Cogniție, Creier, Comportament / Cognition, Brain, Behavior
Vol. IX(3), 239-272, 2005

LANGUAGE AND THE MIRROR SYSTEM: A PERCEPTION/ACTION BASED APPROACH TO COMMUNICATIVE DEVELOPMENT

Michael A. ARBIB[*1], *Erhan OZTOP*[2], *Patricia ZUKOW-GOLDRING*[3]

[1]Computer Science, Neuroscience and USC Brain Project,
University of Southern California, USA
[2]JST-ICORP Computational Brain Project ATR, Computational Neuroscience
Laboratories, Kyoto, Japan
[3] Department of Linguistics, University of Southern California, USA

ABSTRACT
In answering "What are the sources from outside the self that inform what the child knows?", our basic idea is that a shared understanding of action grounds what individuals know in common. In particular, we root the ontogeny of language in the progression from action and gesture to speech or signed language. What then might the evolutionary path to language and the ontogeny of language in the child have in common? We can characterize the source of the emergence of language in both as arising from perceiving and acting, leading to gesture, and eventually to speech or signed language. Rizzolatti & Arbib (1998) argue that the brain mechanisms underlying human language abilities evolved from our non-human primate ancestors' ability to link self-generated actions and the similar actions of others. On this view, communicative gestures emerged eventually from a shared understanding that actions one makes oneself are indeed similar to those made by conspecifics. Thus, what the self knows can be enriched by an understanding of the actions and aims of others, and vice versa. From this view, the origins of language reside in behaviors not originally related to communication. That is, this common understanding of action sequences may provide the "missing link" to language. A corollary of this, not always sufficiently stressed, is that the full pattern of communication and understanding rests on a far richer set of brain functions than the core "mirror system for grasping" said to be shared by monkey and human. We report here on the early stages of a research program designed to integrate empirical cross-cultural studies of infant communicative development (Zukow, 1990; Zukow-Goldring, 1996, 1997, 2001) with a computational approach to the mirror system in monkey, human and robot (Oztop & Arbib, 2002). We stress that mirror neurons are not innate but instead correspond to a repertoire of learned actions and learned methods for recognizing those actions. Our aim is an integrated view of how perceiving and acting ground the emergence of language. Our effort is to integrate analysis of the influences of the environment and, in particular, of the ways in which caregivers attune the child to that environment ("what the head is inside of" [Mace, 1977]) with the study of the neural mechanisms that can learn from these attunements ("what is in the head"). We seek to

---

[*] *Corresponding address:*
E-mail: arbib@usc.edu, erhan@atr.jp, zukow@usc.edu

delineate what children might "know" from birth, and the interplay of perceptual processes with action that might allow them to come to know "what everyone else already knows", including word meaning.

KEY-WORDS: *mirror system, language, communication, perception/action*

A mirror system primer

In this section we review data on the monkey brain and our own modeling thereof to provide the substrate of basic action recognition mechanisms that we believe lie at the core of both phylogenetic and ontogenetic accounts of the development of language capabilities. The neurophysiological findings of the Sakata group on parietal cortex (Taira, Mine, Georgeopoulos, Murata, & Sakata, 1990) and the Rizzolatti group on premotor cortex (Rizzolatti et al., 1988) indicate that parietal area AIP (the Anterior Intra-Parietal sulcus) and ventral premotor area F5 in monkey form key elements in a cortical circuit which transforms visual information on intrinsic properties of objects into hand movements that allow the animal to grasp the objects appropriately (see Jeannerod, Arbib, Rizzolatti, & Sakata, 1995 for a review.) Other studies lead us to postulate that the storage and administration of sequences of manual actions (inhibiting extraneous actions, while priming imminent actions) is carried out by the portion of the supplementary motor area (SMA) known as pre-SMA and the basal ganglia, respectively, which cooperate in the phasing in and out of appropriate F5 activity as a given task unfolds.

Motor information is transferred from F5 to the primary motor cortex (denoted F1 or M1), to which F5 is directly connected, as well as to various subcortical centers for movement execution. For example, neurons located in area F5 discharge during active hand and/or mouth movements (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996a; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Moreover, discharge in most F5 neurons correlates with an action rather than with the individual movements that form it, so that one may classify F5 neurons into various categories corresponding to the action associated with their discharge. The most common are: "grasping-with-the-hand" neurons, "grasping-with-the-hand-and-the-mouth" neurons, "holding" neurons, "manipulating" neurons, and "tearing" neurons.

The FARS model (Fagg & Arbib, 1998) makes clear certain conceptual issues that will be crucial at later stages of the argument. It provides a computational account of what we shall call the *canonical* system, centered on the AIP → F5 pathway, showing how it can account for basic phenomena of grasping. Our basic view is that AIP cells encode (by a population code whose details are beyond the present discussion) “affordances” for grasping from the visual stream and sends (neural codes for) these on to area F5. A*ffordances* (Gibson, 1979) are properties of the object relevant for action, in this case to grasping. In other words, vision here provides perceptual information on how to interact with an object, rather than categorizing the object or determining its identity.

The FARS model posits a crucial role for IT (inferotemporal cortex) and PFC (prefrontal cortex) in modulating F5's selection of an affordance. Here, the dorsal stream (from primary visual cortex to parietal cortex) carries among other things the information needed for AIP to detect that different parts of the object can be grasped in different ways, thus extracting affordances for the grasp system which (according to the FARS model) are then passed on to F5 where a selection must be made for the actual grasp. However, Figure 1 shows "FARS Modificato" in which PFC affects affordance selection in AIP rather than F5. This change is based on anatomical studies (Rizzolatti & Luppino, 2001, 2003). The point is that the dorsal stream does not know "what" the object is, it can only see the object as a set of possible affordances. The ventral stream (from primary visual cortex to

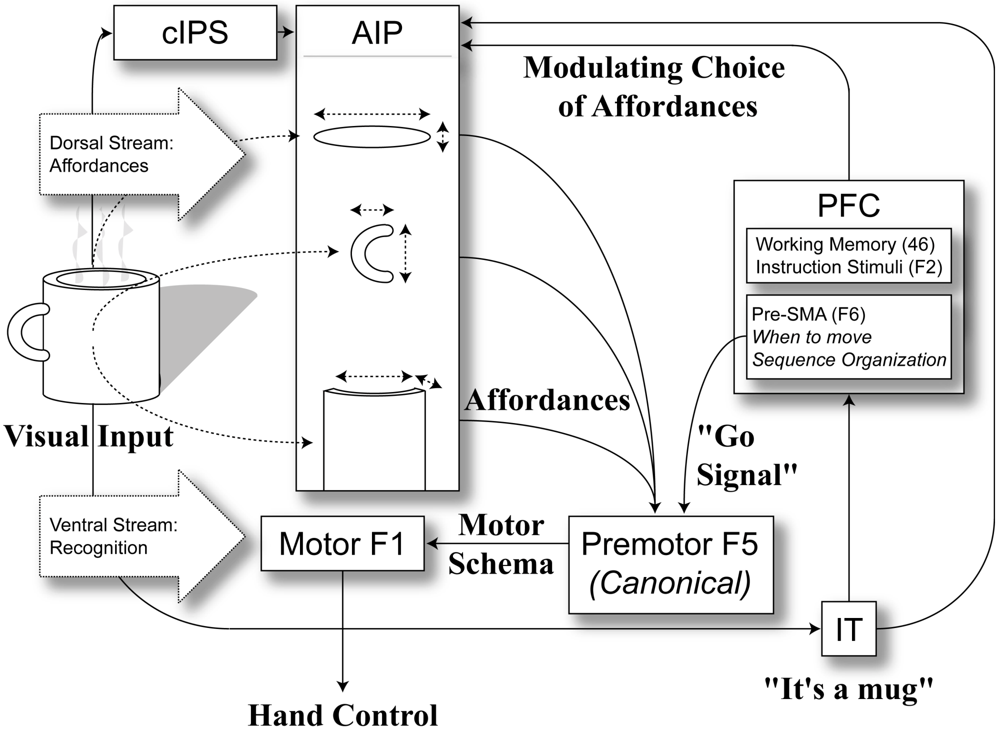

Figure 1. "FARS Modificato": The original FARS diagram (Fagg & Arbib, 1998) is here modified to show PFC acting on AIP rather than F5. The idea is that AIP does not "know" the identity of the object, but can only extract affordances (opportunities for grasping for the object consider as an unidentified solid); prefrontal cortex uses the IT identification of the object, in concert with task analysis and working memory, to help AIP select the appropriate action from its "menu".

inferotemporal cortex), by contrast, is able to recognize what the object is. This information is passed to prefrontal cortex which can then, on the basis of the current goals of the organism and recognition of the nature of the object, bias F5 to choose the affordance appropriate to the task at hand. In particular, the FARS model represents the way in which F5 may accept signals from areas F6 (pre-SMA), 46 (dorsolateral prefrontal cortex), and F2 (dorsal premotor cortex) to respond to task constraints, working memory, and instruction stimuli, respectively (see Fagg and Arbib 1988 for more details).

Further neurophysiological study of F5 revealed something unexpected – a class of F5 neurons that discharged not only when the monkey grasped or manipulated objects, but also when the monkey observed the experimenter make a gesture similar to the one that, when actively performed by the monkey, involved activity of the neuron. Neurons with this property are called "mirror neurons" (Gallese et al., 1996). Movements yielding mirror neuron activity when made by the experimenter include placing objects on or taking objects from a table, grasping food, or manipulating objects. Mirror neurons, in order to be visually triggered, require an interaction between the agent of the action and the object of it. The simple presentation of objects, even when held by hand, does not evoke mirror neuron discharge. Mirror neurons require a specific action – whether observed or self-executed – to be triggered. The majority of them respond selectively in relation to one type of action (e.g., grasping). This congruence can be extremely strict with, for example, the effective motor action (e.g., a precision grip) coinciding with the action that, when seen, triggers the neuron. For other neurons the congruence is broader. For them the motor requirement (e.g., precision grip) is usually stricter than the visual (any type of hand grasping, but not other actions). All mirror neurons show visual generalization. They fire when the instrument of the observed action (usually a hand) is large or small, far from or close to the monkey. They also fire even when the action instrument has shapes as different as those of a human or monkey hand. A few neurons respond even when the object is grasped by the mouth

However, not all F5 neurons respond to action observation. We thus distinguish mirror neurons, which are active both when the monkey performs certain actions and when the monkey observes them performed by others, from *canonical neurons* in F5. Canonical F5 neurons are active when the monkey observes an object and acts upon it, but not when the monkey observes actions performed by others. Mirror neurons receive input from the PF region of parietal cortex encoding observations of arm and hand movements. This is in contrast with the canonical F5 neurons that receive object-related input from AIP. It is the canonical neurons, with their input from AIP, that are modeled in the FARS model. In summary, the properties of mirror neurons suggest that area F5 is endowed with an observation/execution matching system: When the monkey observes a motor act that resembles one in its movement repertoire, a neural code for this action is automatically retrieved. This code consists in the activation of a subset, the mirror neurons, of the F5 neurons which discharge when the observed act is executed by the monkey itself.

Most analyses of the monkey have focused on the idea of a limited "hard-wired" repertoire of basic grasps, such as the precision pinch and the power grasp. However, in this article we emphasize that the child – and so, presumably, the monkey – must learn even the most basic grasps, as well as learn to detect the affordances for which they are appropriate. Thus, development entails cycles of perceiving and acting that engender new skills as children notice how the capabilities of the body relate to affordances of the environment. The basic capabilities are then extended through learning:

1) Developing a further set of useful grasps (extending the repertoire of actions for canonical F5 neurons);
2) Observing new affordances that match with the new grasps (extending the repertoire of AIP neurons);
3) Learning the relation between the self's grasping of an object and that of others grasping (linking F5 mirror neurons with the appropriate visual preprocessing and F5 canonical neurons to match the re-presentations of self-generated actions with similarly goal-oriented actions executed by others).

An interesting anecdote from the Rizzolatti laboratory (unpublished) is suggestive for further analysis: When a monkey first sees the experimenter grasp a raisin using a pair of pliers, his mirror neurons will not fire. However, after many such experiences, the monkey's mirror neurons encoding precision grip will fire when he sees the pliers used to grasp a raisin – the initially novel performance has been characterized as a familiar action.

The notion that a mirror system might exist in humans was tested by two brain imaging experiments (Rizzolatti et al., 1996b; Grafton, Arbib, Fadiga, & Rizzolatti, 1996). The two experiments differed in many aspects, but both compared brain activation when subjects observed the experimenter grasping a 3-D object against activation when subjects simply observed the object. Grasp observation significantly activated the superior temporal sulcus (STS), the inferior parietal lobule, and the inferior frontal gyrus (area 45). All activations were in the left hemisphere. The last area is of especial interest since areas 44 and 45 in the human left hemisphere constitute Broca's area, a major component of the human brain's language mechanisms. Although there is no dataset yet that shows the same activated voxels for grasping execution and grasping observation in Broca's area, such data certainly contribute to the growing body of indirect evidence that there is a mirror system for grasping in Broca's area.

Moreover, F5 in the monkey is generally considered (analysis by Massimo Matelli in R*i*zzolatti and Arbib 1998) to be the homologue of Broca's area in humans, i.e., it can be argued that these areas of monkey and human brain are related to the same region of the common ancestor. Thus, the cortical areas active during action recognition in humans and monkeys correspond very well. Taken together, human and monkey data indicate that in primates there is a fundamental system for action recognition: we argue that individuals recognize actions made by others because the neural pattern elicited in their premotor areas (in a broad sense) during action observation is similar to a part of that internally generated to produce that action. This system in humans is circumscribed to the left hemisphere. This provides the basis for the: **Mirror System Hypothesis** (Rizzolatti & Arbib, 1998; Arbib, 2002, 2005): The brain mechanisms crucial to human language in Broca's area evolved from our non-human primate ancestors' mirror system for grasping which provides the ability to link self-generated actions and the similar actions of others. The Mirror System Hypothesis offers a neural "missing link" for the view that manual gesture preceded speech in the evolution of human symbolic communication, and provides a foundation for the parity property of language, namely that what a message means to the sender will be, in general, approximated by what it means to the receiver.

Imitation and attention: affordances and effectivities

The ability to imitate has profound implications for learning and communication as well as playing a crucial role in attempts to build upon the mirror system hypothesis (Arbib, Billard, Iacoboni, & Oztop, 2000; Iacoboni et al., 1999). Our concern is to understand how imitation, especially assisted imitation, contributes to communicative development.

The empirical literature documents that monkeys do not imitate (Bard & Russell, 1999). Chimpanzees imitate the actions of others in the wild (Quiatt & Itani, 1994), but learn much more complex actions with objects when raised by humans (Tomasello, Savage-Rumbaugh, & Kruger, 1993) but the pace and extent of their imitation is very limited with respect to that of humans. Indeed, the vast majority of human children do imitate, albeit to varying degrees at different ages and for behaviors that differ in modality and complexity of content (Nadel & Butterworth, 1999**;** Eckerman, 1993). But such imitation requires the ability to map the body of the other onto the body of the self, and generate movements which in some sense correspond. If a child know that she is herself like the other (e.g., the caregiver) she may learn to do what the other does to achieve similar benefits or avoid risks. But can such an individual spontaneously or after a delay imitate just any "developmentally appropriate" behavior observed without assistance? We argue, probably not.

Assisted imitation may pave the way to language

Most research investigating the development and implications of imitation focuses on what the child knows, rather than how on the child comes to know. Accounting for these achievements usually takes the form of proposing some combination of cognitive precursors, socio-pragmatic knowledge, or maturing modules hypothesized to be necessary for the activity (Meltzoff & Moore, 1995, 1999; Tomasello, Kruger, & Ratner, 1993; Uzgiris, 1991, 1999). This literature documents the age at which the average child can observe someone else's action and repeat it accurately either promptly or after a delay. In our opinion, this body of research underestimates sources of the infants' accomplishments located in the caregiving environment. Informed by an integrative view of action and perception, we offer a somewhat different perspective that also suggests how imitation may foster the emergence of language.

Greenfield (1972) observed that children imitate those actions that are entering their repertoire. Why might these particular actions be ripe for imitation and not others? Are the children's imitations usually autonomous accomplishments or do they have a robust history of assistance from others? In answer, we provide evidence that caregivers invite infants to imitate. On those occasions, caregivers both direct attention (Adamson & Bakeman, 1984; Tomasello, 1988; Zukow-Goldring, 1989, 1990, 1997) to aspects of the ongoing events and tutor actions to "achieve consensus" (Zukow-Goldring, 1996, 2001). These interactional opportunities give infants crucial practice in (and a refining of) what to notice and do, and when to do it. Further, when demonstrating an activity, the caregiver marks the child's subsequent suitable attempts to imitate with speech and gestures of approval or may elaborate the ongoing activity, whereas repeated and revised messages, dropping the current activity, or remarking on the child's lack of interest follow inadequate responses. These interactions also may be central to communicative development. In particular, engaging in these activities may provide the means to grasp important prerequisites that underlie communicating with language. These basics include knowing that words have an instrumental effect on the receiver of a message (Braunwald, 1978; Braunwald & Brislin, 1979), that words refer (Bates, 1976; Schlesinger, 1982, Zukow-Goldring, 1997, Zukow-Goldring & Rader, 2001), and that coparticipants share or negotiate a common understanding of ongoing events (Macbeth, 1994; Moerman, 1988; Zukow-Goldring, 1990, 1997).

Our normal experience is highly multi-sensory, not restricted to the limited perceptual input of, say, a video clip. Indeed, Stoffregen and Bardy (2001) have argued that "multisensory perception is not merely the primary type of perception; it's the only type of perception". In fact, caregivers and children detect “the something that something is happening to” as well as "the something that is happening" through vision, touch, sound, taste, and touch (Michaels & Carello, 1981; Zukow-Goldring, 1997). Especially relevant to this idea is the young infant's known ability to detect regularities or invariants in the continuous stream of perceptual information (Bahrick & Pickens, 1994).

We have already mentioned J. J. Gibson's notion of *affordances*, and embrace his proposal that creatures detect the perceptual structure that specifies the unchanging invariant aspects of ongoing events as well as the structure specifying transformation and change. However, when we don our "neuroscience hats", we must note that what Gibson calls *direct perception* involves subtle processing of retinal signals by specific brain mechanisms that make the various invariants available for further processing as a basis for action and perception. Gibson claimed that people act to perceive and perceive to act, and much of our work in brain theory (e.g., Arbib, 1989) has stressed action-oriented perception and the action-perception cycle (see also Neisser 1976). The classic example is that as we walk across a room we see more and that more that we see tells us which surfaces will support our walking, what objects block our way and so on. We now add the notion of *effectivities* to the notion of affordances (Shaw & Turvey, 1981; Turvey, Shaw, Reed, & Mace, 1981): acting itself is informed by what the body can do (effectivities expand as an individual gains skill participating in new activities) as well as by what the environment affords for action.

However, the child at first lacks the ability to detect by observation alone how the body relates to the physical layout and to the furniture of the world, except for the most rudimentary actions. In this regard, our empirical studies of communicative development (Zukow-Goldring, 1996, 2001) have stressed, in addition, that during interaction, the perceiving and acting of one person continuously affects the perceiving and acting of the other. We thus propose that caregiver practices guide the infant to perceive possibilities for action (Zukow-Goldring, 1997, 2001; Zukow-Goldring & Ferko, 1994). Our point is not to deny that children can learn certain things for themselves by trial-and-error. Clearly the physical environment or layout affects us and we surely affect it continuously, but we stress too the mutuality of the person and the social environment. That is, by directing the child's attention to its own effectivities in relation to affordances in the environment, the caregiver greatly narrows the search space for learning, and consequently enhances the speed and extent of learning. Further, these caregiver practices or methods may educate infants to notice that the infant is "like the other" through interactions that explicitly foreground the correspondence between the effectivities of the infant-actor's body and that of the caregiver. In any novice-expert interaction, whether infant and caregiver or student and teacher, the perceiving and acting of one person continuously informs that of other interactional partners.

Educating attention: from being a body to becoming a cultural being "like the other"

What do infants have to learn about the world in order to communicate about what’s happening? Infants must learn the most basic things (even about), e.g., taking a bath, eating with utensils, walking. During mundane activities, infants must detect and participate in assembling the structure and organization of everyday events before they can communicate with others about these events. Out of the unceasing perceptual flow, which is quite unlike the highly edited cuts of most movies, *caregivers* continuously educate attention to aspects of ongoing events. This assistance guides infants to notice key elements of what persists and what changes. Caregiver gestures make perceptual structure prominent through translational movements that often occlude other information in a scene. In the same vein, placing objects close to the child’s face ensures attention and the inescapability of details (Zukow-Goldring, 1997).

Caregivers embody or put infants through the motions of activities as well as direct attention to the similarity of the one's own body to those of others, to the relation of the body to specific objects and animate beings, and to what these objects and animate beings afford for action and interaction. In contrast, many studies and theories assume that children know and/or learn autonomously how their bodies move in space and in relation to animate and inanimate things (Piaget, 1962, Thelen & Smith, 1994), and thus do not explore what experiences might underlie eventual adept performance. We thus stress again the role of the caregiver in directing attention to *effectivities* as well as *affordances* – the two sides of the mirror system. These interactive sequences eventually invite imitation.

Caregivers talk about what they are doing as they do it. Often children initially misunderstand these actions and spoken gestures, in part, because words cannot explain unless the child already knows what the words means. Given these circumstances, how is consensus achieved as the child becomes an adept member of the community? Our approach integrates perception, the building of action, and the meaning of words, despite the fact that many studies of language acquisition assume that gestures entail ambiguity of reference (Markman, 1989; Schlesinger, 1982). These authors rely on Quine's classic essay (1960) in which he discussed the ambiguity of reference entailed in, say, speaking about and pointing to a rabbit. But caregivers tend to focus attention with precision. They do not simply say an unfamiliar word (such as Quine's *gavagai)* while pointing. Instead, caregivers may rub a rabbit's fur while saying, “fur”; trace the topography of its ears while saying, “ear”, stroke the entire rabbit or rotate the whole animal when saying, “rabbit”, etc. (Zukow-Goldring, 1990, 1996; Reed, 1993). Successful teaching entails marking the correspondence between what is said and what is happening.

In what follows, we shall illustrate the findings from a number of studies of infant development with some qualitative examples (Zukow-Goldring, 1996, 1997, 2001). The final section will use our modeling of the development of grasping and the grasp-related mirror neuron to advance our understanding on how the mirror system grounds imitation as a core component of communicative and linguistic development within an action/perception framework.

The naturalistic experiments

Infants are immersed in talk: some directed to them, some to others, some to prohibit action, some to direct attention to something new or when the child does not understand an utterance. We argue that at early stages of communicative development, learning that words mean and what they mean entails having an embodied understanding of the organization and structure of relevant aspects of daily life. Concomitantly, the child must notice that others mark the relation between what is said and what is happening, and how they do so. To support this view, we summarize key results from a series of naturalistic experiments we conducted to clarify how children come to comprehend initially misunderstood messages (Zukow-Goldring, 1996, 2001). In these studies we tested the following hypotheses:

1. Providing a child with more perceptual structure will assist caregiver and child to achieve the consensus needed for communication, including, where appropriate, explicit guidance of the child's movements.

2. Additional or more specific verbal information will not enhance understanding when no basis for that understanding has yet been embodied.

The studies reported illuminate how a human child learns about the world. Of course, we are not denying the utility of verbal instruction for older children. Rather, our purpose is to illustrate how the fundamental link between perception and action provides the information upon which communication can build. It is a separate study to understand the later "bootstrapping" that occurs when words can take a far greater role in advancing what the child knows.

Method

Subjects: Five Euro-American families and six Latino families with an infant of 6 months were followed monthly through the one-word period.

Data: We collected twenty-minute monthly videos of naturalistic interaction at home, field notes, diaries and check lists of lexical development, as well as interviews following each video-taping session to ascertain the caregiver's interpretation of ongoing events and of the infant’s utterances.

We selected situations in which caregivers directed infants to notice specific elements, relations, or events over the myriad other possibilities available. This collection of attention-directing interactions included all instances of perceptual imperatives expressed by caregivers, such as *look!/¡mira!, listen!/¡oye!,* and so on, as well as the accompanying gestures, and the gestures alone as well as the infants' subsequent actions. Zukow-Goldring noticed this set of perceptual imperatives when doing field-work in Mexico (1981-1982). They occurred massively. Caregivers constantly said, *¡Mira!/Look!.* While we do not claim there is a *Mira neuron,* the use of perceptual imperatives may help draw the child's attention to the gestures that co-occur with them. These gestures may, in turn, provide specific support for imitation by directing attention to the perceptual information (in touch, smell, taste, vision, movement, and hearing) that lays the groundwork for knowing that the self is like others. Thus, this ability opens up the possibility of learning the actions that others display. A key issue then is to determine the relative importance for the young child of verbal messages versus gestures that educate attention and action. The distinction here is between words that provide explicit instructions, such as *Peel the orange!,* and the pairing of perceptual imperatives like *¡mira!* and gestures that direct the child to attend to the actions of the caregiver. Our conclusion will be that explicit verbal instructions are ineffective in the early stages. During these early stages children notice those affordances and effectivities that provide the referents for the initial development of their lexicon.

Perceptual structure: targets of attention

Caregiver messages combine gestures with targets of attention throughout the prelinguistic and one-word periods. In messages caregivers express what persists and changes as events coalesce and disperse. The targets vary in semantic complexity. The first three levels of caregiver messages that we have observed roughly parallel the three levels of semantic development described by Greenfield and Smith (1976) and Zukow, Reilly, and Greenfield (1982). Level I includes occasions when the caregiver might present a nondynamic object or animate being in the child's line of sight, such as showing a toy or pointing to a person or animal while saying, *Look at the duck/Grandpa!.* Level II consists of messages expressing an object or animate being undergoing some change through action (*Look, push them*) or in which a state or attribute is asserted about someone or something (*Look at the little doll.*). For Level III caregivers talk about more complex relations involving location (patch of dirt), instruments (eating utensils, crayons), part/whole relations (hair/head), possession (my, your), quantity (several of something), class membership (another something), concatenations (stacking objects that share a common surface), nestings (embedding one thing in another), such as saying*, Look, put the bead in the pail!*. Messages at Level IV express the relation between events (*Look, take your ball to your brother, so he can put it away.*). These targets of attention embody structural and transformational invariants in the environment across space/time. Caregiver messages in both cultures communicate perceptual structure and/or semantic functions that are a step or two ahead of those expressed in infants' speech (Zukow-Goldring, 1997).

Attention-Directing Gestures: Infant-Caregiver

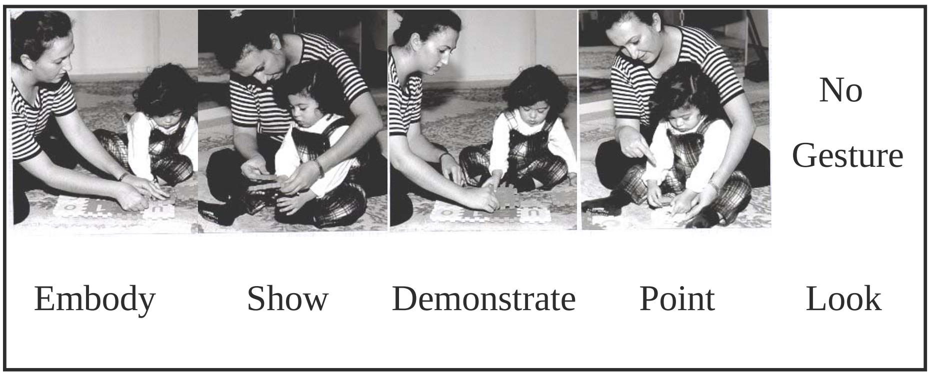

Figure 2. Attention-Directing Gestures

Five gestures that direct attention often accompany caregivers' verbal messages (Figure 2). These gestures encompass varying degrees of other- to self-regulation of attention to the effectivities of the body and the affordances of the environment.

Embody: a caregiver puts an infant through the motions of some activity (e.g., the caregiver takes the child’s hand, using it to press down a lever as he says, *¡por abajo!/ down!*). The child already "knows" the movement, but this is a new way to match or fit the effectivity of the hand/arm movement to the affordance of an object.

Show: caregivers regulate the infant's line of sight with a translational motion or perform some action using a familiar bodily effectivity to introduce a new possibility for action with an unfamiliar object or affordance. For instance, the caregiver looms an object toward the infant, saying, *gwow-wow/wow-wow* while looming a puppy or pushes a button on a new toy, saying, *¡empújalo!/push it!*.

Demonstration: an infant is invited to act/imitate through action and/or speech *(¿Y tu?* or *You do it!).* The infant who watches closely must detect or pick-up in the perceptual flow a familiar coupling of effectivity and affordance to be duplicated. For example, the caregiver may synchronize rhythmically retracting fingers of an upright palm with saying, *adiocito/bye-bye* when catching gaze and smiling. Alternately, an infant may be asked to pretend to avoid the sharp spines of a prickly-pear while the caregiver mimes approaching and pulling away from the fruit's surface, saying, *¡espinoso!/prickly!*.

Point: the infant must detect where a gesture's trajectory through space converges with some target of attention (the caregiver pointing to and saying, *p'acá/over here*).

Whereas adults take following a point for granted, infants must learn to trace its projected trajectory through time/space. As the infant develops, caregivers gradually increase the difficulty and shift the burden of tracing the trajectory from the caregiver to the infant. This work suggests a developmental sequence in which the caregivers of younger infants frequently *point-tap* a proximal target of attention with an index finger, thereby tracing the entire trajectory (point-tap a lever). Next caregivers often *point* to objects or animate beings close at hand. Learning to follow *distal points* to a target across the room or in a large, open area outside appears near the end of the second year. Caregivers may trace the entire trajectory from the corner of the child's eye across a busy environment (a kitchen full of people) to an indistinct target (clothing spattered with beaten egg white) or may throw a small object, such as a rock or seed pod, to trace the trajectory that eventually intersects the place where the target can be found. These preliminary results document that learning to follow a point is another ability that is not a solitary achievement, but one in which caregivers educate attention.

Look: no gestures accompany the caregiver's speech. Instead, only the caregiver's words and gaze direct the infant to correlate attention with that of the caregiver.

Our longitudinal data suggest that caregivers of less advanced infants (not necessarily younger infants) *embody* and *show* most frequently, shifting to *demonstrations, points,* and eventually *looks* as the infant develops (Zukow-Goldring, 1997).

Consensus: We determined whether or not caregivers treated their infants' responses to each message in a sequence as adequate or not. Action or speech indicating approval or embellishing the ongoing activity followed acceptable attempts, whereas inadequate responses were followed by repeated and revised messages, terminating the current activity, or noticing the child's lack of interest.

Perceptual Structure: We scored each message in a sequence after the initial one as providing or failing to offer additional perceptual structure. (For further detail, see Zukow-Goldring, 1996, 2001).

Linguistic Specificity: We assessed each sequence, message by message, for increases or decreases in linguistic specificity. Linguistic messages in ensuing turns can contain more or less explicit expression of nouns and verbs and previously ellipted lexical items. For instance, when adding linguistic specificity, the caregiver might say, *Peel it!* and then *You peel the orange!* Caregivers also can express less specificity in subsequent messages, e.g., *Peel it!* followed by *Do it!.*

Qualitative examples: assisted imitation

The caregiver demonstrates the action(s) and then gives the infant a chance to act. Quite often the infant's attempt is inadequate. Caregivers take care to arrange the physical layout, so that the configuration in space of caregiver, infant, and object(s) makes them suitably aligned in space so that action is within reach. In addition, optimal proximity makes perceptually prominent what the object or some aspect of it affords for action. Frequently the caregiver embodies the infant, so that the child can perceive the relation of his or her body in terms of posture, motor actions, rhythm as it changes over time to accomplish the action or action sequence.

In the following examples, the infants' unsuccessful attempts at "delayed" imitation with toys and a food item display some familiarity with the culturally relevant use of these objects. The three examples below focus mainly on: (1) tutoring effectivities and affordances to learn to concatenate objects (Pop Beads); (2) learning a sequence of actions to consummate an activity (Vibrating Toy); and (3) illustrating that more explicit verbal messages do not usually disambiguate initially misunderstood messages at this early level of lexical development; rather, providing more perceptual structure allows mother and daughter to achieve consensus (Orange Peeling). Even though these objects and their uses are not entirely novel, what is required to imitate apparently is. That is, the ability to notice the relevant affordances and coordinate them with particular effectivities that are necessary to accomplish these tasks is not available to the infant without assistance. Their fragmentary, flawed attempts to imitate actions observed in the past elicit very careful and elaborate tutoring on the part of the caregivers to direct attention to relevant affordances and effectivities. Going further, we need to understand how picking-up the perceptual information that the caregiver is completing an action can provide the basis for detecting the affordances that will guide children in their attempts to imitate that action.

Pop beads (13 months): caregiver tutoring of effectivities and affordances when concatenating beads

Pop beads, easily graspable by infants and toddlers, have affordances that allow concatenation. Play with this toy consists of (a) orienting toward each other the parts of each bead that afford concatenation (the dual complements of protrusion and opening), (ii) moving the appropriately oriented beads toward each other on a converging path, and (iii) applying enough force when the parts meet to embed the protrusion of one in the opening of the other.


PB 1 PB 2 PB 3
PB 4 PB 5 PB 6
PB 7 PB 8 PB 9
PB 10 PB 11 PB 12

Figure 3. Pop Beads.

The infant, Angela, begins by pressing a block lacking the appropriate affordances and a pop bead together. She displays an understanding that completing the task requires two small graspable objects and the application of some force to bring them together (Figure 3, PB1). Her behavior provides no evidence that she knows that a set of objects with specific parts must come

Cogniţie, Creier, Comportament / Cognition, Brain, Behavior
Vol. IX(3), 193-210, 2005
©Romanian Association of Cognitive Sciences

together, nor that they must sustain an orientation as they meet along a converging path. Her mother, Cecilia, provides perceptual information to Angela, gradually foregrounding the affordances of the objects and the effectivities of the body required to put the beads together. At first, Cecilia provides a bit of both, *point-touching* the opening on one bead (but not the protrusion in the other) as she directs attention to an affordance (PB2) and then reorients another bead (PB3), enacting a movement that aligns the beads on the same converging path.

Lacking at this point, however, is information displaying the path itself or the force required to push the converging objects together. After an unsuccessful attempt by Angela, Cecilia *shows* her what the body must do to move the beads along a path with the required orientation as she pushes the protrusion into the opening with appropriate force (PB4). The infant remains unable to put the beads together (PB5) The mother then more elaborately provides perceptual information as she slowly *point-touches* both the affordances of protrusion (PB6) and opening (PB7), followed by *demonstrating* the effectivities required for connecting and then disconnecting of the beads (PB8). As the infant watches more intently, Cecilia eventually invites Angela to imitate, "¿A ver, tu?/Let's see, you (do it)?". She assists her daughter's imitation by *partially demonstrating* what to do by orienting a bead opening toward Angela, making the protrusion easy to see on the second, and holding it in a fixed position (affordances for action) (PB9). Angela moves the bead along the appropriate path (PB10), but she misorients her bead as the beads touch one another (PB11). Cecilia reorients hers to place it back on a converging path. The realignment makes prominent just where Cecilia should push in her bead. As the infant pushes in the protruding end of her bead, the mother pushes from the opposite direction with enough force to link the beads (PB12).

In this case, the caregiver's gestures gradually provided increments in perceptual information that guided the infant to concatenate two objects. Eventually, the caregiver simplified the task by holding an appropriately oriented bead as a fixed target. This assistance allowed the child to bring her slightly misoriented bead (i) along a path toward her mother's (ii). Angela pushed her bead against the other (ii), while her mother subtly reoriented her bead (i) and provided a complimentary push. Notwithstanding Angela's noteworthy improvement on this occasion, bringing together two hands, each grasping a properly oriented object, was not within her "reach". Nor could she by herself apply enough force to connect her beads. It is possible that embodying Angela, putting her through the motions, might have drawn maximal attention to the coordination of the affordances of the beads and the effectivities of the body required to consummate this activity. The point is that the young child cannot simply observe a complex activity and imitate it. Much tutoring is required to build a repertoire on which “true” imitation (if such exists!) can take place.

Vibrating toy (14.5 months) - caregiver and "toy" tutoring of a sequence of actions

This infant, Elsa, and her mother, Kathy, engage in a familiar routine with a reindeer toy that has a hidden affordance, a spring inside the toy to which a string is attached. Family members had played this "game" with Elsa quite frequently during the prior eleven months. However, she had never attempted to imitate the others. In this routine, when the caregiver pulls on a ring that protrudes from the back of the toy, the string unwinds. Releasing the ring/string at the apex of its extension retracts the string so quickly that the toy vibrates strongly accompanied by a loud pulsing noise. Elsa expresses delight when she feels the vibrating toy placed on her stomach. Elsa, however, cannot make the toy vibrate by herself. This "game" entails a sequence of actions: (i) someone grasps the string by the ring, (ii) the string unwinds as it is pulled and (iii) retracts within the toy as the tension on the string lessens. Finally, (iv) someone places the reindeer toy on the infant's stomach. In this example, we place more emphasis on the sequence of actions than on the briefly noted affordances and effectivities that mother and toy make perceptually available. For conciseness sake, we have abbreviated this sequence, omitting repetitions and variations leading to the child's final adept enactment of this activity.

Elsa, sitting in front of her mother, turns to give her mother, Kathy, the toy that she wants her to animate (Figure 4, R1).

Kathy pulls the string out by the ring (R2), releases the string and places the vibrating toy on Elsa's stomach (R3). When Elsa wants her mother to continue, Kathy says, *You do it!,* as she *partially demonstrates* by orienting the back of the toy toward Elsa, making the ring for pulling (affordance) prominent and within the infant's reach. The infant grasps the ring (R4). The mother *embodies* by holding Elsa's hand and the toy steady as she pulls the toy away from them, presumably so the infant can feel the tension (affordance) as the string unwinds (R5). The spring attached to the string *embodies* Elsa by pulling her hand back toward the toy (R6). Subsequently, the infant pulls the string out herself (R7) and holds on as the spring *embodies* by retracting the string back into the toy (R8). At this point, the toy vibrates weakly, if at all. A few seconds later, Elsa pulls the string so quickly and fully out, that the tension on the string *embodies* her by snapping the ring from her fingers. This adept use of her body when pulling the string forcefully (effectivity) allows her quite serendipitously to experience how to take advantage of the vibratory properties of the toy (affordance). Note that her arm recoils from the force moving quickly away from its former position, while the hand holding the strongly vibrating toy moves far in the opposite direction (R9). Within seconds, Elsa first pulls and lets go of the string and then places the base of the toy on her stomach to best perceive the vibrations as she expresses evident joy (R10).

Note the free building up of a sequence, as the child understands each new element. Both caregiver and toy educate Elsa's attention to new affordances and the refining of her actions (effectivities) to fill in the gap between grasping the ring and feeling the vibration of the toy on her stomach. Elsa experiences bodily the tension of the string unwinding as her mother pulls the toy away from her and as the spring hidden in the toy that controls the string retracts as the string appears and disappears. Pulling the string out is within Elsa's grasp. However, the accidental snapping back (letting go) of the string at the apex of its path and at its highest

R 1 R 2 R 3

R 4 R 5 R 6

R 7 R 8 R 9

Figure 4. Vibrating Toy

tension made evident the relation between the effectivity of releasing the string and the ensuing affordance of the toy's vibrations. By the second attempt, Elsa had placed the toy to bring the most enjoyment. Within the next several minutes, she changed her grasp from a finger crooked through the ring to a pincer grip, could pull the string with both right and left hands, and attempted to give her doll the same experience. Although she knew "that" the toy held potential for vibrating, she did not know "how" to make it happen until she received very careful tutoring. This educating of attention and action contrasts sharply with the effort it takes to tutor the attention and action of monkeys as well as autistic children (Fogassi, personal communication; Nadel, Guérini, Pezé, & Rivet, 1999).

Orange peeling (16 months) - caregiver tutoring "when actions speak louder than words"

Peeling an orange with the hand entails penetrating the peel (both zest and pith), grasping the pulled away edge in a pincer grip, pulling the peel away from the flesh, and separating that portion of the peel from the fruit. Tearing off a piece of peel may involve yanking it away or, if the zest is tough, rotating the wrist at a 90 degree angle from the fruit. Thus, what to an adult seems like a single action must for the child be learned as a sequence of grasps and directed actions.

In the first vignette, the infant, who is quite fond of oranges, has seen them peeled and wants to do so herself but cannot. She scratches quite ineffectually at the surface of the peel, but does not know how to remove it. Angela needs her mother's help to learn the actions that go into peeling an orange. However, increasingly explicit verbal messages (*Let's see. Peel it! You peel! Take it off like this, look!*) coordinated with a variety of points, ranging from a subtle head point to proximal points of hand and index finger, do not communicate just how to peel an orange. In contrast, *embodying* the child provides the missing perceptual information (Figure 5A, Vignette 1) regarding how effectivities of her body and the affordances of the orange continuously inform one another. The mother's hands shadow and guide those of the child as together they each support the orange with one hand and pull off the peel with the other together. *Embodying* the infant provides her with perceptual information in vision, touch, and movement that she is like her mother (the other). (For more detail, see Zukow-Goldring, 2001.).

In the second vignette (Figure 5B), Cecilia *partially demonstrates* by pulling the peel nearly free of the orange. Angela easily removes that bit of peeling.

In vignette three (Figure 5C), Cecilia ups the ante by lifting the peel just a small distance away from the flesh. In the third frame below the mother mimes what she says, *Duro duro duro! Fuerte fuerte fuerte!/Hard hard hard! Forcefully forcefully forcefully!* However, she provides perceptual information for both manner of action (hard, forcefully) and trajectory (a change in direction of about ninety degrees) in gesture, but only for manner in speech. Note that trajectory is the information that helps solve the problem, illustrating when and that gestures "speak" louder than words (Kendon, 2002). Although Angela changes the direction of her pulling about 45 degrees (compare frames 2 and 5), Cecilia helps by *showing* how to pull away the peel at a more extreme angle.

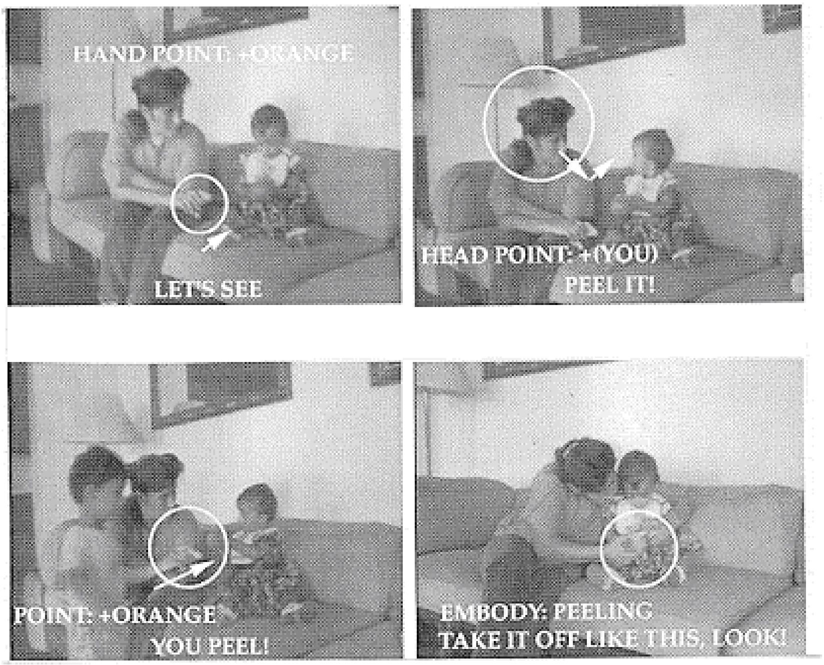

Figure 5A. Orange Peeling, Vignette 1: Hand, head, and finger points and increasingly specific verbal messages do not convey just how child and orange engage in "you peel it". Cecilia embodies her child to provide the missing perceptual information.

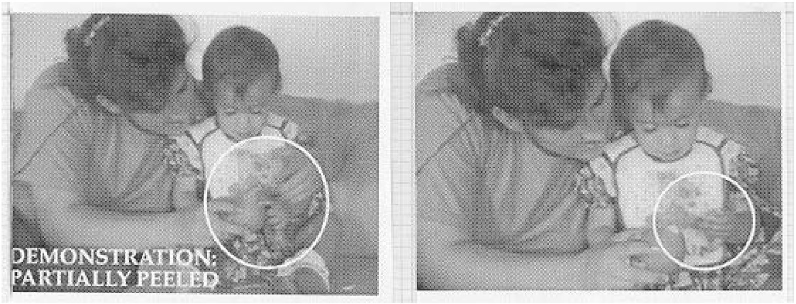

Figure 5B. Orange Peeling, Vignette 2: Cecilia partially demonstrates by lifting the peel almost entirely free of the flesh. She invites Angela to join in, saying *A ver, estírale/Let's see, pull it*. Now, Angela knows how.

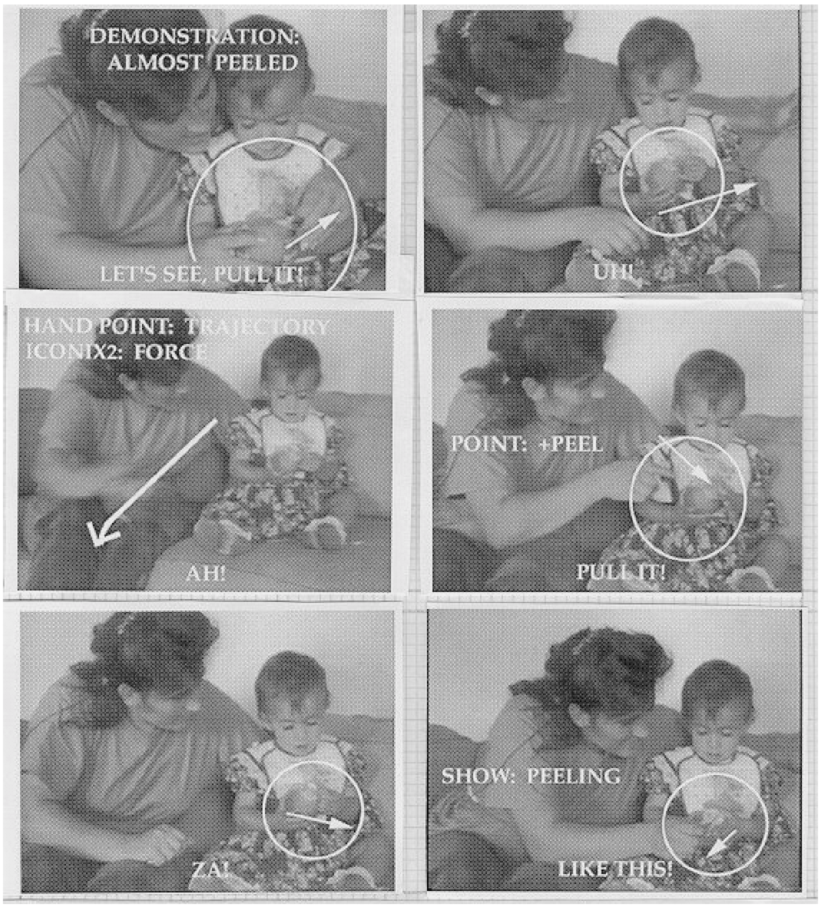

Figure 5C. Orange Peeling, Vignette 3: Cecilia makes the task more difficult, peeling only part way through the pith. Angela cannot tear through the zest. Cecilia mimes "strong and hard" *as well as* change in direction. Angela imitates, but still needs help.

We have argued that ambiguity is the rule, not the exception. Words cannot explain unless a person already knows what words mean. To understand what words mean a person must understand what is happening. Novices do not automatically understand the organization and structure of daily events. Experts show them. These qualitative examples emphasize the finding that caregivers work hard to cultivate their infants' understanding of ongoing events by providing perceptual information to disambiguate their initially misunderstood messages. When there is lack of consensus, providing more perceptual structure tends to resolve misunderstanding, whereas adding specificity to verbal messages did not reduce ambiguity. Perceptual restructuring of messages following communicative breakdowns led to achieving a common understanding whether in Spanish or in English.

In sum, caregivers establish an understanding of what is happening. They provide and direct attention to perceptual structure that makes prominent the relations among animate beings, objects and their actions. These dynamic relations specify the organization and structure of the most mundane daily activities. Caregivers introduce their infants to new effectivities or bodily capabilities and affordances for action and interaction on a daily basis. They assist them to link sequences of actions that comprise more and more complex activities. Caregivers also set aside language training when communication breaks down and, instead, focus on providing the perceptual information that will lead to a consensus. As caregivers educate attention, infants gradually learn to perceive, act, and know in culturally relevant ways.

Modeling development

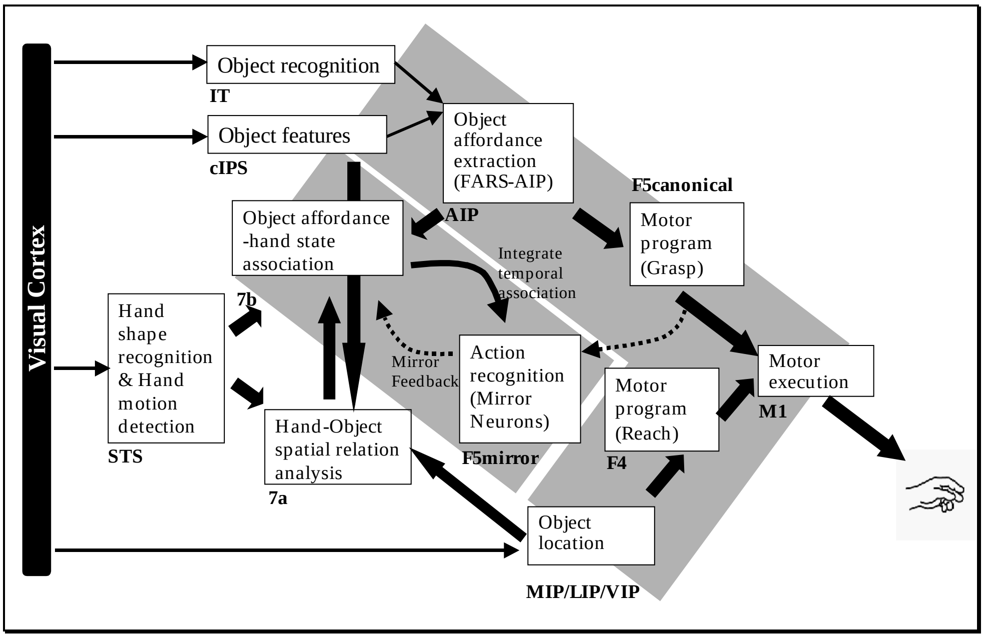

Figure 6. The Mirror Neuron System (MNS) Model, which extends the earlier FARS model. (Oztop and Arbib, 2002)

We have already mentioned the FARS model of the canonical system (Fagg & Arbib, 1998; Figure 1) that shows the importance of object recognition (inferotemporal cortex) and "planning" (prefrontal cortex) in modulating the selection of affordances in the determination of action. Figure 6 provides a glimpse of the schemas (functions) involved in the monkey mirror system[1]. In this section we first briefly review the elements of Figure 6 and then indicate the stages of modeling required to place this system in a developmental perspective. This is followed by the briefest review of completed work on modeling fragments of these stages. This review grounds our analysis of the challenges for future modeling which unites these modeling efforts with analysis of the data reviewed above on a perception/action based approach to communicative development, and how these considerations may provide a developmental counterpart to the Mirror System Hypothesis (Rizzolatti and Arbib, 1998) on the evolution of brain mechanisms supportive of language.

First, we look at those elements involved when the monkey itself reaches for an object. Areas IT and cIPS provide visual input concerning the nature of the observed object and the position and orientation of the object's surfaces, respectively, to AIP. The job of AIP is then to extract the affordances the object offers for grasping. The upper diagonal in Figure 6 corresponds to the basic pathway AIP → F5canonical → M1 (primary motor cortex) of the FARS model, but we will not dwell here on the role of PFC in action selection that was so important in the FARS model. The lower right diagonal (VIP → F4) completes the "canonical" portion of the MNS model, since motor cortex must not only instruct the hand muscles how to grasp but also (via various intermediaries) the arm muscles how to reach, transporting the hand to the object. However, we stress here that MNS (and the ILGM model described below) are "non-neurophysiological" models in the sense that we do not formulate separate models for each of these brain regions, coupling them in a way which represents the anatomy of the brain, and decomposing them into neural networks whose structure matches the observed connectivity of neurons within that region of the given brain. Instead, we simulate several "schemas", which represent the functional equivalent of the aggregate of several such brain regions without matching their internal structure. Nonetheless, even these models have implications for neurophysiology since they do contain populations of simulated neurons which correspond well with actual neural populations of the monkey brain, even when no such "neural matching" need hold for the simulation that provides their input and output.

The rest of the diagram presents the core elements for the understanding of the mirror system. As we have seen, mirror neurons do not fire when the monkey sees the hand movement or the object in isolation – it is the sight of the hand moving appropriately to grasp or otherwise manipulate a seen (or recently seen; Umilta et al., 2001) object that is required for the mirror neurons attuned to the

---

[1] For our present purposes, it is not necessary to pursue the hypotheses on where different schemas are located in the brain. In this paper, it will be the development of function that is of concern, rather than where in the brain the neural changes occur that mediate that development. Thus, in what follows we will use the abbreviations for brain regions without further explanation. For some readers, the abbreviations may just be seen as bizarre labels for the different functions diagrammed in Figure 6. Those readers wanting to see the abbreviations spelled out, as well as a brief exposition of data related to the hypothesized linkage of functions to brain structures, are referred to Oztop and Arbib (2002) and Arbib & Bota (2003).

Cogniție, Creier, Comportament / Cognition, Brain, Behavior
Vol. IX(3), 193-210, 2005

given action to fire. This requires schemas for the recognition of both the shape of the hand and analysis of its motion (ascribed in the figure to STS), and for analysis of the relation of these hand parameters to the location and affordance of the object (7a and 7b in the figure). In the MNS model (Oztop and Arbib, 2002), the "hand state" was accordingly defined as a vector whose components represented the movement of the wrist relative to the location of the object and of the hand shape relative to the affordances of the object.

Learning in the mirror system

Oztop and Arbib (2002) showed that an artificial neural network corresponding to 7b → $F5_{mirror}$ could be trained to recognize the grasp type from the hand state trajectory, with correct classification often being achieved well before the hand reached the object. The situation simulated was that of training mirror neurons for grasps already in the repertoire of the simulated monkey. More precisely, we assume that the neural equivalent of a grasp being in the monkey's repertoire is that there is a pattern of activity in the F5 canonical neurons that commands that grasp. During training, then, the output of the F5canonical neurons, acting as a code for the grasp being executed by the monkey at that time, was used as the training signal for the $F5_{mirror}$ neurons. As a result of this training, the appropriate mirror neurons can fire in response to the hand state trajectory even when the trajectory is not accompanied by $F5_{canonical}$ firing – and thus the F5 mirror neurons are prepared to respond to hand state trajectories even when the hand is of the "other" rather than the "self".

In other words, the simulated F5 mirror system could learn to produce the neural code for a grasp even when the $F5_{canonical}$ neurons were silent, correctly classifying the encoding of the hand state trajectory (the object-hand relation encoded by 7a and the hand motion information encoded by STS) that it received after recoding by 7b, with this recoding being itself dependent on learning. This provides "action recognition" because the hand state is defined in such a way that the relevant data can be based on the movement of any hand, whether that of self or other, relative to the object. Of course, what makes the modeling worthwhile was that the trained network responded not only to the training set trajectory (the object-hand relation encoded by 7a and the hand motion information encoded by STS) that it received after recoding by 7b, with this recoding being itself dependent on learning), but also exhibited interesting responses to novel hand-object relationships. Despite the use of a non-physiological neural network, simulations with the model revealed a range of putative properties of mirror neurons that suggest new neurophysiological experiments. To close this discussion of MNS, we stress that although it was constructed as a model of the development of mirror neurons in the monkey, we believe that it serves equally well as a model of the development of mirror neurons in the human infant. A major theme of the future modeling that this article delineates, then, is to clarify which aspects of human development are generic for primates, and which are specific to the human repertoire. As we shall now see, further work is required to understand how it is, for example, that an infant appears to understand the hand actions of a parent preparing to feed him long before the infant can feed himself – the infant anticipates and opens his mouth before the spoon arrives.

Learning to grasp

The MNS model makes the crucial assumption that the grasps which the mirror system comes to recognize are already in the (monkey or human) infant's repertoire. But this raises the question of how grasps entered the repertoire. To simplify somewhat, there are two answers: (i) Children explore their environment and as their initially inept arm and hand movements successfully contact objects, they learn to reliably reproduce the successful grasps, with the repertoire being tuned through further experience. (ii) With more or less help from caregivers, infants come to recognize certain novel actions in terms of similarities and differences from movements already in their repertoires and on this basis learn to produce some version of these novel actions for themselves. In terms of Figure 6, we might say that if MNS were augmented to have a population of mirror neurons which could acquire population codes for observed actions not yet in the repertoire of self-actions, then in case (ii) the mirror neurons would provide training for the canonical neurons, reversing the information flow see in the MNSI model. We note that this raises the further possibility that the infant may come to recognize movements that are not only not within the repertoire but which never come to be within the repertoire. In this case, the cumulative development of action recognition may proceed to increase the breadth and subtlety of the range of actions that are recognizable but cannot be performed by children. These considerations will prove especially important for our further work on the phylogeny and ontology of language, and make clear why that work goes under the slogan "Beyond the Mirror" to emphasize that the functionality common to monkey F5 mirror neurons and human Broca's area will be a small part, no matter how crucial, of the final analysis. However, for the present paper, we limit ourselves to two further accounts,

(a) the description of two views of a completed models for autonomous grasp development, and
(b) a high-level view of prospects for modeling the learning of new actions with the assistance of a caregiver.

Our first view of the model of how the infant or monkey learns to grasp, ILGM, the Learning to Grasp Model (Oztop, Bradley and Arbib, 2004; see Oztop, Arbib and Bradley, 2005, for further exposition) takes seriously the claim that this basic stage is common to monkey and human infants, and uses as its database what is known about how human infants develop grasping skills. From the first, an infant learns its own possibilities for action in the environment, and the affordances of objects, through exploratory behavior. By 2-3 months, infants start exploring their bodies as they move in the environment, they babble and touch themselves and also start to stare at their hands (Bayley 1969). Infants progress from a crude ability of reaching at birth to finer reaching and further grasping ability around four months of age. Infants learn to overcome problems associated with reaching and grasping by interactive searching (von Hofsten 1993; Berthier, Clifton, Gullapalli, McCall, & Robin, 1996).The precision grasp appears around 12-18 months of age (Berthier, Clifton, McCall, & Robin, 1999). To grasp successfully infants have to learn how to control their arms and, further, to match the abilities of their limbs with affordances presented by the environment (Bernstein 1967; Gibson, E. J., 1969; Gibson, J. J., 1988; Thelen 2000). At first, the poorly controlled arm, trunk and postural movements make it very difficult for the young infant to generate consistent feedback to form stable links between perceptual and motor schemas. However, Rochat and Morgan (1995) have shown that infants are aware of a variety of visual, proprioceptive and haptic consequences of their limb movements. The child's grasping can be affected by haptic cues by 4 months (Newell 1986), while infants as young as 5 months may abort their reaches if vision of the hand is removed (Lasky 1977). All this yields a well established set of grasps, including the precision grip, with preshaping to visual affordances by 12-18 months of age (Berthier et al. 1999).

ILGM models the discovery of grasps that match the affordances presented by the objects in the environment. The advent of voluntary grasping of objects is preceded by several weeks in which the infant engages in arm movements and fisted swipes in the presence of visible objects (von Hofsten 1984). An infant, once contacting an object, will occasionally try to grasp it (Clifton, Muir, Ashmead, & Clarkson, 1993). Normal infants very quickly (in a few weeks) acquire the ability to habituate and sculpt hand movements in the presence of a palmar stimulus – cutaneous reflexes initially aid or increase the probability in securing grasp of an object and then come to selectively use them to assist their movements. By around seven moths the infant is able to stabilize the grasp (Clifton et al. 1993). However, infants do not readily demonstrate control over fractionated finger movements before the end of the first year, even though fractionated finger movements may occur spontaneously and much earlier in the noise of random movements – we may speak of "motor babbling", comparing it to the infant's (vocal) babbling which contains many sounds that will later emerge in purposeful language. Adults preshape the hand during hand transport, e.g., adjusting the distance between the thumb and other fingers according to the size of the object. In contrast, before nine months of age, infants adjust their grasps after touching the object, lacking anticipation of the orientation and size of the object (Rosenbaum 1991). This holds even though infants younger than nine months old are physically able to vary their grip size, for they can spread their fingers farther apart once they have felt a large object (von Hofsten and Ronnqvist 1988). Newell (1986) identified rudimentary hand shaping after contact starting at 4 to 6 months, whereas 7 to 8 month olds did not appear to need contact to initiate shaping. Von Hofsten and Ronnqvist (1988) found that children would start shaping the hand midreach by 9-13 mos. It appears that in early infancy the fractionated control of fingers is mainly driven by somatosensory feedback. Newell (1986) find that the older infants' visually programmed and younger infants' haptically adjusted grasp configurations are very similar. This strongly suggests that the earlier haptic grasping phase serves to train the visual grasp planning circuits in the infant brain.

Such data constrained the design of ILGM, and enabled us to evaluate its relevance to infant learning through explicit comparisons. The model interacts with its environment (plans and executes grasp actions) and observes the consequences of its actions (grasp feedback) and modifies its internal parameters (corresponding to neural connections) in such a way that certain patterns (grasp plans) are selected and refined amongst many other possibilities. The Learning to Grasp Model (ILGM) models development in two stages. The first stage is the period when infants are unable to incorporate object affordance into grasp plans while the second phase is when infants start incorporating object information into grasps. ILGM has been analyzed via simulation experiments that predict behavioral responses which allow us to make comparisons where experimental data is available. When no data are available, we produce useful predictions that can be experimentally tested.

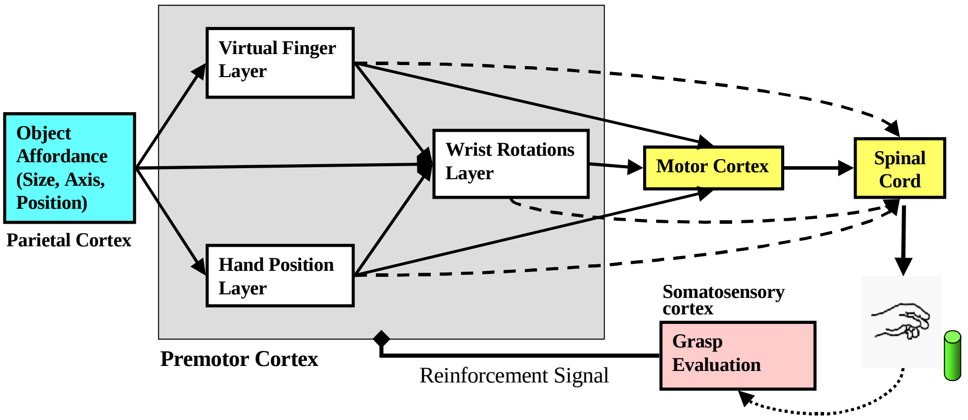

Figure 7. The structure of the Infant Learning to Grasp Model. The individual layers are trained based on somatosensory feedback.

The details of the model are beyond the scope of the present review but with reference to Figure 7 we note that the crucial element of the model is the built-in grasp evaluation mechanism which uses kinesthetic information to evaluate that the extent to which a successful contact with the object is made as a result of the infant's initially more or less random contact with an object and consequent grasping. The "better" the resultant grasp, the greater the reinforcement signal that makes the premotor network more likely to generate the grasp again. As a result, through time the system becomes less likely to produce a random grasp and more likely to produce one from its repertoire of successful grasps. Five comments about this simulation:

- It requires not only the simulation of neural networks for the "premotor cortex" shown in Figure 7 but also the simulation of an arm and hand so that the simulation can compute data about the contact between hand and object needed for grasp evaluation.
- ILGM models how the child may discover grasps through "motor babbling", having already acquired the ability to project the hand in the general direction of an object and make an uncoordinated swipe at the object. Learning is driven by "the joy of grasping", as signaled by the grasp evaluation, not by any explicit training signal.
- In this round of modeling, we have assumed that affordance information is already encoded in the brain. An immediate goal for further modeling is to understand how the activity of AIP might itself be shaped by experience, as it comes to recognize and encode those visual features of objects local to the part of the object where a successful grasp has taken place, with those visual features becoming cross-indexed to the kinesthetic features of the associated grasp.
- Even this takes us only as far as finding a stable grasp appropriate to the observed affordance of the object. This says nothing of the adaptive value of a grasp or the context (dependent on object and task) in which it will be used. This returns us to the loop via IT and PFC in the FARS model of Figure 1, showing how other parts of the brain complement the canonical and mirror neurons of F5 in placing a grasp in "semantic context".
- The ILGM model is non-neurophysiological, designed to explain the development of the infant's behavior, rather than analyze the changes in neural activity that might be observed in the monkey.

We close by noting that ILGM says nothing about the mirror system – rather it shows how the infant brain may acquire the basic repertoire of grasps that "gets the mirror neurons started" along the lines delineated in the MNS model (and its future, more neurophysiologically realistic, variants). Our task in the rest of this article is to give a prospectus for future modeling of the learning of new actions with the assistance of a caregiver, once the basic motor repertoire and mirror system for that repertoire are in place.

Imitation and attention: challenges for future modeling

Our Naturalistic Experiments delineated some of the ways in which imitation, especially assisted imitation, contributes to communicative development. From a neuroscience perspective, the present sections will summarize some of our observations in terms of the future modeling they suggest along with a neo-Gibsonian view of child development. (A challenge we have set ourselves is to unpackage aspects of our naturalistic and computational approaches that do not at first converge.) But first we note two major deficiencies in the MNS model overview of Figure 6 and the related modeling which we have just outlined:

(a) In each of our present models, the input to the model is already focused on the task at hand: visual input solely concerning the object for the FARS model; visual input solely concerning the object and the one hand for the MNS model; and visual input concerning object affordances plus somatosensory input related to the success of the grasp for the ILGM model. Clearly, models that address the naturalistic experiments must encompass a wider range of sensory data concerning a range of objects and information about the body of the child, the behavior of the caregiver, and interactions between child and caregiver. We must then model the attentional processes of the child, seeking to explain why certain caregiver behaviors are more effective than others in focusing the attention of the child on relevant affordances and effectivities. On the one hand, such new elements provide a major challenge for future modeling. On the other hand, they will pay off in greatly improved efficiency in learning, since the successful focusing of attention by the caregiver means that the child's "search space" is limited to the neighborhood of successful grasps and manipulations, rather than involving a time-consuming trial-and-error process that includes many configurations far removed from those required for successful completion of the task.

(b) Both neurophysiological studies of the mirror system and the above modeling get as far as recognizing the similarity of actions whether conducted by the self or another. As already noted, we seek to model not only how movements within the child's repertoire become recognizable even when performed by others, but also how the movements performed by others may become, not only recognizable but imitatable and thus added to the child's repertoire. However, this still ignores the question of agency: What information allows the child to know whose hand it is that acts? Going further, we need to understand how picking-up the perceptual information that specifies that the caregiver is completing an action can provide the basis for detecting the affordances that will guide children in their attempts to imitate that action. In any case, we reiterate that successful imitation requires attention to the pattern of regularities in the other's behavior (which involves the relation of the agent to other agents and to objects). This requires not only the concern with attention outlined in (a) but also recognition that "perceiving regularities" is always a function of what the child already knows. At any time, the child can recognize those actions that share regularities with ones in its repertoire. If the difference between the known action and the desired action is "perceivable" then the child can quickly detect the "new" affordance that will guide the ensuing movement, adding the new motion to its repertoire. Otherwise, a time-consuming process of trial-and-error is required to gradually shape a successful action.

(c) At a more technical level, the "biomechanical" model of the arm and hand must be refined to allow compliant motion, e.g., the motion of fingers conforming to the affordances of an object rather than being preprogrammed to match those affordances. Reliance on compliance is another important factor in reducing the search space.

(d) Similarly, the emphasis must shift from single actions to sequences of actions, where the shaping of action and the transition from one action to the next depends crucially on the response of objects in the environment to manipulation. In the vibrating toy, it is the toy that determines how far out the ring can be pulled, and the child learns that the resistance of the fully extended string is the signal to release the ring. This dimension of the proposed work will build upon and extend studies of the role of the supplementary motor area and basal ganglia in sequential behavior (e.g., Dominey, Arbib, & Joseph, 1995; Bischoff-Grethe, Crowley, & Arbib, 2003). However, most of the modeling here has focused on how a sequence may be learned as the result of many, many repetitions. This is certainly appropriate for analysis of certain aspects of the child's behavior and even adult skills may require much practice to be honed, though this process of tuning, as distinct from assemblage, may place more emphasis on the cerebellum than on the basal ganglia. However - and the vibrating toy is a case in point - there are cases where the process of learning is extremely rapid, and may be characterized as "sequence editing", with an unsuccessful element A of a sequence being replaced by a successful element B within a trial or two once that success is recognized.

Earlier, we countered Quine’s “gavagai” argument by noting that caregivers can direct attention to the desired referent - they may rub a rabbit's fur while saying “fur”; trace the topography of its ears while saying “ear”, and so on. Rather than outline here specific models for building language atop the basic structure of action recognition and imitation, we simply stress again the multimodality of perception. That is, crucial to this achievement is the child's ability to detect the higher-order perceptual regularities that mark the correspondence between the caregiver's utterance and the ongoing action to which he is attending (Zukow-Goldring, 1997; Zukow-Goldring & Rader, 2001). Evidence that suggests the evolutionary origins of this ability comes from Gallese (2001). He has found that mirror neurons can be activated by the sound that co-occurs with the action. For example, since there is no reason to believe that there is any a priori neural linkage between the sight and sound of, say, breaking a peanut, there is every reason to believe that detecting these regularities can be extended to sounds that co-occur with the action, as distinct from the sounds for the action itself. This sets the goal for extending the modeling of the mirror system from hand movements to speech gestures. Of course, as the mirror system hypothesis suggests, this requires many developments that extend the mirror system of the monkey-human ancestor to support not only imitation of hand movements but more general forms of pantomime, leading on to sequences of manual signs and then to protospeech as the control systems for protosign extend to the comprehension and control of intentional patterns of vocalization.

All this is in the context of understanding how the methods of the caregiver correspond to the expanding capabilities of the child. Often developmental researchers and scholars study affective, motor, perceptual and cognitive development separately. Caregivers do not. During the prelinguistic and one-word periods, caregivers prepare infants to imitate by assisting them "to see what to do" before they can "do what they see" others doing. Day-in and day-out, they cultivate imitation within mundane daily activities with gestures. They animate and direct their infants' attention to their own and others' bodily movements as well as making prominent what the environment offers for action. Thus, this investigation suggests that caregiver practices are crucial to the development of the important "like the other" phenomenon (cf. Meltzoff and Moore, 1999).

Humans who eventually learn/understand that the self is "like the other" cultivate abilities in their young that contribute to imitating, tutoring, communicating, and representing events. The mirror system offers a means to clarify in what manner human and nonhuman primates understand what they see other conspecifics and other primates doing, what abilities and perceptual information underlie learning to do what they see others do, and much more.

♣ This paper expands upon an earlier version whose abstract was prepared for discussion at the conference on "Perspectives On Imitation: From Cognitive Neuroscience to Social Science", 23-26 May 2002, Royaumont Abbey, France.

REFERENCES

Adamson, L., & Bakeman, R. (1984). Mothers' communicative acts: Changes during infancy. *Infant Behavior and Development, 7*, 467-478.

Arbib M. A. (1989). *The metaphorical brain 2 : Neural networks and beyond*. New York: Wiley.

Arbib, M. A., Billard, A., Iacoboni, M., & Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks, 13*, 975-997.

Arbib, M. A. (2002). The mirror system, imitation, and the evolution of language. In C. Nehaniv & K. Dautenhahn (Eds.), *Imitation in animals and artifacts.* Cambridge, MA: The MIT Press.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, in press.

Arbib, M. A., & Bota, M. (2003). Language evolution: Neural homologies and neuroinformatics. *Neural Networks,16,* 1237-1260.

Bahrick, L. E., & Pickens, J. N. (1994). Amodal relations: The basis for intermodal perception and learning in infancy. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bard, K. A., & Russell, C. L. (1999). Evolutionary foundations of imitation: Social, cognitive and developmental aspects of imitative processes in non-human primates. In J. Nadel, & G. Butterworth (Eds.), *Imitation in infancy: Progress and prospects of current research*. Cambridge, UK: Cambridge University Press.

Bates, E. (1976). *Language and context: The acquisition of pragmatics.* New York: Academic Press.

Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development.* New York: The Psychological Corporation.

Bernstein, N. A. (1967). *The coordination and regulation of movements*. Oxford: Pergamon Press.

Berthier, N. E., Clifton, R. K., Gullapalli, V., McCall, D. D., Robin, D. J. (1996). Visual information and object size in the control of reaching. *Journal of Motor Behavior 28*, 187-197

Berthier, N. E., Clifton, R. K., McCall, D. D., Robin, D. J. (1999). Proximodistal structure of early reaching in human infants. *Experimental Brain Research, 127*, 259-269.

Bischoff-Grethe, A., Crowley, M. G., & Arbib, M. A. (2003). Movement inhibition and next sensory state predictions in the basal ganglia. In A. M. Graybiel, M. R. Delong & S. T. Kitai, (Eds.), *The basal ganglia VI.* New York: Kluwer Academic/Plenum Publishers.

Braunwald, S. R., & Brislin, R. W. (1979). On being understood: The listener's contribution to the toddler's ability to communicate. In P. French (Ed.), *The development of meaning: Pedo-linguistic series.* Japan: Bunks Hyonron Press.

Braunwald, S. (1978). Context, word and meaning: Toward a communicational analysis of lexical acquisition. In A. Lock (Ed.), *Action, gesture, and symbol: The emergence of language*. London: Academic Press.

Clifton, R. K., Muir, D. W., Ashmead, D. H., Clarkson, M. G. (1993). Is visually guided reaching in early infancy a myth? *Child Development, 64*, 1099-1110.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*, 176-180.

Dominey, P. F., Arbib, M. A., & Joseph, J.-P. (1995). A model of corticostriatal plasticity for learning associations and sequences, *Journal of Cognitive Neuroscience*, *7,* 311-336.

Eckerman, C. O. (1993). Toddlers' achievement of coordinated action with conspecifics: A dynamic systems perspective. In L. B. Smith & E. Thelen (Eds.), *A dynamic systems approach to development* (pp. 333-357). Cambridge: MIT Press.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks, 11*, 1277-1303.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

Gallese, V. (November, 2001). Echo mirror neurons: Recognizing action by sound. Paper presented at the *Human Frontiers Science Program Workshop "Mirror System: Humans, Monkeys and Models"*. University of Southern California, Los Angeles, CA.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting and acquiring of knowledge. *Annual Review of Psychology, 39*, 1-41

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Bston: Houghton Mifflin.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Grafton, S. T., Arbib, M. A., Fadiga, L., Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography 2: Observation compared with imagination. *Experimental Brain Research, 112*, 103-111.

Greenfield, P. M., & Smith, J. (1976). *The structure of communication in early language development.* New York: Academic Press.

Greenfield, P. M. (1972). Cross-cultural studies of mother-infant interaction: Toward a structural-functional approach. *Human Development*, *15*, 131-138.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526-2528.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends in Neurosciences, 18,* 314-320.

Kendon, A. (2002). Historical observations on the relationship between research on sign languages and language origins theory. In D. Armstrong, M. Karchmer, & J. Van Cleeve (Eds.), *The study of sign languages. Essays in honor of William C. Stokoe.* Washington, D.C.: Gallaudet University Press.

Lasky, R. E. (1977). The effect of visual feedback of the hand on the reaching and retrieval behavior of young infants. *Child Development, 48,* 112-117.

Macbeth, D. (1994). Classroom encounters with the unspeakable: "Do you see, Danelle?". *Discourse Processes*, *17*, 311-335.

Mace, W. (1977). Ask now what's in your head, but what your head's inside of. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Erlbaum.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: The MIT Press.

Meltzoff, A. N., & Moore, M. K. (1995). Infants' understanding of people and things: From body imitation to folk psychology. In J. Bermúdez, A. J. Marcel, & N. Eilan (Eds.), *The body and the self* . Cambridge, MA: MIT Press.

Meltzoff, A. N., & Moore, M. K (1999). Persons and representation: Why infant imitation is important for theories of human development. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy*. Cambridge, UK: Cambridge University Press.

Michaels, C. F., & Carello, C. (1981). *Direct perception*. Englewood Cliffs, N. J.: Prentice-Hall.

Moerman, M. (1988). *Talking culture: Ethnography and conversation analysis*. Philadelphia: University of Pennsylvania Press.

Nadel, J., & Butterworth, G. (1999). *Imitation in infancy.* Cambridge: Cambridge University Press.

Nadel, J., Guérini, C., Pezé, A., & Rivet C. (1999). The evolving nature of imitation as a format for communication. *Imitation in infancy.* Cambridge: Cambridge University Press.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. W.H. Freeman.

Newell, K. M. (1986). Motor development in children: Aspects of coordination and control. In M. G. Wade & H. T. A. Whiting (Eds), *Motor development in children: Aspects of coordination and control*. Bston: Nijhoff.

Oztop, E., & Arbib, M.A. (2002). Schema design and implementation of the grasp-related mirror neuron system, *Biological Cybernetics, 87*, 116-140.

Oztop, E., Arbib, M. A., & Bradley, N. (2005). The development of grasping and the mirror system. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system.* Cambridge: Cambridge University Press.

Oztop, E., Bradley, N. S., & Arbib, M. A. (2004). Infant grasp learning: A computational model. *Experimental Brain Research, 158*, 480–503.

Piaget, J. (1962). *Play, dreams, and imitation in childhood*. New York: Norton Library.

Quiatt, D., & Itani, J. (1994). *Hominid culture in primate perspective.* Niwor, CO: University Press of Colorado.

Quine, W. V. O. (1960). *Word and object.* New York: Wiley.

Reed, E. S. (1993). The intention to use a specific affordance: A conceptual framework for psychology. In R. H. Wozniak & K. Fischer (Eds.), *Development in context: Acting and thinking in specific environments.* Hillsdale, NJ: Erlbaum.

Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences 21*, 188-194.

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

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

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F, (1996b). Localization of grasp representations in humans by PET .1. Observation versus execution. *Experimental Brain Research, 111,* 246-252.

Rizzolatti, G., & Luppino, G. (2003). Grasping movements: visuomotor transformations. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed). Cambridge, MA: The MIT Press.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system, *Neuron*, *31*, 889-901.

Rochat, P., & Morgan, R. (1995). Spatial determinants in the perception of self-produced leg movements by 3-5 month-old infants. *Developmental Psychology, 31*, 626-636.

Rosenbaum, D. A. (1991). *Human motor control*. San Diego: Academic Press.

Schlesinger, I. M. (1982). *Steps to language: Toward a theory of native language acquisition.* Hillsdale, NJ: Erlbaum.

Shaw, R., & Turvey, M. (1981). Coalitions as models of ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.

Stoffregen, T., & Bardy, B. (2001). On specification and the senses. *Behavioral and Brain Sciences*, *24*, 195-261.

Taira, M., Mine, S., Georgopoulos, A. P., Murata, A., & Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Experimental Brain Research, 83*, 29-36.

Thelen, E. (2000). Motor development as foundation and future of developmental psychology. *International Journal of Behavioral Development, 24*, 385-397.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language Sciences, 10*, 69-88.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16,* 495-552.

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. (1993). Imitative learning of actions on objects by children, chimpanzees and enculturated chimpanzees. *Child Development, 64,* 1688-1705.

Turvey, M., Shaw, R., Reed, E., & Mace, W. (1981). Ecological laws of perceiving and acting. *Cognition, 9,* 237-304.

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001) I know what you are doing: A neurophysiological study. *Neuron, 31*, 155-165.

Uzgiris, I. (1991). The social context of imitation. In M. Lewis & S. Feinman (Eds.), *Social influences and socialization in infancy*. New York: Plenum Press.

Uzgiris, I. (1999). Imitation as activity: Its developmental aspect. In J. Nadel & Butterworth (Eds.), *Imitation in infancy.* Cambridge: Cambridge University Press.

von Hofsten, C. (1984). Developmental changes in the organization of prereaching movements. *Developmental Psychology, 20*, 378-388.

von Hofsten C (1993). The structuring of neonatal arm movements. *Child Development, 64*, 1046-1057.

von Hofsten, C., & Ronnqvist, L. (1988). Preparation for grasping an object: A developmental study. *Journal of Experimental Psychology: Human Perception & Performance, 14*, 610-621.

von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology 18*, 450-461.

Zukow, P. G. (1989). Siblings as effective socializing agents: Evidence from Central Mexico. In P. G. Zukow (Ed.), *Sibling interactions across cultures: Theoretical and methodological issues*. New York: Springer-Verlag.

Zukow, P. G. (1990). Socio-perceptual bases for the emergence of language: An alternative to innatist approaches. In C. Dent, & P. G. Zukow (Eds.), *The idea of innateness: Effects on language and communication research. Developmental Psychobiology, 23*, 705 - 726.

Zukow, P. G., Reilly, J. S., & Greenfield, P. M. (1982). Making the absent present: Facilitating the transition from sensorimotor to linguistic communication. In K. E. Nelson (Ed.), *Children's language, Vol. 3*. Hillsdale, NJ: Erlbaum.

Zukow-Goldring, P., & Ferko, K. R. (1994). An ecological approach to the emergence of the lexicon: Socializing attention. In V. John-Steiner, C. Panofsky, & L. Smith (Eds.), *Sociocultural approaches to language and literacy: Interactionist perspective*. New York: Cambridge University Press.

Zukow-Goldring, P., & Rader, N. (2001). Perceiving referring actions: A commentary on Gogate, Walker-Andrews, & Bahrick. *Developmental Science.*

Zukow-Goldring, P. (1996). Sensitive caregivers foster the comprehension of speech: When gestures speak louder than words. *Early Development and Parenting, 5*, 195-211.

Zukow-Goldring, P. (1997). A social ecological realist approach to the emergence of the lexicon: Educating attention to amodal invariants in gesture and speech. In C. Dent-Read & P. Zukow-Goldring (Eds.), *Evolving explanations of development: Ecological approaches to organism-environment systems*. Washington, D. C.: American Psychological Association.

Zukow-Goldring, P. (2001). Perceiving referring actions: Latino and euro-american infants and caregivers comprehending speech. In K. L. Nelson, A. Aksu-Koc, & C. Johnson (Eds.), *Children's language, Vol. 11*. Hillsdale NJ: Erlbaum.

Zukow-Goldring, P., Romo, L., & Duncan, K. R. (1994). Gestures speak louder than words: Achieving consensus in latino classrooms. In A. Alvarez & P. del Rio (Eds.), *Education as cultural construction*. Madrid: Fondacion Infancia & Aprendizaje.