appears to be important in changing mathematical thinking in adults. In ongoing work, we have developed an adult math task that is comparable to the equivalence task that we use with children. In this paradigm, adults learn to solve mathematical equivalence problems in a new mathematical system – a commutative group of order three (adapted from Kaminski, Sloutsky, & Heckler, 2008). We varied the input that adults received during learning to include gesture or not include gesture. Our results suggest that, just like children, adults benefit from seeing gesture during math instruction (Fenn & Cook, 2010). Moreover, similar information presented via a laser pointer does not provide the same benefit as hand gesture, suggesting that gesture may be unique in facilitating listener understanding (Cook, in preparation).

Producing gesture may also be important in changing thinking in adults, although there has not been as much work with adults as has been done with children. In a recent study, Beilock and Goldin-Meadow (2010) demonstrated that the particular gestures speakers produced while explaining the Tower of Hanoi problem-solving task influenced the subsequent problem solving solutions of the speakers. Speakers who produced more one handed gestures during their explanation had more difficulty transferring their experience with the Tower of Hanoi to a new Tower that required two handed lifting actions. Speakers who did not produce an explanation or who did not switch to a new tower did not show difficulty transferring knowledge. These findings suggest that the gestures that adult speakers produce have an influence on their mental representations and the actions that they subsequently perform.

More generally, problem solving in mathematics is influenced by visual representations, even in adults who are practiced problem solvers. For example, McNeil, Rittle-Johnson, Hattikudur and Peterson (2010) reported that adults can be induced to incorrectly solve mathematical equivalence problems when their basic arithmetic knowledge is activated, because the visual context of arithmetic supports incorrect conceptualization of equality. Landy and Goldstone (2007) similarly reported that adults were more accurate at judging algebraic expressions when the
physical spacing of the elements in the problem matched the underlying mathematical structure, although this spacing is irrelevant in formal notational systems. More generally, Kirschner and Awtry (2004) report how knowledge of algebra depends on appropriate recognition of visual patterns in the environment rather than on acquiring abstract rules for combining symbols, and Kellman and colleagues (2008) found that providing training in visual pattern recognition supported learning of both algebra and fractions. If adults are generally sensitive to the visual context of mathematical problem solving, there is little reason to suspect that adults are different from children in their sensitivity to gesture in supporting changes in mathematical thinking.

**SOME POTENTIAL CRITICISMS**

Taken together, the findings discussed here suggest that even mathematics, often considered the most abstract of the sciences, may have its roots in more basic systems for perception and action, and that we may find evidence for this in the hand movements that accompany speech. However, there are some findings that appear difficult to integrate with the perspective reported here. In one example, Nicoladies, Pika and Marentette (2010) recently reported that preschool children were more successful with number words than number gestures in both the give-a-number task and the how-many task. However, given the greater experience that children have with number words compared with number gestures, it is not clear that these results truly reflect differences attributable to the nature of the representation, rather than the greater frequency of exposure to number words. Nonetheless, these findings suggest that gesture may not always be helpful. In fact, the early number words are phonologically quite distinct from one another, while the number gestures overlap considerably. Thus, unlike pointing gestures, number gestures may not be as helpful for developing an understanding of exact quantity.

On the other hand, finger representations have been implicated in developing an understanding of numeric quantities in a variety of studies. Butterworth (1999) has proposed that because of extensive experience counting, finger representations are inexorably linked to number (although see Penner-Wilger & Anderson, 2008 for an alternative explanation). There is some evidence consistent with this claim. First the ability to individuate the fingers via touch, finger gnosis, can predict subsequent numerical performance (Fayol, Barrouillet, & Marinthe, 1998; Noël, 2005). And, training students on finger representations can support learning of numeric representations (Gracia-Bafalluy & Noël, 2008). Moreover, children make a greater-than-expected number of errors in addition of magnitude five, precisely what would be expected if internal hand-based representations were supporting this process (Domahs, Krinzinger, & Willmes,
All of these data point to the conclusion that bodily representations are involved in numerical understanding, at least at the early stages of learning.

An alternative possibility, however, is that gestures produced in mathematical tasks are simply task artifacts and thus not actually involved in the reasoning process. Explaining math with a visible problem available may make it particularly likely that people gesture, without actually involving the gestures in the underlying conceptualization. Yet, we know that gestures are produced in many situations where there are no available references (e.g. Bavelas, Gerwing, Sutton, & Prevost, 2008; Ping & Goldin-Meadow, 2008; 2010), indicating that gestures are not simply elicited in highly supportive contexts. And, like the findings reported for mathematical reasoning, gestures facilitate working memory and conceptual development during Piagetian conservation tasks even when visual referents are not available (Ping & Goldin-Meadow, 2008; 2010). As I have argued, the available evidence suggests that producing gestures while explaining mathematical reasoning influences ongoing cognitive processing and is thus functionally involved in the task.

A final, perhaps more troubling, argument against the idea that gestures reveal the embodied nature of mathematical reasoning, is the suggestion that gestures reflect properties of communicating about mathematical ideas that are distinct from the ideas themselves. As any mathematician will report, the process of communicating mathematical ideas in proofs is quite distinct from the process of generating these ideas in thought. Gestures may be important not for the ideas themselves, but for the packaging of these ideas so that others may understand them. Yet even if this were the case, gestures would still be important for mathematical thinking. Most mathematical knowledge comes to us communicated from someone else, and so the communication of complex mathematical ideas is a central part of mathematical knowledge.

Moreover, mathematicians do report thinking in images. Albert Einstein famously said, “Words and language, whether written or spoken, do not seem to play any part in my thought processes. The psychological entities that serve as building blocks for my thought are certain signs or images, more or less clear, that I can reproduce and recombine at will (1945, p. 142).” Keith Devlin reports, “To me, then, learning new mathematics is like constructing a mental house in my mind; understanding that new mathematics is like becoming familiar with the interior of my mental house; and working on a mathematical problem is like arranging the furniture. Thinking mathematics is like living in the house. As a mathematician, I create a symbolic world in my mind and then enter that world (2000, p. 127).” Clearly, mathematicians report that their thinking is situated and embodied. If they are to be believed, gestures might be just the right tool for developing mathematical thinking.
ACKNOWLEDGEMENT

Susan Wagner Cook would like to thank Larissa Samuelson for stimulating conversations related to the ideas discussed here, as well as for specific feedback on the manuscript.

REFERENCES

Cook, S. W. (in preparation). Don’t show me the light! The effects of laser pointers and hand gestures on listeners’ understanding.


---

*Cognition, Brain, Behavior. An Interdisciplinary Journal*

*15 (2011) 553-570*


ABSTRACT THINKING IN SPACE AND TIME: USING THE ENVIRONMENT TO LEARN WORDS

Larissa K. SAMUELSON*

Department of Psychology & Delta Center, University of Iowa, Iowa City, IA, USA

ABSTRACT

A substantial body of work has examined the gestures children and adults make when they talk and found them to be a revealing window on the processes of cognitive change. In her paper, Susan Wagner Cook (this volume) reviews this work along with her own recent work examining the gestures children and adults produce when they talk about math. She argues that the combined data point to a new view of our mathematical knowledge as embodied. Here I comment on Cook’s arguments, highlighting how this view of math as embodied offers new insights for our understanding of classic developmental themes, in particular, the continuity versus discontinuity dichotomy. In addition, I present a brief summary of recent work on how children use their bodies in another realm typically thought of as abstract—understanding referential intent. I present an embodied account of how children disambiguate speaker intent in novel naming situations and argue that, as in the case of embodied math, an embodied view of cognition can help elucidate developmental mechanism.

KEYWORDS: Cognitive Development, Word Learning, Mathematical representations, Binding, Embodiment

The last twenty years of developmental theory and empirical work has seen a return to the possibility that cognitive development is grounded in the sensorimotor (see, for example, Barsalou, Breazeal & Smith, 2007; Iverson, 2010; Pereira, Smith & Yu, 2008; Port & van Gelder, 1995; Thelen & Smith, 1994; Spivey, 2007; Smith & Sheya, 2010). Like Piaget, researchers are again linking abstract cognitive activity to the physical realities of the child’s active behaviors and abilities. One prominent area of research in this regard is the now extensive body of work examining the

* Corresponding author:
E-mail: larissa-samuelson@uiowa.edu
gestures children (and adults) make while doing cognitive tasks. This work has examined gesture as both a revealing window on cognitive change and as a potential motivator of that change. In what follows I comment on Susan Wagner Cook’s (this volume) synthesis of recent work examining the role that gesture might play in the development of children’s, and adults’, mathematical knowledge. Following this, I highlight how this view of math as embodied offers new insights for our understanding of classic developmental themes, in particular, the continuity versus discontinuity dichotomy. In addition, a review of recent work on how children use their bodies to disambiguate speaker intent in novel naming situations suggests how an embodied view of cognition can help elucidate developmental mechanism.

EMBODIED MATH: INSIGHTS FOR DEVELOPMENTAL SCIENCE

The synthesis presented by Cook (this volume) on the role of gesture in the communication, development and use of mathematical knowledge presents a new and radical view of how we use our bodies—in particular our hands—to think. This is a significant departure from previous discussions of the relation between gesture and cognition. Reported interest in gesture and what it might reveal about cognition goes back to Quintilian in AD 100 (Quintilian, 1924). The significance of this work for cognitive development blossomed in the early 1990s with the seminal work by Goldin-Meadow on gesture-speech mismatches (Perry, Church & Goldin-Meadow, 1988, Goldin-Meadow, Alibali & Church, 1993; Perry, Church & Goldin-Meadow, 1992). Studies demonstrating that differences in children’s knowledge as expressed in speech versus their hands could predict which children would learn from training in conservation and math problems fit emerging ideas of the importance of variability in cognitive development (e.g., Siegler, 1996; Thelen & Smith, 1994). Nevertheless, the prevailing view was that gestures provided another window on to the process of change in cognition—not that gestures were actually part of thinking.

Clearly, this view of gestures as a useful window provided the basis for important work in numerous areas of cognitive development research (see Goldin-Meadow, 2004 for a review). Nonetheless, the view of gesture as a useful window on cognitive change has not led to a radical departure from more traditional ways of thinking about development. Research on the role of gesture certainly supports a view of the child as an active participant in the developmental process, and highlights individual differences in the expression of thought via variability seen in the more frequent gestures of some children and adults as opposed to the more restrained hands of other participants. However, this view does not recommend a change in the way we view the classic developmental dichotomies. Likewise, gesture-speech mismatches provide a connectedness between the child’s level of cognition prior to training and her abilities following training, as does the fact that
both children and adults benefit from the ability to gesture while reasoning provides a connectedness between the processes of cognitive change in children and adults. Yet, in both cases the connectedness can, and has been, understood without embracing an embodied view of cognition.

The more recent work, by Cook and others, however, shows something different — that having participants gesture in particular ways actually helps them learn. In this work we can start to see gestures as a mechanism of change in their own right. And further, we begin to see hints that we need changes in our own thinking about development. In particular, this work suggests there is continuity in the system, and not only in processes that create development across the timescales of child to adult and pre- to post-training, but more generally. For example, the fact that children learn more when multiple pieces of information are distributed across modalities (Singer & Goldin-Meadow, 2005) suggests a level of continuity that allows flow and integration of information presented in contrasting forms. Similarly, the fact that some children taught to gesture when solving problems spontaneously transferred the information contained in the gesture to their speech (Goldin-Meadow, Cook & Mitchell, 2009) suggests that the information contained in the gestures and speech is continuous enough to be accessible by both the linguistic and sensorimotor systems.

Of course, some will argue that these examples show only the ability to quickly translate representations from one cognitive subsystem to another — a view that comfortably stays within the information-processing boxology of the last 50 years. However, the increasing theoretical and empirical literature demonstrating the continuous, coupled, and dynamic nature of the cognitive system (see above and Barsalou, Brazeal & Smith, 2007; Port and Van Gelder, 1995; Thelen & Smith, 1994; Smith & Sheya, 2010, for reviews and examples) fits with the more radical perspective that the continuity seen in these gesture studies is due to the embodiment of thought.

This view also points toward an answer to the question of how it is that gesture actually helps us think. In particular, a view of cognition as embodied highlights a new medium for thought that is not often considered when we try to understand complex cognitive processes — space. The idea is this — bodies live in, and move in, and take-up space. Space is fundamental to all our behaviors, from knowing to project your voice in a large lecture hall to cover the distance to a listener, to directing eye movements to novel locations as you gather information about a cluttered environment, to knowing how much acceleration to allow in a movement of a hand in order to push the elevator button without bumping the person standing next to the door. And, critically, an embodied view of cognition brings space into the realm of consideration as a possible promoter, mechanism, and grounding substrate for cognition.

Examples of our use of space to think abound, once you start considering it — we can often remember where on a page a critical section of text was, American
Sign Language uses space to denote grammatical relations, and of course, gestures use space to depict information (see also Richardson and Spivey, 2007). And children can use space to think and communicate from an early age (see also Kirkham and Richardson, 2010). As an example, Figure 1 illustrates a one-year-old child’s use of space to refer to a desired food. As is clear in the sequence, the favorite food at this meal was associated with a particular location on the plate and in pointing to that location the child could clearly indicate that he wanted more of that food.

Figure 1.
A one-year-old child using space to refer at one year of age. In the left panel he reaches for his favorite food, bananas, which he finishes (center). In the right panel he asks for more by pointing to the location in space associated with the bananas. Importantly, the position of the plate has shifted on the highchair tray, suggesting that he is not simply moving his arm to the same exact place in space.

This is not to say that children (and adults) come equipped to reason about space innately — studies on hiding and finding tasks from AnotB to DeLoache’s scale models to Plumert’s object placement tasks (for review see chapters in Plumert & Spencer, 2007) – all confirm that there is a protracted developmental course in spatial cognition. Rather, the argument here is that if cognition is embodied, then the properties of the body — including the relation of that body in space — should be able to be used to serve cognition. Recent work by my colleagues and I demonstrate that this is in fact the case in another form of very abstract cognition — learning the referent of novel words. In parallel with what has been shown by Cook and others for gesture and mathematics, we show that space is a powerful cue that helps children link abstract words with their physical referents.
**USING SPACE AND THE BODY TO LEARN WORDS**

Linda Smith, Lynn Perry, John Spencer and I have recently been examining the possibility that young children can use space to bind novel words to objects (Samuelson, Smith, Perry & Spencer, 2011). Our studies are based on a seminal study by Baldwin (1993) examining young children’s ability to read the referential intent of a speaker. A schematic of the task is presented in Figure 2. A novel object is presented to a 20-month-old child for exploration and manipulation on one side of a table. This object is then removed and a second novel object is presented on the other side of the table and the child is again allowed to reach for, grasp and explore the object. This is repeated for a set of familiarization trials. Both objects are then placed in separate opaque buckets on either side of the table. The experimenter looks into one bucket and says “Modi!” The object from the other bucket is then taken out and placed on its side of the table. It is removed after the child examines it and the other object is placed on the table. After examination, this item is also removed. Both objects are then placed on a tray on the center of the table. The tray is pushed toward the child, and the experimenter asks, “Can you get me the modi?” Children retrieve the object that was in the bucket the experimenter was looking in when she said the novel word .70 of the time. Baldwin interpreted this result as suggesting children understood the pragmatic use of eye gaze as an intentional cue (Baldwin, 1993).

In contrast, we have argued this result was due to children’s use of spatial memory to bind words to objects (Samuelson et al., 2011). We have implemented this proposal in a Dynamic Neural Field (DNF, see Spencer & Schöner, 2009) model that provides a process account of how children can use the shared space of social interactions to link the novel name to the novel object, even when the two are not presented simultaneously. The details of the model are beyond the scope of the current brief paper. However, in short, the children act on the objects during the familiarization trials. They look at the objects in particular locations in space. They reach for them. They manipulate them. And then they attend as each object is removed from its side of the table. These actions create associations between each of the novel objects and their unique locations via the child’s bodily actions towards those objects and locations. Then, when the experimenter looks into a bucket placed at one of those unique locations and says the name, the child’s memory of the object previously seen and acted on at that location is recalled and bound to the novel name. Thus, the child is able to link the novel name to the correct object via the space that her body, her attention, her actions and the object, occur in.
Figure 2.
Schematic of the experiments by Samuelson and colleagues examining children’s ability to use space to disambiguate reference. The task is based on Baldwin’s 1993 study which was replicated in the first experiment. Subsequent experiments tested the proposal that space was critical to children’s performance in this task and supports the suggestion that early reference is embodied.

We have tested several predications of this account and quantitatively simulated children’s behavior with our model. In our first experiment (see Figure 2) we replicated Baldwin’s task in a control condition (No Switch) and disrupted space as a cue in our experimental condition (Switch) by changing the location of the objects on the second familiarization trial. The remainder of the procedure followed that of the Baldwin task. If children’s actions in space help to bind labels to objects, then disrupting the consistency between the objects, actions, and particular locations in space, should weaken this performance. Children performed identically to those in Baldwin’s study in the replication condition, choosing the object from the named bucket on .73 of test trials (compared to .70 in Baldwin’s study). In contrast, children in the Switch condition performed at chance levels (see Figure 3). Note that if binding the object and label depended on understanding the intentions of the experimenter at the time of labeling, it should not matter where the objects had been beforehand. That it did, demonstrates the importance of space in binding labels to objects.
Figure 3.
Results from Samuelson et al.’s (2011) studies of embodied reference. Children’s performance is shown in the black bars along with the results of simulations of a DNF model which provided a strong fit to children’s behavior and elucidates the processes that enable children to bind names to unseen objects (see text and Samuelson et al., 2011 for details).

In our second experiment we went a step farther and removed the hidden object component of the task by not placing the objects in buckets during the naming event. Rather, following the familiarization presentation, the experimenter pointed to the place on the table where one of the objects had been and said, “Modi!” The rest of the procedure was identical to Baldwin’s study. Children demonstrated linking of the object that corresponded to the named location at the same rate as those in the No Switch condition of Experiment 1 (see Figure 3).

In our third experiment we pitted space against temporal congruence. The same basic no-bucket procedure was used with two exceptions. First, there were four familiarization trials for each object. Second, during the labeling event, only one object was present on the table, but in the opposite of its previous location, and the experimenter directly pointed at this object and said, “Modi!” Children chose the temporally linked object significantly less than predicted by chance. In other words, they selected the object that had been in the labeled location earlier even though it was not there during the labeling itself. As is clear in Figure 3, in a control condition during which the object and label were presented together at that location without prior familiarization (corresponding to starting the experiment at the naming event), children bound the name and object at high levels.
Finally, as a critical test that an embodied spatial cue is central to this phenomenon, we examined whether another salient cue, color, would have the same result. Instead of presenting and letting the child engage with each object in the same unique spatial location each time, the objects were familiarized on their own colored tray each time, at a central location (see Figure 2). Thus, each object was associated with a distinctive cue, but the cues did not afford differences in children’s actions. Rather on all familiarization trials they oriented towards and looked at the same central location for both objects, and reached for the same location relative to their bodies. During the labeling event, no objects were present (as in the prior experiments) and one of the two colored trays was placed at the center of the table. The experimenter pointed at the tray and said “Modi!” The test was the same as in the other experiments. Children performed at chance levels in this task. And importantly, on memory check trials inserted after the main test, children were able to indicate which object went with each colored tray at high levels well above what would be expected by chance (.70 correct). Thus, the color cue was distinctive and memorable, but it was not sufficient to allow children to bind the objects and labels. We believe that this is because unlike the spatial cues, the color ones did not afford differential actions towards the objects. That is, the spatial cue, but not the color one, was embodied.

Together with the DNF simulations, these studies show that children can use consistency in spatial location to bind a novel name to a novel object in an ambiguous naming situation. These studies suggest that the body’s movements can be used as deictic references to bind objects in the physical environment to variables in cognitive operations. Thus, this work provides another example of the proposal that even cognition as abstract as understanding reference and mathematical reasoning is embodied. Further, this work suggests a possible way that gestures in Susan Cook’s training studies may be doing cognitive work. That is, the DNF model suggests that the way space works to bind names to objects is by serving as an embodied associative cue. When the name is presented while the location of the target object is cued, the child’s memory of the object associated with the cue is retrieved and then bound to the new name. Similarly, the hands doing the gestures in Susan Cook’s training studies may serve as a cue to the relevant concept — a placeholder for some of the resource-heavy information necessary to learning. Just as a single small souvenir can serve as a reminder of a week-long trip, hand gestures may have the potential to activate and re-activate abstract conceptual knowledge in new situations.
CONCLUSIONS

Clearly the extent to which one is moved by these arguments will rest heavily on one’s acceptance of the definition of embodiment proposed by Cook (this volume) and others, combined with one’s view of how well the empirical examples presented meet the standards set by that definition. For this reason, it is worth reconsidering exactly what embodied cognition is and why it matters.

The critical difference between situated cognition and embodied cognition outlined by Cook (this volume) was that embodied cognition required that something about the affordances of one’s body affect cognition. While Gibson’s term can be very difficult to interpret and apply from the perspective of more traditional information processing approaches, at its heart affordances are the set of possible behaviors and actions given the fit between the actor’s body — including all its receptors and effectors — and the impinging environment. Thus, to act is to have your affordances involved. From this perspective, then, saying that we should see the affordances of the body affecting cognition should not be misinterpreted as suggesting we should expect to see a measurable difference in a behavior as a signature of embodiment. Every behavior — from perceiving that a toy is on one side of the table while your mom is smiling at you from across the table, to the movements of your hands in an arch as you try to explain how a t-test determines the significance of a difference between means with respect to variability — already contains that signature because it was generated by your body. The implication of this for cognitive development, then, is that we do not have to go to increasingly abstract places, or ungrounded dichotomies, for our understanding of change. Rather, we can look to the body and to the continuous nature of the system as a possible source of information grounding, understanding and, indeed, explaining cognition.

ACKNOWLEDGEMENTS

Writing of this paper was supported by Award Number R01HD045713 from the Eunice Kennedy Shriver National Institute Of Child Health & Human Development to Larissa Samuelson. The content is solely the responsibility of the author and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute Of Child Health & Human Development or the National Institutes of Health. Larissa would like to thank Susan Wagner-Cook for stimulating discussions of the research and ideas presented in the paper as well as specific feedback about the manuscript. She also thanks John Spencer for helpful discussions of the topic.
REFERENCES


Cook, S.W., (this volume). Abstract thinking in space and time: Using gesture to learn math.


Cognition, Brain, Behavior. An Interdisciplinary Journal
15 (2011) 571-581