Connectionism in an Artificial Life Perspective:

Simulating Motor, Cognitive, and Language Development

Matthew Schlesinger

Southern Illinois University Carbondale

Domenico Parisi

National Research Council

1. Introduction

"Classical" connectionist models of development (e.g., Elman Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996) tend to view behavioral change in the individual as due only to the experience of the individual; i.e., as the outcome of learning by the individual. In contrast, the Connectionism in an Artificial Life Perspective (CALP) considers behavioral change as the outcome of adaptation at the level of the population as well as the individual. In this view, individual organisms possess inherited genotypes that specify some aspects of their neural network and behavior, and in turn, determine how the neural network and behavior change during life. Inherited genotypes are the present result of a process of change at the population level that has taken place in the preceding generations.

Thus, the CALP studies how populations of individually different networks change in successive generations because of the selective reproduction of individuals and the constant addition of new variant individuals due to sexual recombination and random mutations in inherited genotypes. Given differences in the inherited genotype, development in the individual is not just due to the experiences encountered by the individual. Instead, it results from the interaction between individual experiences and the inherited genotype. From this theoretical perspective, we refer to changes that are mainly (but never exclusively) due to the inherited genotype as "maturation", while changes that are mainly (but never exclusively) due to experience are called "learning", and changes in which the role of genetic information and experience seem to be equally important are called "development".

Both versions of connectionism can be said to instantiate a (Piagetian) "genetic epistemology" according to which, in order to understand X, one should try to reconstruct how X has become what it now is. However, unlike "classical" connectionism, CALP goes beyond

Piaget's emphasis on change in the individual because its "genetic epistemology" perspective includes both phylogeny and ontogeny (Parisi & Schlesinger, 2002).

CALP uses genetic algorithms (Holland, 1975) to study evolution. However, genetic algorithms can be viewed as abstract models of change in general, and as such they can be used in two different ways: (1) they can be used together with learning algorithms to study how evolution *across* individuals creates innate bases for learning and development in the individual, but (2) they can also be used to study development *within* an individual, with the behavior of successive generations of organisms representing the successive developmental phases within the individual.

Another important difference between "classical" connectionism and CALP is that individuals are always embodied and situated. "Classical" neural networks do not have a body and effectively live in a void. More precisely, their only "environment" is the researcher who decides which input arrives to the network's input units and who evaluates the network's output and provides a teaching input for learning. However, "natural" learning consists to a large extent of learning how the environment responds to the organism's own actions. This cannot be studied if there is no environment. In contrast, in CALP, a neural network is just one physical component of the organism's entire body. The organism lives and interacts within a physical environment. In a CALP simulation, one not only simulates a neural network and what takes place inside the network, but one also simulates the organism's body, with its specific physical properties such as size and shape and nature and location of sensors and effectors, and the physical environment in which the organism happens to live. The researcher does not decide upon the inputs to the organism's neural networks arbitrarily. They depend on the physical characteristics of the environment and of the organism's body and on the current location of the organism in the environment. This has the important consequence that the network's outputs, in particular the movements controlled by the network, tend to change either the physical relation of the organism's sensors to the environment or the environment itself. Thus, these organisms (at least in part) determine their own inputs (Nolfi & Parisi, 1993).

Because they are embedded within an environment, CALP neural networks may be called "ecological" networks. Supervised learning algorithms such as the backpropagation algorithm are rightly criticized as ecologically implausible because in many cases it is not clear where the teaching input that specifies the correct output comes from (Parisi, 1996; Parisi, Cecconi, & Nolfi, 1990). This criticism appears to be justified but not in all cases. For example, if a neural network is learning to predict the next state of a changing environment or the consequences of the organism's own actions on the environment, the next input arriving from the environment can be used as teaching input in an ecologically plausible way. Another example is a neural network that learns to imitate the output of another neural network in response to some input. The "imitator" uses the "model's" output as teaching input in a perfectly ecological manner.

In this chapter we provide an overview of modeling within the CALP framework. We then describe three simulations that illustrate the approach with examples from motor, cognitive, and language development.

2. Modeling Overview

In this section we provide a general introduction for non-specialists into the technical aspects of designing and studying a CALP simulation. For additional details, the interested reader may also consider Schlesinger and Parisi (2001b) and Schlesinger (in press); similarly, a comprehensive introduction with several examples is provided by Parisi, Cecconi, and Nolfi (1990).

We begin the overview by first describing the "key ingredients" of a simulation, while stressing the integrative aspects of the model. Next, we focus on the role of the autonomous agent or organism, and different levels at which the organism can be designed and studied. We then consider the task of designing an environment for our simulated organisms to inhabit. A related problem concerns the choice of how organisms will adapt to their environment (i.e., evolve, develop, etc.), that is, the learning algorithm.

2.1. Basic Principles

The Artificial Life framework that we employ is general and versatile enough to allow for formulating and investigating a wide range of research questions, varying across both a spatial and temporal scale. A particularly useful aspect of our approach is that it is consistently multi-leveled. For example, a typical simulation will include at least four distinct spatiotemporal scales: (1) a genetic level, including biological structures such as a genome, (2) a neuroanatomical or neurophysiological level, represented by an artificial neural network, (3) an ontogenetic level, at which an individual organism spends its life, and finally (4) a phylogenetic level, across which a population of organisms adapts on an evolutionary timescale.

Note that it is ultimately up to the model-builder to decide which of these levels should be included in their simulation, and if so, how much detail (i.e., biological and physical realism) needs to be incorporated into the model. To illustrate the flexibility of this approach, consider just a few examples of the kinds of simulations that we have investigated, and which vary widely in scope. For instance, one line of research (Schlesinger & Parisi, 2001b, Schlesinger, 2003; Schlesinger & Young, in press) simulates infants' perceptual activity (e.g., eye movements, visual scanning, etc.) in the specific context of an object permanence study (Baillargeon, 1986). This work has a relatively narrow focus on microgenetic learning experiences that occur in individual children as they participate in a particular experiment. Another series of studies investigates the development of reaching (Schlesinger & Parisi, 2001a; Schlesinger, Parisi, & Langer, 2000). Although operating on a longer developmental timescale, this work also focuses on ontogenetic changes within the individual. Finally, we have also studied numerous simulations which investigate changes in a population of artificial organisms, including maturation, learning, development, and evolution (e.g., Parisi, Cecconi, & Nolfi, 1990; Nolfi & Parisi, 2002).

2.2. Constructing the Organism

The concept of an autonomous, embodied organism plays a central role in our research. However, it is important to note that our use of the term embodiment extends well beyond the idea of possessing a physical body (e.g., a torso, arms, legs, etc.). Instead, consistent with the notion of a multi-leveled model, we conceptualize our organisms as embodied on several, simultaneous levels. Each level of embodiment provides a distinct set of physical constraints that influence developmental and evolutionary processes in a unique manner.

One level of embodiment is the physical shape and size of the organism, including how they occupy space, and where possible, how they move through it. At the same spatiotemporal scale, another closely-related level includes both the sensors (e.g., visual, auditory, haptic, etc.) that provide information about the current state experienced by the organism, as well as effectors or motor systems (e.g., arms, legs, eyes, mouth, etc.) for manipulating or changing the environment. At a more fine-grained level, there is also an artificial nervous system, which not only links the sensors to the effectors (i.e., input to output), but also provides a physical structure upon which learning, development, and evolution can operate (e.g., by changes in network connection weights, network topology, etc.). In virtually all of our simulations, we have employed an artificial neural network as a model of the nervous system.

Finally, in models that simulate evolutionary change, the model-builder must also construct a genome. Note that virtually any characteristic of the organism can be encoded in the genome, including not only the "congenital" properties of the neural network (i.e., those acquired at birth, e.g., the network architecture, the initial connection weights, etc.), but other inherited physical characteristics that may either not emerge at birth, or may change over time (e.g., size, shape, growth rates, learning rates, etc.). While there are a variety of techniques for representing these characteristics in a genetic code, a common strategy is to use binary strings of varying lengths to encode each property. For example, the strength of a connection between two units in the neural network can be encoded as a string of 8 ones and zeros. The entire network is then represented genetically by this collection of successive binary strings, analogous to a string of DNA.

2.3. Designing the Environment

In most of our work, we study simulated environments, though some evolutionary robotics studies include real environments (e.g., Nolfi, Floreano, Miglino, Mondada, 1994). Note that the process of designing the physical environment in which the organism is situated, in many ways, compliments how the organism is constructed. For example, if the organism has visual sensors, then there should naturally be features in the environment that can be seen and possibly discriminated, such as objects of different shapes, sizes, and colors. Similarly, if the organism is capable of locomotion, then the environment needs to be large enough, in relation to the organism, so that movement is possible. This synergy between the organism and the environment is no accident, of course, but instead follows logically from the idea of an embodied and situated organism.

In much of our work, we employ a minimalist principle in designing the environments that our organisms inhabit. That is, we try to avoid the mistake of making the environment "too real", for several reasons. First, overly complicated environments are not only difficult to program, but also consume large amounts of computing time and resources. A second and related concern is that unnecessarily complex environments often cloud or confuse the key relations we hope to study in simulation. Indeed, one of the main points of simulating a given phenomenon is to investigate it while stripping away extraneous details.

On the other hand, a critical feature of the environment is that it should incorporate as many physical and/or geometric constraints as necessary. In the case of our model of reaching, for example, this means carefully working out the equations which describe the position of the organism, and how a new position or posture is achieved by a combination of arm and eye movements. This kinematic model can be made more complicated by the addition of dynamic forces (e.g., gravity, inertia, etc.), which interact with each other in complicated ways. Similarly, models that employ vision as one of the organism's sensory modalities may have to account for various geometric properties of light propagation, such as occlusion and reflection.

2.4. Selecting the Learning Algorithm

Broadly construed, the learning algorithm is a set of computational rules which evaluate the behavior of the organism (or population of organisms), and produce a set of changes in the organism's neural network such that desired behaviors improve or become more likely. From an Artificial Life perspective, the learning algorithm takes on additional meaning. Not only is it a technique for studying adaptive behavior, but different algorithms can also be interpreted as adaptation over different timescales. For example, a genetic algorithm may operate by selecting the "most fit" organisms in a generation (i.e., those most successful on a given task, such as finding food), and using them to produce the next generation of organisms. Copying errors (i.e., mutations) in the neural networks of the offspring guarantee continued variability in performance, from which more successful organisms may be selected. This scenario, of course, is intended to be an implementation of neo-Darwinian processes via simulation.

It is important to remember that using a genetic algorithm is rarely intended to be a literal representation of how evolution occurs in biological organisms, nor more specifically, of how neural structures arise and change over time. Instead, as we have argued elsewhere (e.g., Schlesinger, in press), genetic algorithms play a more metaphorical role within CALP, providing a computational framework that (a) operates at the evolutionary timescale, (b) creates the opportunity to simulate innate (i.e., genetically-specified) behaviors, and (c) is versatile enough to allow multiple levels of change to be systematically studied and compared (e.g., related changes at the behavioral, neural, and genetic levels; see Miglino, Nolfi, & Parisi, 1996).

We have also investigated other learning algorithms on different timescales. As a proxy for learning during the individual lifespan, for example, we have implemented the backpropagation-of-error algorithm in a population of artificial organisms. An interesting question arises when multiple algorithms are combined: For example, what happens when we simulate a population that evolves via a genetic algorithm, but is also comprised of individuals who learn via backprop? Our work suggests that adding the capacity for learning or development, over and above the changes produced by evolution, can accelerate or facilitate the appearance of the evolved behavior (Parisi, Cecconi, & Nolfi, 1990; see also Hinton & Nolan, 1987). Interestingly, in some cases learning may facilitate evolution even when the learned behavior (e.g., learning to predict the sensory consequences of planned movements) is different from the evolved behavior (e.g., foraging for food).

Our view, more generally, is that there is a wide array of learning methods appropriate for connectionist models, including not only genetic algorithms and backprop, but also reinforcement learning, autoassociators, Hebbian learning, and adaptive resonance theory (for recent reviews, see Parisi & Schlesinger, 2002; Schlesinger & Parisi, in press). To our knowledge, only the genetic algorithm approach seems well-suited for simulating evolutionary processes, while each of the others might be used interchangeably for investigating ontogenetic processes such as learning, maturation, and development.

In order to illustrate how the concept of an embodied and situated organism can play a central role in developmental research, we next present a brief survey of our work from three developmental domains: motor development, cognitive development, and language development.

3. Motor Development

Unlike models of cognitive processes (e.g., object recognition or categorization) which do not necessarily need to simulate an embodied organism, models of motor activity are almost inevitably embodied. This is not only because there is a broad consensus among researchers that motor development is an interactive process (e.g., Berthier, 1996; von Hofsten, 1984; Thelen, Corbetta, Kamm, Spencer, Schneider, & Zernicke., 1993), but also because motor activity requires both sensing one's current state in the environment (e.g., the location of the hands with respect to a nearby object) and also acting on the environment (e.g., reaching, moving, looking, vocalizing, etc.).

As an example of how embodiment plays a central role in motor development, consider the problem of learning to reach. Within approximately their first three months, young infants acquire the ability to move their hands efficiently toward desired objects (efficient grasping, on the other hand, develops more slowly). This ability involves the coordination of at least two major sensory systems (i.e., vision and proprioception) and three major motor systems (i.e., movement of the head, eyes, and arms). What physical constraints play a role in the acquisition of this new skill?

We addressed this question by designing and studying a population of artificial neural networks, which represent the neural control systems that underlie the development of hand-eye coordination in young infants (Schlesinger, Parisi, & Langer, 2000). The model includes constraints on the development process that are both embodied and situated. First, the simulated physical body includes a monocular visual system, a two-segment arm (with a ball joint at the shoulder and a hinge joint at the elbow), and a feedforward neural network (see Box 1). One geometric constraint imposed by this design is that rotation of the body axis or shoulder joint causes the arm endpoint to move, even when the elbow joint is held in place. As we see below, a developmental consequence is that control of the arm proceeds in a center-to-periphery, or proximodistal pattern, with consistent use of the shoulder joint emerging before use of the elbow joint. Second, a related, situated constraint follows from placing the organism in an environment in which it can rotate its visual field 180°. Consequently, when neither the hand nor the target object are visible, the organism can learn to either (a) first find the target and then bring the hand into the visual field (i.e., the more efficient strategy), or (b) first find the hand, and then “drag” it around while searching for the target. Indeed, the model reliably acquires the former strategy, mirroring the developmental pattern in human infants (e.g., Clifton, Rochat, Robin, & Berthier, 1994).

Several important conclusions emerge from this work. First, note that the physical design of our organism includes a two-segment arm (i.e., an upperarm and forearm) with rotation limits comparable to a human arm; the organism can also rotate its torso while reaching. Thus, the organism effectively has three movement degrees of freedom (DOF), while the target object is located on a two-dimensional plane, leaving the organism with one redundant DOF (this is known as the "DOF problem", see Bernstein, 1967). Instead of using all three DOF, the model instead quickly learns to "freeze" the shoulder joint and use only torso and elbow rotations to reach a target. Interestingly, the strategy of freezing redundant joints is a common phenomenon in adult novices as they acquire new motor skills (e.g., learning to ski; see Vereijken, Emmerik, Whiting, & Newell, 1992).

A second major result that emerges from our reaching model concerns how opposing pairs of muscles (i.e., flexors and extensors) that are used to rotate each joint are coordinated. In the model, we observe a proximodistal pattern of development, in which the muscles controlling the shoulder joint become coordinated before those controlling the elbow joint. This result not only corroborates kinematic analyses of infants' reaching movements (i.e., changes in hand position and velocity, e.g., von Hofsten, 1984; Trevarthen, Murray, & Hubley, 1981), but also suggests that the initial stiffness or co-contractions observed throughout young infants' arms and shoulders as they begin to reach (Thelen et al., 1993) should diminish first in proximal segments of the infant's body before more distal segments (e.g., in the upperarm before the forearm).

More recently, we have used the reaching model to investigate the sensory encoding strategies that are available to infants for locating the position of their arms in space, and for moving their hand toward a desired target (Schlesinger & Parisi, 2001a). A number of studies suggest that, contrary to earlier theories, infants do not learn to reach by visually guiding their hand toward a target (e.g., Clifton, Muir, Ashmead, & Clarkson, 1993; Clifton, Rochat, Robin, & Berthier, 1994). Instead, by the onset of reaching, infants are equally successful at reaching for a target, whether or not their hand is visible (i.e., under normal illumination vs. an illuminated target in a darkened room). This and other empirical observations (e.g., infants do not attend to their hands as they reach) suggest that vision and proprioception become coordinated through a modality other than vision.

In an extension of our original reaching model (Schlesinger & Parisi, 2001a), we hypothesized that tactile input, made during contact with the target, might provide a third sensory channel that helps "triangulate" vision of the target with proprioception from the arm. A series of simulations provided strong support for our hypothesis. First, we compared learning across two conditions (i.e., in two populations of artificial neural networks): in one condition, a tactile sensor becomes active during contact with a target, while in the second condition the tactile sensor is permanently deactivated. Performance was superior, across a number of measures, when tactile input was available. Specifically, in this condition (1) learning was more rapid and reached a higher level of performance, (2) latency to contact with the target was shorter, and (3) after reaching the target, more sustained contact with the target was achieved.

Second, we designed a training scenario analogous to the Clifton et al. (1993) study, in which reaching under normal lighting conditions was compared to reaching toward luminous targets in the dark. In particular, we simulated a condition, comparable to the first described above (i.e., with tactile input from the hand), but with one important change: the hand and arm were made invisible, so that while the target could be seen, the hand could only be localized through proprioceptive feedback from the arm. Note that this simulation provides a critical test of our hypothesis. Specifically, if sight of the hand is necessary in the model for learning to reach, then performance should be severely disrupted by making the hand invisible. Alternatively, if contact with the target during reaching is sufficient for coordinating vision and proprioception, then making the hand invisible should not affect performance.

Our results provided unequivocal support for our hypothesis. That is, we found that even when the hand is made invisible, there are no disruptions in either the learning trajectory or final performance. Effectively, the model learns equally well to coordinate vision and proprioception, regardless of whether the hand is visible. A number of empirical predictions follow from these results. First, if touching an object acts like a "reward function" that links vision and proprioception, we should predict that postures (i.e., head, eye, and arm configurations) that are near to those during successful reaching trajectories should be more salient to infants, and therefore easier for them to discriminate than postures that are far from reaching trajectories. Second, if infants' visual-proprioceptive coordination is perturbed (e.g., by wearing displacing prisms that make objects appear in non-veridical locations; see McDonnell and Abraham, 1975), we should also predict that infants will adapt more rapidly as they reach for real objects, that is, when there is tactile feedback, than when they reach for virtual objects (e.g., holographic images).

4. Cognitive Development

Most computational models of cognitive development do not employ an embodied or situated approach (but see Jones, Ritter, & Wood, 2000; Sporns, this volume; Thelen, Schöner, Scheier, & Smith, 2001 for recent exceptions). Instead, the prevailing approach is to simulate a specific neural circuit or processing loop, while sidestepping the issue of designing both the physical structure of the organism, as well as the structure of the environment. In this case, the environment is treated as a stream of sensory information, which is pre-parsed or segmented into a set of input patterns. These patterns then comprise a training set, from which a pattern is selected at random and presented to the model. We referred above to this training regime as "classical connectionism" (Parisi, 1997).

Unfortunately, this approach is somewhat at odds not only with developmental theories that stress the importance of exploration and interaction with the environment (e.g., Gibson, 2000; Piaget, 1952), but also with developmental research that highlights the role of self-directed activity (e.g., Bertenthal, Campos, & Kermoian, 1994; Held & Hein, 1963). A related concern is that many models also focus on the role of internal representations, while often overlooking the contribution of implicit knowledge that is encoded or represented through skillful interaction with the environment (e.g., procedural knowledge, sensorimotor representations, etc.; see Ballard, Hayhoe, Pook, & Rao, 1997; Brooks, 1991; O'Regan, 1992).

In contrast, the notion of an embodied and situated organism provides an ideal framework not only for investigating the role of self-produced activity in cognitive development, but also for identifying a wide array of constraints that influence representations as they emerge early in development. To illustrate our point, consider the development of object permanence, which involves the understanding that an object continues to exist while it is out of sight (Piaget, 1952). As a specific example, imagine an infant confronted with the problem of visually tracking a moving object, which occasionally disappears and then reappears as it passes behind a screen.

One strategy for addressing this problem is to assume that infants form internal representations of hidden objects, and that they use these representations to guide their behavior (e.g., where to search for the hidden object), when perceptual contact with the object is lost. This approach has inspired a number of connectionist models (e.g., Mareschal, Plunkett, & Harris, 1999; Munakata, McClelland, Johnson, & Siegler, 1997), which seek to explain the development of infants' search behaviors in terms of progressively more detailed and accurate internal representations or mental models.

As an alternative approach, we might assume that infants' earliest representations are not internal "copies" of external experience, but instead are pragmatic systems of knowledge about how one can act in the current situation, and how ongoing experience may be transformed by those actions (for related accounts, see O'Regan & Noe, 2001; Thelen & Smith, 1994). From this perspective, the developmental task is not to learn to construct a (disembodied, general-purpose) mental image, model, or representation of hidden objects and events, but rather, to construct a set of task-specific representations that link perception with action. Naturally, these links form the basis not only for adaptive, prospective, goal-directed behavior, but also for knowledge about the physical world (including the concept of object permanence). Note that this knowledge therefore is always embedded in action, and is situated and embodied.

In order to illustrate how this alternative approach might work in practice, we designed a connectionist model of an infant as it learns to visually track moving objects (Schlesinger & Barto, 1999; Schlesinger & Parisi, 2001b). In contrast to our other models, the tracking model focuses on change at the individual or ontogenetic level, during a relatively brief time span (e.g., during an experimental session). There are three important aspects of the tracking model.

First, it implements a relatively simple, but nevertheless embodied design by simulating an oculomotor control system, including: (1) an "eye" with a 2-dimensional retina, (2) a feed-forward neural network, and (3) a motor system that drives eye movements, allowing the model to shift its gaze from one location to another (i.e., to control its visual input over time). Second, there are neither built-in memory systems (e.g., a delayed input line, recurrent connections from one layer in the network to another, etc.) nor sensory prediction systems. In other words, the model's ability to act prospectively (e.g., to anticipate when and where a hidden object will reappear) is not based either on input that has been temporally-coded, or on an internal model that computes past or future states. Third, and perhaps most importantly, the overall design of the model is based on bottom-up principles; in particular, there is no built-in knowledge or abilities, and therefore, no top-down influence on the model's perceptual processing from previous experiences or expectations.

We began our evaluation of the tracking model by assessing its ability to keep a moving target within view. Specifically, we trained the model to track the target across a variety of conditions (e.g., normal tracking, tracking with a brief occlusion of the target, etc.). The performance of the model during training, as well as its generalization to novel tracking events, replicated several key aspects of visual tracking development in young infants (e.g., Aslin, 1987; Bremner, 1985). First, the model learns to track fully visible targets before tracking targets that are briefly occluded. Second, as it learns to track occluded targets, the model's initial strategy is to continue to "fixate" the location where the target was last seen (i.e., at the point where it becomes occluded). Third, like human infants, as the model tracks occluded targets it also learns to generate anticipatory eye-movements or saccades to the location where the occluded target reappears. Finally, the model also generates a pattern of gaze-shift errors comparable to those produced by 9-month-old infants (Berthier et al., 2001), when the target encounters an obstacle during occlusion (Schlesinger & Parisi, 2001b).

A key result of the tracking model is that when the target is occluded, like human infants, the model generates a series of anticipatory eye movements, toward the location where the target will reappear (Johnson, Amso, & Slemmer, 2003). However, given the design of the model, it is clear that this prospective or future-oriented behavior is not the result of an explicit prediction or computation, but is instead simply an adaptive by-product of acquiring an effective tracking strategy.

Given these initial positive findings, we wondered how the tracking model might inform the debate on infants' early object knowledge (e.g., Baillargeon, 1999; Haith, 1998; Smith, 1999;

Spelke, 1998). One of the questions raised by this debate is what kind of knowledge is tapped by methods such as the violation-of-expectation (VOE) paradigm, which measure infants' looking time during possible versus magical or physically impossible events? In particular, why do infants often look longer at impossible events (e.g., Spelke, 1985)? We addressed this question by elaborating the tracking model in a number of ways, so that it could "watch" a series of possible and impossible events comparable to those shown to infants in a study that investigates the development of object permanence (Baillargeon, 1986).

In Baillargeon's "car study", infants see a car that rolls down a track, behind a screen, and then out the other side (for additional details, see Box 2). After several repetitions of this event, infants then watch a series of possible and impossible events, in which an obstacle is placed either behind or on the track, and is then hidden by the screen. After the obstacle is hidden, the car then appears and moves along the track as before. Baillargeon found that infants looked significantly longer when the obstacle was placed on the track (i.e., during the impossible event), suggesting that they were surprised to see the car reappear when its path was obstructed by a hidden object.

We began by first presenting the tracking model with a computer-animation event comparable to the event that infants first watch in the car study. Specifically, the model was trained to fixate the car, and to follow its movement. After training, the model was then presented with animated versions of the possible and impossible events. A number of important results emerged from this work (Schlesinger, 2003).

First, we found that the model was significantly better at tracking the car during the possible event. That is, the appearance of an object in the car's path disrupted the model's ability to track the car. Second, as we monitored the internal activity (i.e., hidden layer activations) of the tracking model, we also found that the model "represented" the possible and impossible events differently. In particular, relative to the initial training event, the impossible event was significantly more novel or unfamiliar. Note that each of these findings provides an indirect replication of Baillargeon's results, and also suggests why infants may look longer at the impossible event in her study: (1) because the car is more difficult to track during the impossible event, and/or (2) because the impossible event is also more perceptually novel or unfamiliar.

The latest version of the tracking model has generated a variety of empirical predictions regarding infants' moment-to-moment visual activity during VOE studies like the car study (Schlesinger, 2003). For example, the model not only predicts that infants will distribute their gazes differently during the possible and impossible events, but also suggests a specific qualitative pattern of gaze-direction during these events. In order to evaluate these predictions, we recently conducted a replication of Baillargeon's original car study with 6-month-old infants (Schlesinger & Casey, 2003). We not only found, like Baillargeon, that infants looked significantly longer at the impossible event, but also found that the tracking model successfully predicted qualitative features of infants' visual activity during the possible and impossible events.

5. Language Development

Artificial Life simulations can be used to study how organisms can acquire an ability to respond to signals appropriately (i.e., understanding or comprehension) and to produce the various signals in the appropriate circumstances (i.e., production). More specifically, Artificial Life simulations can be used to investigate how different classes of signals (words) emerge in the experience of language users and how smaller signals (words) can make up larger signals (sentences).

Imagine an organism whose neural network has visual input units encoding the location and perceptual properties of objects present in the environment and auditory input units encoding the sounds produced by a conspecific. The network has to respond to the inputs by generating the appropriate locomotory movements. The environment contains edible and poisonous mushrooms and each individual mushroom is different from all others (Cangelosi & Parisi, 1998). If an encountered mushroom is sufficiently near to the organism, the organism can perceive both the location and the perceptual properties of the mushroom and therefore it can recognize if the mushroom is edible or poisonous and respond appropriately, i.e., it approaches and eats the edible mushrooms and avoids the poisonous ones. However, because of the organism's sensory limitations, if the mushroom is more distant, only the mushroom's location is perceived by the organism's neural network but not the mushroom's perceptual properties. Hence, in these circumstances the only option open to the organism is to approach every mushroom until the mushroom is sufficiently close that the organism can recognize if it is of the edible or poisonous variety. But of course the resulting behavior would not be very efficient.

Language can emerge to solve this problem. Another individual, who is nearer to the mushroom or has fewer sensory limitations than the first individual, recognizes the mushroom and generates one of two sounds, X1 and X2, which co-vary with the two categories of mushrooms. The first individual hears the sound produced by the conspecific and responds by approaching the mushroom if it hears the sound co-varying with edible mushrooms and going away from the mushroom if it hears the other sound.

Are the two sounds of this very simple language, X1 and X2, nouns or verbs? There is no real justification for answering one way or another. The two sounds can be paraphrased by using either two English nouns (actually, noun phrases), "edible mushroom" and "poisonous mushroom", or two English verbs, "approach" and "avoid". Hence, calling the two sounds either nouns or verbs would be arbitrary.

Now imagine another, somewhat more complex, scenario. The organism has a retina and a 2-segment arm, like in the simulations previously described. The environment contains only two objects that have different shapes. At any given time the organism sees both objects at the same time and it hears one of two different sounds, N1 or N2. The organism responds to the input by moving its 2-segment arm. When it hears N1, the organism is rewarded if it reaches for the object possessing a given shape, while when it hears N2 it is rewarded if it reaches the object with the alternative shape. Now we would be somewhat justified in saying that the two sounds are nouns, if we define a noun as a sound which co-varies with the particular object on which a given action (reaching, in our case) is executed, independently of the particular action which is executed on the object.

We can also define verbs if we imagine a variation of our scenario. The organism sees a single object, which may be of any shape, and it hears one of two sounds, V1 and V2. When the organism hears V1, it is rewarded if it pushes the object away from itself, and when it hears V2 it is rewarded if it pulls the object toward itself. The sounds now are verbs, not nouns, where a verb is defined as a sound which co-varies with the particular action which must be executed on an object, whatever the object.

If the two scenarios can be put together, then we will see the emergence of simple sentences. At any given time the organism sees two objects that have different shapes, and hears a sequence made up of two sounds. The organism can hear one of 4 possible sequences: N1+V1, N1+V2, N2+V1, N2+V2. The N sounds allow the organism to select the object on which the action must be executed, while the V sounds allow the organism to select the action that must be executed on the object.

Variations of the simulation scenario can easily be constructed to study the emergence of further aspects of language (Parisi, Cangelosi, & Falcetta, 2002). Imagine that a multiplicity of objects varying in both shape and color are simultaneously seen by the organism. To know on which of the objects a given action must be executed, our neural network will need not a just single sound but a sequence of two sounds, one allowing the network to select the class of objects with a given shape and the other one for selecting the object with a given color within that class. The sequence of two sounds is a noun phrase. The former sound is a noun because it can occur alone and because in language nouns tend to co-vary with the shape of objects, not with their other properties (Landau & Smith, 1998). The latter sound is a noun modifier, more precisely, an adjective. Noun modifiers can be locatives if the sound accompanying the noun co-varies with the location of the object in the visual field or with respect to another object, not with other, more permanent, properties of objects such as their color.

Language emerges at the end of the first year of life when infants begin to notice how specific sounds co-vary with specific aspects of their experience, i.e., of what they perceive and what they and other individuals do. Prior to that, during their entire first year, they have learned both how to map nonlinguistic inputs into nonlinguistic outputs (sensory-motor skills) and how to map heard sounds - which may be produced by others or self-produced - into produced sounds (imitation of sounds, including self-imitation). One can hypothesize that at least two functionally separate networks exist in the brain for doing these two different things for the entire first year of life. At around 1 year, the two networks begin to become connected together. This means that nonlinguistic input (e.g., an object is seen) may be mapped into linguistic output (a sound is produced) and linguistic input (a sound is heard) may be mapped into nonlinguistic output (an action is executed). This is the beginning of, respectively, language production and language comprehension.

As suggested by our simulations, one may also hypothesize that for some time after the onset of language at around one year the linguistic signals that are produced and understood by children cannot be assigned to distinct classes from the children's point of view, i.e., in their brain, because infants have not had sufficient linguistic experience to notice systematic differences in the co-variation patterns of different classes of signals. On the other hand, many of the early linguistic signals produced by children tend to be recognized and interpreted by adults as nouns (i.e., the names of things). This suggests that "noun" as a linguistic category may have a developmental advantage over other categories of words. Indeed, our simulations can be interpreted as indicating that nouns - and noun phrases - are, basically, means for directing the attention of the hearer to specific portions of what is currently perceived by the hearer - or *should* be perceived, when the hearer is asked to search for objects with given characteristics. A crucial skill that infants are developing at this age is the ability to selectively attend to portions of what they perceive and, especially, to sharing selective attention with others (e.g., Butterworth & Jarret, 1991). We believe that nouns are well-suited for playing an important role in this process.

6. Summary

In this chapter, we have introduced Connectionism in an Artificial Life Perspective (CALP). The two key elements of our approach are (1) the use of neural networks, and (2) autonomous organisms that learn, develop, and evolve through Artificial Life. In combination, these elements not only highlight the interactive nature of developmental processes, but also provide a broad, flexible framework for simulating and investigating situated and embodied knowledge. We have also argued, more generally, that by conceptualizing the individual organism as an autonomous agent, complex interactions that span multiple spatiotemporal scales can be systematically studied. One of the valuable benefits of this approach is that the influences of learning, maturation, development, and evolution, often considered in isolation, can all be incorporated into the same modeling process.

To situate our work within a larger perspective, we now raise and respond to what we view as four fundamental issues that can be used to evaluate CALP. First, which features of the embodied and situated organism may constrain the development of representations?

Although a variety of embodied constraints play a role in each of the models that we described (i.e., motor, cognitive, and language), the tracking model is particularly well-suited to address the development of representations. Specifically, this work questions the assumption that prospective, object-oriented activity in infants requires an internal model of the external world. Therefore, rather than building in structures that explicitly support memory-based or predictive processes, our key design strategy is instead to build a sensorimotor (i.e., oculomotor) system in which sensory feedback informs each action. Although the tracking model "lives" in a snapshot world, it quickly learns to generate actions which are implicitly predictive. In particular, it efficiently tracks moving objects, adjusting its gaze as necessary to keep the target object within view. In addition, we also note that during occlusion, the model shifts its gaze to the location where the target will reappear, although it is not explicitly taught or instructed to do so (see Box 2).

How are these future-oriented behaviors achieved without some kind of internal model? There are at least three levels of constraint that operate in the tracking model. First, the model "selects" its own input. As we highlight elsewhere (e.g., Schlesinger & Parisi, 2001b), we believe that sensory self-selection is a crucial element of development, and more specifically, that it shifts the question of representation development away from "how an internal copy of the environment is learned", and instead toward "how an adaptive sensorimotor mapping of the environment is learned". It is this mapping, we propose, that implicitly encodes key features of the environment. Second, the physical design of our one-eyed organism incorporates a number of complementary constraints, including in particular two visual processing streams, analogous to the fovea and visual periphery. Finally, the specific learning algorithm employed (i.e., reinforcement learning) also constrains the process of representation development, or perhaps more accurately, *enhances the likelihood* that the tracking model will exploit sensorimotor strategies of representation, by shaping the model's performance with evaluation (i.e., a global assessment of behavior as a whole) rather than by specific, instructive feedback (i.e., by supervised learning).

This suggests a second, fundamental issue: in general, how are embodied and situated constraints implemented in CALP? As we have stressed repeatedly in this chapter, each level of "physicality" provides a specific set of constraints on development. One level is the morphological design of the organism, which allows it to recruit and exploit geometric features or regularities in the environment. Corresponding to this design, of course, is the structure of the environment, which affords various opportunities for learning and development. For example, the reaching model learns that only two of its three joint DOF are necessary for reaching along a 2-dimensional, planar surface. Another level of constraint is imposed by the architecture of the neural network, which in turn is constrained by the design and implementation, where appropriate, of the genetic code that determines the internal and external structure of the organism.

A third fundamental issue concerns the major lessons that CALP may offer for developmental researchers. One implication of our modeling work is a renewed emphasis on the role of action during development, in contrast to recent alternative approaches that have focused on passive, statistical-learning processes (e.g., Marcus, Vijayan, Rao, &Vishton, 1999; Saffran, 2003). Note that action plays a particularly important role in our language work, where goal-directed actions are acquired in the context of a linguistic signal. We hypothesize that it is the coordination of these actions with ongoing linguistic input that enables the infant to construct implicit categories for "verb" and "noun", and to begin recognizing and ultimately using these proto-words in conventional ways.

Another implication of CALP is the notion of an "emergent constraint". This is the idea that the outcome of a developmental process need not be programmed in by maturation, but instead may occur as the result of successive learning experiences that the organism determines or selects for itself. For example, in our reaching model, the upperarm and forearm segments are controlled by parallel, analogous systems, and yet the model consistently learns to efficiently move the upperarm first. This temporal bias suggests a proximodistal developmental pattern, as the organism first learns to control limb segments that are closest to the central body axis or torso. Therefore, the proximodistal pattern constrains future developmental experiences, without being explicitly pre-programmed into development itself.

Finally, what concrete predictions, if any, do our CALP models suggest? First, with respect to motor development, our reaching model predicts a proximodistal pattern of muscle control, specifically beginning first in the shoulder and upperarm muscles, followed by the forearm muscles. The reaching model also predicts that while wearing displacing prisms, infants should learn to adapt or recalibrate their reaches toward real, 3-D objects before they adapt their reaches toward virtual objects.

Second, the tracking model predicts that in preferential-looking studies that implement the violation-of-expectation paradigm, infants should generate different gaze patterns during possible and impossible events. In particular, the model predicts that infants' gaze patterns during the habituation (or familiarization) event and the possible event should be most similar, while the impossible event should provoke largely novel scanning or gaze patterns.

Third, one of the predictions of the language model is that separate neural structures for (1) mapping nonlinguistic, sensorimotor input onto sensorimotor output, on the one hand, and (2) mapping linguistic, auditory input onto linguistic, vocal output, on the other hand, initially develop along parallel, but independent trajectories. The model also predicts that these structures or pathways become coordinated (e.g., directly or indirectly connected) at roughly the end of the first year of life, enabling the infant to use their existing sensorimotor knowledge to begin inferring meaning or intent from linguistic signals.

In conclusion, despite the many unique advantages of studying neural networks in an Artificial Life perspective, we acknowledge that our approach is not only one of many different modeling techniques, but more importantly, that it also faces a number of significant challenges (e.g., Parisi & Schlesinger, 2002; Schlesinger & Parisi, 2001b). For example, how can an organism that is capable of only sensorimotor representation ultimately acquire the capacity for symbolic mental activity (e.g., metaphor, deduction, counterfactual reasoning, etc.)? While there are a variety of theories about how abstract mental symbols may develop out of embodied knowledge (e.g., Johnson, 1987; Piaget, 1952), it is certainly a more difficult task to demonstrate this process convincingly through simulation. Nevertheless, we are confident that the CALP will be able to address these kinds of questions, and more generally, continue to play a vital role in developmental research.

Acknowledgments

This work was supported in part by a grant to the first author from the National Institute of Child Health and Human Development (1R03 HD40780-02).

Glossary

DOF problem – Systems with multiple free parameters are often underconstrained: they have more degrees of freedom (DOF) than are necessary to solve a given task. During arm movements, for example, humans require only 3 DOF (i.e., moveable limb segments) to reach for or point to any object in a 3-dimensional workspace. In this case, the DOF problem arises because of redundant or unnecessary limb segments and their corresponding joints (e.g., waist, shoulder, elbow, wrist, etc.), which create an unlimited number of possible movement paths.

genetic epistemology – A formal study of cognitive development, made famous by Piaget, that defines knowledge acquisition as a constructive process influenced by both genetic and experiential factors.

genome – The entire pool of genetic material in a population of organisms.

genotype – The set of genetic material inherited by an individual. Genotype also refers to specific portions of genetic material (e.g., a gene) that encode a particular characteristic (e.g., the genotype for brown eyes).

ontogeny – The process of change during the lifespan (i.e., from conception to death) of an individual organism.

phenotype – The observable characteristics of an individual, determined in part by their genotype.

phylogeny – The process of change that spans multiple generations (i.e., evolution).

References

Aslin, R.N. (1987). Visual and auditory development in infancy. In J.D. Osofsky (Ed.), *Handbook of infant development*, *2nd ed*. (pp. 5-97). New York: Wiley and Sons.

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition*, *23*, 21-41.

Baillargeon, R. (1999). Young infants' expectations about hidden objects: A reply to three challenges. *Developmental Science*, *2*, 115-132.

Ballard, D.H., Hayhoe, M.M., Pook, P.K., & Rao, R.P.N. (1997). Deictic codes for the embodiment of cognition. *Behavioral & Brain Sciences*, *20*, 723-767.

Bertenthal, B.I., Campos, J.J., & Kermoian, R. (1994). An epigenetic perspective on the development of self-produced locomotion and its consequences. *Current Directions in Psychological Science*, *3*, 140-145.

Berthier, N.E. (1996). Learning to reach: A mathematical model. *Developmental Psychology*, *32*, 811-823.

Berthier, N.E., Bertenthal, B.I., Seaks, J.D., Sylvia, M.R., Johnson, R.L., & Clifton, R.K. (2001). Using object knowledge in visual tracking and reaching. *Infancy*, *2*, 257-284.

Bremner, J.G. (1985). Object tracking and search in infancy: A review of data and a theoretical evaluation. *Developmental Review*, *5*, 371-396.

Brooks, R.A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139-159.

Butterworth, G., & Jarret, N. (1991). What minds have in common in space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55-72.

Calabretta, R., Nolfi, S., Parisi, D., & Wagner G.P. (2000). Duplication of modules facilitates the evolution of functional specialization. *Artificial Life*, *6*, 69-84.

Cangelosi, A. & Parisi, D. (1998). The emergence of 'language' in an evolving population of neural networks. *Connection Science*, *10*, 83-97.

Cangelosi, A., Parisi, D., & Nolfi, S. (1994). Cell division and migration in a 'genotype' for neural networks. *Network*, *5*, 497-515.

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.

Clifton, R.K., Rochat, P, Robin, D.J., & Berthier, N.E. (1994). Multimodal perception in the control of infant reaching. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 876-886.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.

Elman, J.L. (2003). Development: it's about time. *Developmental Science*, 6, 430-433.

Georgopolous, A.P., Schwartz, A.B., & Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science*, *233*, 1416-1419.

Gibson, E.J. (2000). Perceptual learning in development: Some basic concepts. *Ecological Psychology*, *12*, 295-302.

Haith, M.M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development*, *21*, 167-179.

Held, R., & Hein, A. (1963). Movement produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, *56*, 872-876.

Hinton, G.E., & Nolan, S.J. (1987). How learning can guide evolution. *Complex Systems, 1*, 495-502.

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

Holland, J.H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of imagination, reason, and meaning*. Chicago: University of Chicago Press.

Johnson, S.P., Amso, D., & Slemmer, J.A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences*, *100*, 10568-10573.

Jones, G., Ritter, F.E., & Wood, D.J. (2000). Using a cognitive architecture to examine what develops. *Psychological Science, 11*, 93-100.

Landau, B., Smith, L.B. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *2*, 291-321.

Marcus, G.F, Vijayan, S, Rao, S.B., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77-80.

Mareschal, D., Plunkett, K., & Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviours in infancy. *Developmental Science*, *2*, 306-317.

McDonnell, P.M., & Abraham, W.C. (1979). Adaptation to displacing prisms in human infants. *Perception*, *8*, 175-185.

Miglino, O., Nolfi, S. & Parisi, D. (1996). Discontinuity in evolution: How different levels of organization imply pre-adaptation. In R.K. Belew & M. Mitchell (Eds.), *Adaptive individuals in evolving populations* (pp. 399-415). Readings, MA: Addison-Wesley.

Munakata, Y., McClelland, J.L., Johnson, M.H., & Siegler, R.S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, *104*, 686-713.

Nolfi, S., Floreano, D., Miglino, O. & Mondada F. (1994). How to evolve autonomous robots: Different approaches in evolutionary robotics. In R. Brooks and P. Maes (Eds.), *Proceedings of the International Conference Artificial Life IV*. (pp. 190-197). Cambridge Mass: MIT Press.

Nolfi, S., & Parisi, D. (1993). Self-selection of input stimuli for improving performance. In G.A. Bekey & K.Y. Goldberg (Eds.), *Neural networks in robotics* (pp. 403-418). Boston: Kluwer.

Nolfi, S., & Parisi, D. (2002). Evolution of artificial neural networks. In M. A. Arbib (Ed.), *Handbook of brain theory and neural networks* (pp. 418-421). Cambridge, MA: MIT Press.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461-488.

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

Parisi, D. (1996). Computational models of developmental mechanisms. In R. Gelman, & T.K. Au (eds.), *Perceptual and Cognitive Development* (pp. 373-412). San Diego: Academic Press.

Parisi, D. (1997). Active sampling in evolving neural networks. *Human Development*, *40*, 320-324.

Parisi, D., Cangelosi, A. & Falcetta, I. (2002). Verbs, nouns, and simulated language games. *Italian Journal of Linguistics*, *14*, 99-114.

Parisi, D, Cecconi, F, & Nolfi, S. (1990). Econets: Neural networks that learn in an environment. *Network*, *1*, 149-16.

Parisi, D., & Schlesinger, M. (2002). Artificial Life and Piaget. *Cognitive Development*, *17*, 1301-1321.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110-114.

Schlesinger, M. (2003). A lesson from robotics: Modeling infants as autonomous agents. *Adaptive Behavior*, *11*, 97-107.

Schlesinger, M. (in press). Evolving agents as a metaphor for the developing child. *Developmental Science*.

Schlesinger, M., & Barto, A. (1999). Optimal control methods for simulating the perception of causality in young infants. In M. Hahn & S.C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 625-630). New Jersey: Erlbaum.

Schlesinger, M., & Casey, P. (2003). Visual expectations in infants: Evaluating the gaze-direction model. In C.G. Prince, L. Berthouze, H. Kozima, D. Bullock, G. Stojanov, & C. Balkenius (Eds.), *Proceedings of the Third International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems* (pp. 115-122). Sweden: Lund University Cognitive Studies.

Schlesinger, M., & Parisi, D. (2001a). Multimodal control of reaching: The role of tactile feedback. *IEEE Transactions on Evolutionary Computation: Special Section on Evolutionary Computation and Cognitive Science*, *5*, 122-128.

Schlesinger, M, & Parisi, D. (2001b). The agent-based approach: A new direction for computational models of development. *Developmental Review*, *21*, 121-146.

Schlesinger, M., & Parisi, D. (Eds.). (in press). Beyond backprop: Emerging trends in connectionist models of development. [Special section]. *Developmental Science*.

Schlesinger, M., Parisi, D., & Langer, J. (2000). Learning to reach by constraining the movement search space. *Developmental Science*, *3*, 67-80.

Schlesinger, M., & Young, M.E. (2003). Examining the role of prediction in infants' physical knowledge. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*.

Smith, L.B. (1999). Do infants possess innate knowledge structures? The con side. *Developmental Science*, *2*, 133-144.

Spelke, E.S. (1985). Preferential looking methods as a tool for the study of cognition in infancy. In G. Gottlieb & N. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life* (pp. 323-363). Norwood, NJ: Ablex.

Spelke, E.S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, *21*, 181-200.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J.P., Schneider, K., & Zernicke, R.F. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. *Child Development*, *64*, 1058-1098.

Thelen, E., Schöner, G., Scheier, C. & Smith, L., B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*, 1-34.

Vereijken, B., Emmerik, R.E.A. van, Whiting, H.T.A., & Newell, K.M. (1992). Free(z)ing degrees of freedom in skill acquisition. *Journal of Motor Behavior*, *24*, 133-142.

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

Box Text

Box 1. Our “econet” (ecological neural network) model of reaching (Schlesinger, Parisi, & Langer, 2000) was designed to simulate the development of hand-eye coordination in young infants. A key focus of the model is on the role of trial-and-error learning mechanisms during the development of reaching. Accordingly, we implemented a population-based approach in which a number of alternative movement patterns are produced and evaluated, with selection and reproduction of the most successful movements.

This work not only highlights the role of trial-and-error learning, but it is also loosely based on the concept of a population of motor neurons, in which subsets of neurons each code a particular movement direction (Georgopolous, Schwartz, & Kettner, 1986). Consequently our model represents an individual child by simulating a population of neural networks, each of which “competes” for the control of the individual’s eye and arm movements.

Nevertheless, we should stress in this case that not only the population-based approach, but also our use of a genetic algorithm are intended as a metaphor or proxy for the trial-and-error strategies employed by an individual child, rather than as an explicit assumption regarding the actual neural mechanisms involved (for a related discussion, see Schlesinger, in press).

As Figure 1A illustrates, each econet is initially represented by a binary string, which specifies the physical characteristics of the econet. In the first generation of econets, each string is a random sequence of 1’s and 0’s. The string is then used to “build” the econet, by converting portions of the string into corresponding characteristics. For example, initial connection weights in the econet’s neural network (not shown here) are encoded as eight-bit segments in the binary string.

Note that other physical features, such as size and shape, might also be encoded in the same way. By repeating this process a number of times, as Figure 1B illustrates, an entire population of econets is created. Because each is created with a different binary string, the first generation of econets is composed of unique individuals.

The next step in the modeling process involves evaluating the fitness of each individual. This is accomplished by placing each econet in its environment—in this case, a simple world with one object (see Figure 1C)—and giving it several trials to reach for the target object. At the start of each trial, both the econet and the object are randomly positioned. The econet's movements are determined by its neural network, which receives visual, proprioceptive, and tactile sensory inputs and generates eye and arm movements. Trials end when the econet either reaches the target, or fails to reach it after a predetermined number of movements. After each successful trial the econet's fitness is increased by 1.

Learning is accomplished in the model by selecting the most successful econets in each generation (i.e., those with the highest fitness), and making multiple copies of each to produce the next generation. Occasional copying errors (i.e., mutations) in the offsprings' binary strings introduce variations in behavior, with the possibility that some offspring will be more successful than their parents at reaching. By repeating these steps in an iterative fashion, we eventually obtain a population of econets that are successful at reaching efficiently.

Box 2. Baillargeon's "car study" (Baillargeon, 1986; Baillargeon & DeVos, 1991) was designed to show that infants are able to mentally represent hidden objects and their properties. Figure 2A illustrates the general sequence of events in this study. Infants begin by watching the habituation event, in which (a) a screen is lifted and then lowered, and then (b) a car appears and rolls down a track, passing behind the screen. After several repetitions of the habituation event, infants watch two test events. During both the possible and impossible test events, a box is now revealed when the screen is lifted; the box is behind the track during the possible event, but on the track during the impossible event. During both events the car rolls along the track as it did during the habituation event.

Baillargeon found that young infants looked significantly longer at the impossible event, and interpreted this result as evidence that infants were surprised to see the car reappear when its path was obstructed by the box. She concluded that infants not only knew that the box and car existed while occluded, but also knew that both objects could not be in the same place at the same time.

Our work raises the possibility of another interpretation. Specifically, infants might look longer at the impossible event because the location of the box (i.e., in the car's trajectory) draws the infant's attention, causing them to scan the event in a novel way, and therefore making the impossible event more salient or interesting.

We investigated this alternative account by modeling infants' gaze patterns during the car study. First, we created computer-animation events that corresponded to those in the original car study. Figure 2B illustrates selected frames from these animation sequences. Second, we designed a multi-layer network, which "watches" the habituation event and learns to track the movements of the car.

As Figure 2C illustrates, each animation frame is processed into two input streams. In the "peripheral" stream, a low-resolution pattern is obtained by averaging large areas of pixels in the animation frame. In the "fovea" stream, a high-resolution pattern is obtained by sampling a small region of the input while preserving the pixel values from the animation frame. In addition, a third input identifies the center of the fovea stream (i.e., the fixation point, indicated by a green x in the animation frame) in Cartesian coordinates. Input values are then propagated forward through the network, generating an eye-movement (i.e., a shift in the fixation point); the two banks of output units correspond to horizontal and vertical shifts, respectively, which are superimposed.

During training, the model is rewarded for keeping the car within the fovea. Specifically, a reinforcement-learning algorithm is used to modify the connection weights in the network, increasing the likelihood that the model will track the car (for additional details, see Schlesinger, 2003). After training, the model is then presented with the possible and impossible test events (see Figure 2B).

Our gaze-direction model generates a number of empirical predictions concerning infants' gaze patterns during the test events. First, the model predicts that overall, infants will gaze most toward the center (i.e., toward the screen), with fewer fixations toward the right and the left (i.e., toward the car's exit and entry points, respectively). This prediction was confirmed in a replication study of 6-month-old infants (Schlesinger & Casey, 2003).

Second, the gaze-direction model also predicts that infants will scan the possible and impossible events differently. In particular, when infants increase their looking time during the impossible event, most of these "additional" fixations (i.e., those over and above the looking time spent in the possible event) should be toward the center of the display. Again, this prediction was also supported by the evidence (Schlesinger & Casey, 2003).

These results suggest that infants' gaze patterns may not only provide a more detailed picture of their reactions to possible and impossible events, but they may also be driven in part by perceptually salient aspects of the display. Our long-term goal is to continue to develop a modeling approach that includes both perceptual-processing and knowledge-based systems, which will allow us to ultimately tease these two levels of explanation apart.

Box 3. A primary strength of CALP is its versatility and flexibility. In particular, any given model can easily be modified to study a new developmental question. Take as an example our econet reaching model (see Box 1), which we reconfigured to address the problem of acquiring proto-linguistic signaling systems (Parisi, Cangelosi, & Falcetta, 2002).

Imagine a population of artificial organisms with a single arm that lives in an environment which contains different objects that they have to reach or to push/pull in order to survive and reproduce. As Figure 3A illustrates, each individual also has an "ear" which can receive auditory signals. Using a genetic algorithm, the organisms learn in a succession of generations to execute the appropriate action on the appropriate object in response to both the particular object(s) that they see and the particular linguistic signal(s) that they hear. An organism's behavior is controlled by a neural network (not shown here) with (a) visual input units encoding the perceptual properties of the objects, (b) proprioceptive input units encoding the current position of the organism's arm, and (c) and auditory input units encoding heard linguistic signals, plus output units encoding movements of the arm.

In one simulation (Figure 3B) the organisms each time see a single object which can belong to one of two categories and they have to push the objects that belong to one category and to pull the objects that belong to the other category, where the objects are encoded in the input visual units as deviations from two prototype binary patterns. The addition of two linguistic signals that co-vary with the two categories makes the organisms' performance more efficient. However, in this simulation the signals cannot be classified as either nouns or verbs since they co-vary at the same time with both the action to be executed and the category of the object on which the action is to be executed. This is what the first signals understood or produced by children are likely to be.

In a second simulation (Figure 3C), the organisms see two objects at the same time, each belonging to one of the two categories and they learn to reach one object if they hear one signal and the other object if they hear the other signal. The two signals can be interpreted as proto-nouns if nouns are defined as signals that direct action (or attention) to a specific portion of the current input.

In a third simulation (Figure 3D), a single object is presented and the organisms learn to push the object in response to one signal and to pull the object in response to another signal, whatever the object. These signals can be interpreted as proto-verbs if verbs are defined as linguistic signals that co-vary with the action to be executed independently of the object on which the action is executed.

Finally, in a fourth simulation (Figure 3E) two objects and two signals are presented at the same time and the organisms learn to push or pull one of the two objects in response to the first signal (verb) and they execute the action on one or the other object in response to the second signal (noun). These sequences of two signals can be interpreted as proto-sentences made up of one verb and one noun.

Figure Captions

Figure 2. The reaching model. (A) A binary string is used to generate the physical characteristics (i.e., phenotype) of each organism. (B) The first generation is composed of a population of organisms, each created by a random binary string. (C) Each organism is evaluated by placing it in the environment, and testing its ability to make contact with a target object. (See Box 1 for details)

Figure 2. The tracking model. (A) Baillargeon (1986) investigated infants' reactions in the "car study", in which a car rolls briefly behind a screen; infants watch the habituation event, followed by the possible and impossible test events. (B) The tracking model is presented with analogous, computer-animated events. (C) The architecture of the tracking model's oculomotor control system; each frame of animation is processed along parallel periphery and fovea input streams, leading to the production of a shift in the fixation point. (See Box 2 for details)

Figure 3. The language model. (A) As in the reaching model, a binary string encodes the characteristics of each organism; in addition to one eye and arm, each organism also has an "ear", which receives auditory inputs. (B) The organism learns to push objects of one category and pull objects of another category, which are presented at different times. (C) The organism learns to reach one object or another when presented at the same time. (D) The organism learns to push or pull a single object. (E) The organism learns to push one object or pull another when both are present. (See Box 3 for details)

Figure 1

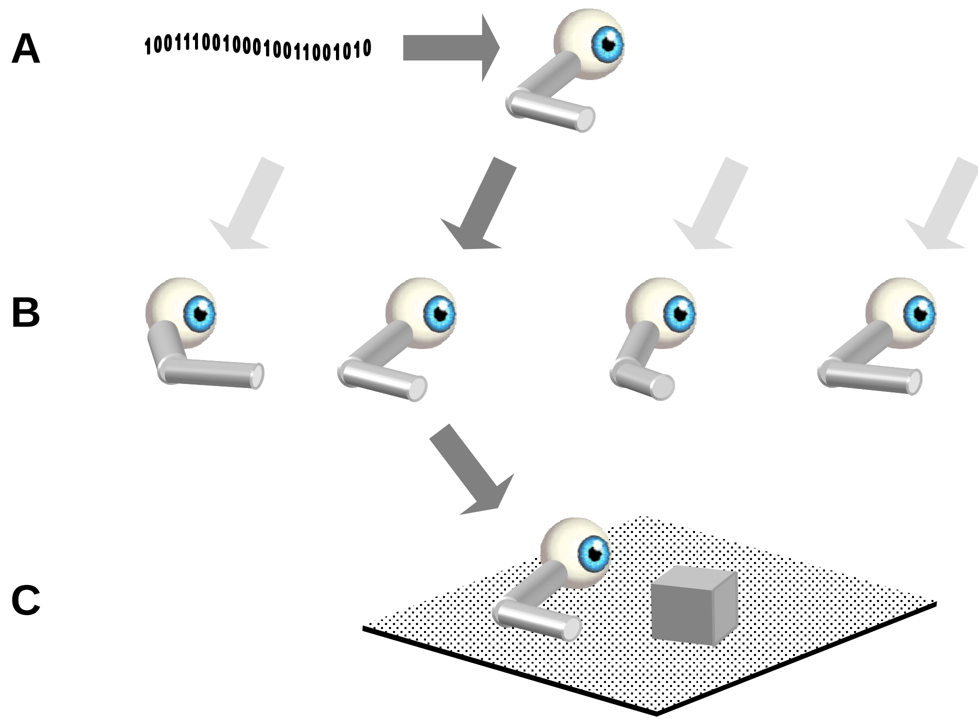

Figure 2

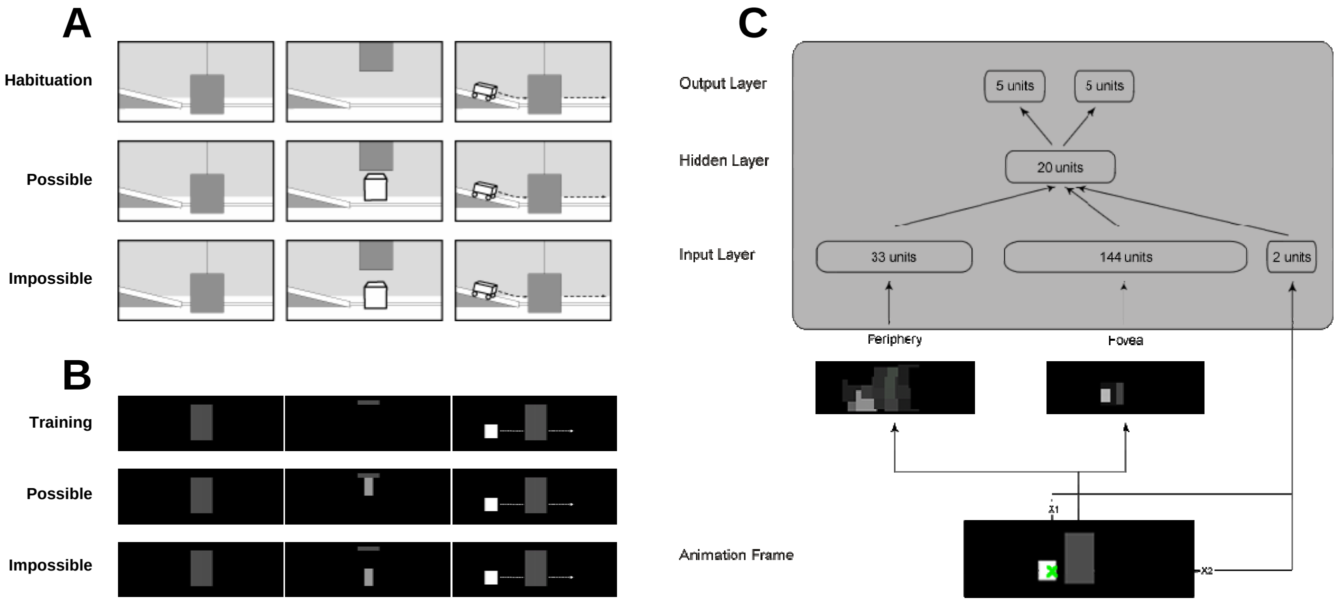
A
Habituation
Possible
Impossible
B
Training
Possible
Impossible
C
Output Layer
Hidden Layer
Input Layer
5 units
5 units
20 units
33 units
144 units
2 units
Periphery
Fovea
X1
X2
Animation Frame

Figure 3