Toward a New Grand Theory of Development?
Connectionism and Dynamic Systems Theory Re-Considered

J.P. Spencer, M. S. C. Thomas, & J. McClelland (Editors).
To be published by Oxford University Press.

Contributors and Book Structure

Part I: Overview
Introduction: Spencer, Thomas & McClelland
Chapter 1: Jay McClelland—overview of connectionist approaches **[Note: Rough draft only]**
Chapter 2: Gregor Schöner—overview of dynamic systems approaches

Part II: Case Studies
Chapter 3: Karl Newell, Liu & Mayer-Kress
Chapter 4: Whit Tabor
Chapter 5: John Spencer, Jeff Johnson & John Lipinski **[Note: Outline only]**
Chapter 6: Matthew Schlesinger
Chapter 7: Yuko Munakata & Bruce Morton
Chapter 8: Linda Smith **[Note: Abstract only]**
Chapter 9: Daniela Corbetta
Chapter 10: Han van der Maas & Maartje Raijmakers
Chapter 11: Denis Mareschal, Robert Leech & Rick Cooper
Chapter 12: Bob McMurray, Jessica Horst & Larissa Samuelson

Part III: Reactions from the "outside"
Chapter 13: Guy Van Orden & Heidi Kloos
Chapter 14: Lisa Oakes, Nora Newcombe & Jodie Plumert
Chapter 15: Tim Johnston & Bob Lickliter

Part IV: Reactions from the "inside"
Chapter 16: Jay McClelland
Chapter 17: Gregor Schöner
Chapter 18: Kim Plunkett
Chapter 19: Michael Thomas & Fiona Richardson
Chapter 20: Paul van Geert & Kurt Fischer

Conclusion: Thomas, Spencer & McClelland

Connectionist Models of Development:
Mechanistic Dynamical Models with Emergent Dynamical Properties

James L. McClelland and Gautam Vallabha

The symbolic paradigm of cognitive modeling, championed by Minsky and Papert, Newell and Simon, and other pioneers of the 1950's and 1960's, remains very much alive and well today. Yet an alternative paradigm, first championed in the 1950's by Rosenblatt, in which cognitive processes are viewed as emergent functions of a complex stochastic dynamical system, has continued to have adherents. A cooling of interest in such approaches in the 1960's did not deter Grossberg (1973, 1976) or James Anderson (1973), and the approach emerged again full force in the 1980's. By the end of that decade, many others had begun to explore the implications of this paradigm for development (McClelland, 1989; Plunkett and Marchman, 19??). A parallel and closely related movement, emerging from the physical sciences, also began to gain adherents during the same decade (Schoner & Kelso, 1988) and began to attract the interest of developmental psychologists in the early 1990s (Thelen and Smith, 19??).

The present chapter represents our attempt to underscore the common ground between connectionist and dynamical systems approaches. Central to both is the emergent nature of system-level behavior and changes to such behavior through development. Essential to this exploration of emergentism is a framework that allows consideration of the dynamics of state variables over time.

Mechanistic vs Emergent dynamics

In what follows, we will distinguish between mechanistic and emergent dynamics. This distinction parallels a distinction offered in their seminal article on dynamical systems by Schoner and Kelso (1988; Schoner, this volume). Importantly, for both connectionist models and the models of the type investigated by Schoner and Kelso (1988), the dynamical systems involved can be described at two different levels: A macroscopic or emergent level, and a microscopic of mechanistic level. This differentiates both approaches from some others in which the focus has been restricted to a descriptive characterization on the macroscopic dynamics.

For connectionist models, the state evolutions at the emergent, macroscopic, or systems level are determined by several things, among them the mechanistic microdynamics specified by the modeler and used in simulations. Other crucial factors are the network architecture and the ensemble of experiences on which the network is trained. These state evolutions exhibit emergent properties congruent with those postulated by Dynamical Systems Theory (DST), as articulated by Schoner and Kelso (1988, Schoner, this volume).

To give a few examples: The cascade model (McClelland, 1979) describes the time-evolution of activation across the units in a linear, feed-forward network in terms of a simple differential equation:

$$\Delta a_i/dt = \lambda(a_i - net_i)$$

In words, the change in activation of a unit is driven by the difference between its current activation and the net input it receives from other units. The purpose of the model was to explore the consequences for overall system behavior of changes in parameters of these equations at the mechanistic or microstructural level. The analysis reveal, for example, that changes in the rate constants at two different layers in a cascaded, feedforward network had additive effects on reaction time at the systems level.

More complex equations are often used in other models (e.g., Grossberg, 1976)

$$\Delta a_i/dt = \alpha E(M-a_i) + \gamma I(a_i-m) + \gamma(a_i-r)$$

Grossberg's formulation separates the excitatory input E and the inhibitory input I, and lets each drive the activation up or down by an amount proportional to the distance from the current activation to the maximum M or minimum m; also operating is a restoring force driving the activation of units back toward rest r; each force has its own strength parameter (α, β, and γ). A similar function was used in the interactive activation model (McClelland and Rumelhart, 1981). In his seminal work in the 1970's Grossberg examined in detail the conditions under which these microdynamic equations gave rise to stable emergent behaviors such as attractors and various other important emergent structures (Grossberg, 1978).

But these equations simple express the dynamics of a connectionist network as it responds to an input in real time. Typically, such equations are used to model activation processes that are assumed to be complete in under a second, such as, for example, the process involved in settling on a perceptual interpretation of a visual input, such as a line drawing or a word. But hidden in the equations, as presented thus far, are other dynamical variables that really form the heart of these models: their connections. For example, the *net* input in our first equation above represents the sum, over all connections projecting to a given receiving unit *r*, of the activation of the unit sending the projection, *s*, times the weight to *r* from *s*:

$$net_r = \Sigma_s w_{rs} a_s$$

And similarly the *E* and *I* in Grossberg's equation are essentially the sum of the excitatory inputs (mediated by excitatory connections) and inhibitory inputs (mediated my inhibitory connections). While some units receive direct stimulation from the outside world, most units receive their input from other units, and thus, what determines their activation is the the values of their incoming connection weights.

According to the connectionist approach to cognitive development, it is largely the time evolution of the strengths of the connection weights that governs the process of change over developmental time. Connectionist 'learning rules' are essentially differential equations that describe the changes of these connections. For example Linsker suggested that the connection weight to a receiving neuron r from a sending neuron s might obey a Hebbian-type learning rule:

$$dw_{rs}/dt = \varepsilon(ar - \Theta_r)\ (as - \Theta_s) - \beta w_{rs}$$

Here a_r and a_s are the activations of the receiving and sending units, β is a constant regulating a tendency for weights to decay toward 0, ε is the learning rate, and Θ_r and Θ_s

are critical values that may themselves be adaptive (another word for ‘governed by mechanistic differential equations indicating how the variables change as a function of their own or other variables’). As with activation functions, many variations have been proposed for connectionist learning rules. What is crucial for the moment is that the time-evolution of the connections in the system --- the parameters which essentially determine both the outcome and the time-course of processing --- are themselves defined in terms of differential equations. The emergent results of these connection dynamics are emergent cognitive structures such as feature detectors, phonetic and other types of categories, and behavioral competencies, as well as emergent dynamical trajectories of change in such properties.

Within PDP approaches to development, the mechanistic and emergent dynamical properties of connectionist models are relevant at two time scales. On a short time scale, connection weights are usually treated as fixed, contributing along with the architecture of the system and the specified mechanistic dynamics of the activation process to the emergent time-course and outcome of a single act of information processing and/or behavior, such as reaching to one of two locations in which an object may have been hidden by an experimenter. On a longer time scale, connection dynamics within the same system are relevant to the gradual change in the way the system responds to inputs over developmental time, usually measured in months or years. Many connectionist models simplify the activation dynamics in ways that may be misleading about the underlying theory. For example, in most connectionist models of past-tense verb inflection or single word reading, an input pattern is propagated forward through one or more layers of connection weights (and 0 or more hidden layers of units) to an output layer. This output represents the asymptotic state that would be achieved in an identical cascaded feedforward network or even a recurrent network (see Plaut et al, 1996, for an explicit comparison of one-pass feedforward and recurrent versions of the a network for single-word reading). Likewise, while intrinsic variability is assumed always to be at work in processing (McClelland, 1991, 1993), it is often left out of specific models for simplicity. A similar approach is often taken in simple recurrent networks (Elman, 1990), which can be viewed as chunking time into relatively large-scale, discrete units, and setting the state for the next discrete unit based on the state of the one before. While the networks look discrete, this is to be understood as an approximation that increases the feasibility of simulation; allowing a network to settle iteratively over many time steps multiplicatively scales the time it takes to run a developmental simulation.

Since the focus of many connectionist models that have been applied to development has been on change over developmental time, the emphasis tends to fall on the time-evolution of the connections. Indeed, within multi-layer connectionist networks, this time evolution can have complex dynamical properties, undergoing both accelerations and decelerations. This can occur, even though the connectionist learning rule looks simple and does not itself undergo any change; thus there are emergent, system-dependent dynamics that depend in complex ways on the state of the connections. We shall consider such issues in detail later in our paper. For now the essential point to make is this: *Connectionist systems are dynamical systems, and like other dynamical systems they can have complex emergent properties which themselves may be described by dyamical variables – rates of change with respect to time.*

In what follows, the points above will be elaborated through the presentation of a series of models. We begin with the recent model of activation dynamics proposed by Usher and McClelland (2001), showing how even a very simple architecture can give rise to interesting emergent dynamical properties (including attractors and path-dependence) which we will treat collectively as the emergent response dynamics of the connectionist system. We will then consider a more complex model proposed by Vallabha and McClelland (in press) which embeds the Usher and McClelland activation dynamics within a network whose connections also evolve according to a simple differential equation or learning rule, allowing it to learn in response to experience and exhibit interesting developmental dynamics in the domain of acquisition of phonological distinctions relevant to both first and second languages. Further emergent dynamic properties will be explored within this context. Finally we will consider a model of by Rogers and McClelland (2004) which addresses the time evolution of conceptual category representations. This model simplifies out the real-time activation / response dynamics (as in the past-tense verb inflection and reading models mentioned above) to address the gradual evolution of semantic category representations over the first ten years of life. Once again, simple learning rules give rise to complex emergent dynamical properties, now seen at a developmental time scale. Overall we hope to bring out how these models instantiate many of the properties of the dynamical models that have been proposed by many of the other contributors to this volume.

Short-Term Dynamics in a Connectionist Model of Perceptual Classification

We begin with Usher and McClelland's (2001) model of the time-course and accuracy of responding in speeded perceptual classification tasks. The goal of this model was to provide an account of classification responses, including time-accuracy tradeoffs and reaction-time distributions, as a function primarily of the difficulty of the classification. Its architecture is very similar to that used in PDP models of reading (McClelland and Rumelhart, 1981) and speech perception (McClelland & Elman, 1986; McClelland, 1991), and incorporates the basic principles of graded, interactive and stochastic processing. At its simplest level, the model consists of a layer of units, in which each unit corresponds to one response (Figure 1a). Each unit *i* receives excitatory activity ρ_i from an input layer that represents the external evidence or "support" for that particular response. In addition, each unit has an excitatory connection to itself and inhibitory connections to all other units in that layer. The key premise of the model is that each unit takes up the external evidence in a cumulative and noisy manner while competing with the other units. It is important to keep in mind that this is an abstract description of a system of mutual constraints that does not impose a specific division between perceptual and motor processing. For example, each "unit" may be instantiated as a spatially dispersed pattern of neural activity that encompasses both perceptual and motor systems.

The dynamics of this system are governed by a set of stochastic nonlinear differential equations:

$$dx_i = [\rho_i - \lambda x_i + \alpha f_i - \beta \sum_{i \neq j} f_j] \cdot dt + \xi_i \sqrt{dt} \qquad (1)$$

where ρ_i is the external input to unit *i*, x_i is the unit's instantaneous net input (its "current"), f_i is the instantaneous output (its "firing rate"), λ is the spontaneous decay rate of the unit's net input, β is the strength of the inhibitory connection between units, and ξ_i is the standard deviation of the integration noise. For simplicity, Usher and McClelland assumed that $x_i = \max(x_i, 0)$, and that $f_i = x_i$ which reduces Eq. 1 to the following set of linear equations:

$$dx_i = [\rho_i - kx_i - \beta \sum\nolimits_{i \neq j} x_j] \cdot dt + \xi_i \sqrt{dt} \qquad (2)$$

where $k \equiv \lambda - \alpha$. It should be noted that while Eq. (2) is linear, the overall system is not because of the floor on the activity of x_i. The behavior of this system can be understood by considering a network with two units, with the added constraint that $\rho_1 + \rho_2 = 1$. The state of the network then depends on three parameters: k, the net amount of activity leakage, β, the inter-unit inhibition level, and ρ_1, the external input to unit 1. Figure 1b shows the dynamics of the units in two sample regimes (both with k=0.2 and ρ_1=0.52). With β=0.0, there is no inhibition, so the two units find a balance between the leakage current and the sustained external inputs. This results in a slight advantage for x_1 because $\rho_1 > \rho_2$. With β=0.4, the lateral inhibition results in competitive interaction that depresses both x_1 and x_2. However, once x_1 overcomes x_2 (at $t \approx 11$s), its rate of increase is similar to that in the β=0.0 condition.

The main point to take from Figure 1b is that the system's dynamics tend toward a stable state, with the precise nature of that state being governed by the three parameters. In order to map these stable states, we consider the dynamics of $\upsilon \equiv x_1 - x_2$. For each parameter setting, we numerically integrated Eq. (2) for 25s (ξ_1=ξ_2=0), for initial conditions ranging from $\upsilon = -5.0$ to $\upsilon = +5.0$. From these trajectories, we calculated $d\upsilon/dt$ and estimated the locations of the fixed points of υ for that parameter setting. Figure 2 shows the resulting bifurcation diagrams for various values of ρ_1, k, and β. The diagrams show that the system has a wide variety of dynamics. For $k \leq 0$, there is no leakage, so both units' activity continually increase (the sign of υ is determined by the unit with the faster increase). As a result, the system only has unstable fixed points. Once $k > 0$, the net leakage dominates the external inputs and restricts the overall levels of υ. This regime is dominated by the value of β, and its effect is to introduce *bistability* into the dynamics. (see for example, k=0.2, β=0.4, $0.35 < \rho_1 < 0.65$). Here, the final state of the system is dependent on the initial condition υ_0, with preference for υ^+ gradually increasing with ρ_1. If the inhibition is increased (e.g., β=0.6), υ_0 is more likely to overcome ρ_1, and consequently the bistable regime is wider. The bistability also implies that the system is capable of *hysteresis*. Say ρ_1

is initially 0.0, so that system settles at υ^-. If ρ_1 is then gradually ramped up, the system will stay at υ^- until the end of the bistable regime, at which point it will snap to υ^+. If ρ_1 is then gradually decreased, the system will stay at υ^+ during the bistable regime, indicating path dependence in the system dynamics.

A key point to keep in mind here is that the U-M model was not designed to produce the dynamics in Figure 2. Its purpose was to account for ensemble statistics of categorization performance using the principles of interactive, stochastic and graded processing, and the dynamical properties fell out as one consequence of these principles. Furthermore, if Eq (2) is augmented with an additional depression term (so that a unit is less able to compete if it has recently been active), its dynamics become remarkably similar to those of an explicitly dynamical model of categorization (Tuller, Case, Ding and Kelso, 1994; Ditzinger, Tuller and Kelso, 1997). This kinship suggests that the PDP approach (as instantiated in the U-M model) and the dynamical approach (as instantiated through the dynamic field or synergetic approaches) may are in fact closely related approaches to the characterization of response dynamics.

Developmental Dynamics

There are two broad ways in which developmental dynamics have been explored within the connectionist approach, one based on supervised error-driven training and the other based on unsupervised Hebbian learning. The power of the former method is that it can re-represent the input in sophisticated ways to match the task at hand (for example, a network may assign finer-grain representations to salient inputs and coarser-grain ones to unimportant or rare inputs). However, in many cases it is not plausible to assume that outcome information is consistently available, and unsupervised learning based on Hebbian learning can provide insight into such problems (Rumelhart & Zipser, 1985; Petrov et al., 19xx). In addition, there is substantial evidence for the biological plausibility of Hebbian learning (Ahissar, 1998; Syka, 2002). Finally, Hebbian learning with bidirectional weights typically results in symmetric connections that facilitate interactive attractor dynamics (Grossberg, 19xx).

The relation between Hebbian learning and developmental dynamics was explored in a model proposed by Vallabha and McClelland (submitted). The motive for the model was to provide an account for the initial acquisition of speech categories in infancy, the establishment of such categories as attractors, the role of such attractors in creating difficulty in learning new speech categories in adulthood, and how acquisition is affected by different training conditions. Here we present a simplified version of this model that focuses on the emergence of a "perceptual magnet effect" (Kuhl, 1991) as a result of perceptual experience. The effect is marked by a decrease in discriminability between adjacent stimuli as they get closer to the category prototype, and it is developmental in that it only shows up for sound categories that are distinguished in the native language (Iverson et al., 1996). We will focus on how exposure to native like sound structure results in the formation of perceptual categories, and how these in turn affect discrimination, producing a perceptual magnet effect. Below we describe how the Vallabha-McClelland model addresses this issue.

The model consists of a two layers of units: L1 and L2 (Figure 3), with 80 units in L1 and 2 units in L2. Each unit has excitatory connections with every unit in the other layer, and inhibitory connections with all other units in its layer. The pattern of activity over L1 is taken to be the "perceptual representation", and the activity in L2 is taken to be the "category representation", with each unit standing for a distinct response category. The dynamics for the units are similar to those for the U-M model:

$$\begin{aligned} dx_i &= [\rho_i - x_i + \textstyle\sum_j wscale_{ij} \cdot w_{ij} f_j] \cdot dt + \xi \sqrt{dt} \\ f_i &= \min\big(0, \tanh(x_i \cdot gain_{act})\big) \end{aligned} \tag{3}$$

where w_{ij} are the incoming weights to unit i, $gain_{\mathrm{act}}$ is the gain of the activation function, and $wscale_{kj}$ is a “weight scaling” parameter set to 5.0 for L2→L1 weights and to 1.0 for all other weights (it simulates the effect of a group of similarly-wired L3 units that act in concert).

The external inputs consisted of bumps of activity presented to the L1 units. The center of the bump was the "input location x", with $\rho_i \equiv 0.8 \cdot \exp(-(i - x)^2 / 17)$. The input locations were drawn from two Gaussian distributions, $N(29,3)$ and $N(51,3)$, with 400 samples from each distribution. On each trial, the network was allowed to settle for 30 time steps with dt

= 0.2, $\xi = 0.04$, and $gain_{act} = 0.5$, with the external inputs being presented the entire time. Once the settling was completed, the weights between L1 and L2 were updated using a Hebbian rule, $\Delta w_{ij} = \eta \cdot f_i \cdot f_j$. Following the update, the weights were multiplicatively normalized in a graded manner, i.e., for small weight-vector magnitudes there was no normalization, but as the weight-vector magnitude increased, the normalization was applied with greater force such that the maximum magnitude was 1.0. In order to ensure that the L1→L2 and L2→L1 weights are approximately symmetric, the normalization was done over incoming weights for L1→L2 projections, and over outgoing weights for L2→L1 projections (Grossberg, 1988).

The above dynamics resulted in competitive learning between the units in L2 (cf. Rumelhart & Zipser, 1985; Carpenter & Grossberg, 1987). For example, one L2 unit developed strong connections to L1 units around input location 29 (the center of one of the input distributions), with more active input locations having higher connection strengths than less-active locations. In doing so, it inhibited the other L2 unit from becoming sensitive to those same input locations. Now consider what happens when an input stimulus is presented to the network. The stimulus causes a small bump of activity on L1. Due to the within-L1 interactions, this bump coalesces and becomes more prominent. Concurrently, it activates the L2 unit that is sensitized to that input location. The L2 unit recurrently excites the input locations that are most-frequently active for that category, i.e., it excites the "prototypical" L1 representation for that category. As a result, the L1 activity becomes skewed towards the category prototype. Thus, the categorization ability of the network (reflected in L2 activity) changes the perceptual representations (reflected in L1 activity), and shapes the discrimination of input stimuli.

Figure 4 illustrates the above process by showing the evolution of L1 activity at different stages of training. The key point to note is the *skew* of the L1 activity at the final time step of the settling. Before training, the skew is negligible. After 1500 weight updates, the final activities are noticeably skewed toward the center of the category. A consequence of this skew is that the final L1 activities for adjacent input stimuli (e.g., at input locations 25 and 27) become more similar to each other. If this similarity, as measured by overlap or by Euclidean distance, is used as a measure of discriminability, then we get the "perceptual magnet effect": after extensive exposure to exemplars of a category, more prototypical stimuli are harder to distinguish than less prototypical ones (because the amount of skew is greater with the former than with the latter).

The developmental dynamics of the learning may be visualized through the relation between the input location and the "amount of skew". We calculated the skew as follows. For each input location *k*, we presented an input centered at *k* and ran the network for 30 timesteps. Then we took the final L1 activity vector $y^{(k)}$ and calculated its "center of mass" c_k:

$$c_k = \sum_{i=1}^{80} i \cdot \left(y_i^{(k)} / \Sigma_j y_j^{(k)} \right) \qquad (4)$$

The amount and direction of skew are indexed by $c_k - k$ (rightward and leftward skews are indicated by positive and negative values, respectively). Figure 5a shows the skew over the input space at different stages of training. It can be seen that the centers of the distributions (locations 29 and 51) function like a dynamical attractor and the center of the space (input location 40) functions like an unstable fixed point. If we calculate the skew in L1 representation at each timestep (rather than just the final timestep), an interesting pattern emerges. Figure 5b shows the skew for input location 26 as a function of processing stage (the number of timesteps) *and* the developmental stage (the number of weight updates). The effect of training is to accelerate the response dynamics – a skew of 0.5 took 16 timesteps to develop after 1500 updates, 13 timesteps after 2000 updates, and only 10 timesteps after 4000 updates.

Figure 5b suggests how the response and developmental dynamics may be linked. A system without learning (such as the Usher-McClelland model) displays a rich variety of response dynamics (Figure 2). Some of these dynamics may be particularly relevant for accomplishing a task such as categorizing an input stimulus. The developmental dynamics may be seen as facilitating (or entrenching) just those task-relevant dynamics in the system. Upon this view, response and developmental dynamics are different timescales of activity that can coexist in the same system. Furthermore, these two scales are linked – the developmental effects facilitate certain response dynamics, which in turn shape further development – leading to a kind of circular causality. One consequence of this linkage is that the developmental changes need not be imposed on the system (by stipulating a change in learning rate or in the lateral-interaction kernel, for example), but can rather emerge from the operation of the system.

Developmental Dynamics in Models with Simplified Short Time-Scale Dynamics

We now turn to a model introduced by Rumelhart (1990; Rumelhart & Todd, 1993) and studied in detail by Rogers and McClelland (2004). This model is one of those that simplifies the short-time scale dynamics down to a single pass through a feed-forward network, to concentrate on the emergent micro-dynamics of the system. The model is illustrated in Figure X. We describe the model and the processing that occurs in it at a level of detail that we hope is sufficient to allow readers not already familiar with such networks to understand the basics of their operation. Additional details of the implementation and of the training patterns are presented in Rogers and McClelland (2004).

[FROM HERE DOWN THE TEXT IS CURRENTLY AN UNEDITED EXURPT FROM ROGERS AND MCCLELLAND, AND NO FIGURES ARE PROVIDED. THIS TEXT WILL BE SHAPED AND CONDENSED AND SHOULD NOT BE TREATED AS FINAL. IT IS PROVIDED TO GIVE A FLAVOR FOR WHAT WILL BE COVERED IN THE FINAL VERSION OF THIS ARTICLE].

Rumelhart’s initial goal was to demonstrate that the propositional content contained in a traditional taxonomic hierarchy could also be captured in the distributed representations acquired by a PDP network trained with backpropagation, and that the network could support the kinds of inferences that can be performed in Quillian’s hierarchical propositional network. Rumelhart used individual nodes (or *localist units*) in the network’s input and output layers to correspond to the consituents of the propositions—the items or concepts that occupy the first (*subject*) slot in each proposition, the relation terms that occupy the second (*relation*) slot, and the attribute values

that occupy the third slot. Each item is represented by an individual input unit in the layer labelled *Item*; different relations are represented by individual units in the layer labelled *Relation*; and the various possible completions of three-element propositions are represented by individual units in the layer labelled *Attribute*. When presented with a particular *Item* and *Relation* pair in the input, the network's task is to turn on the attribute units in the output that correspond to valid completions of the proposition. For example, when the units corresponding to *canary* and *can* are activated in the input, the network must learn to activate the output units *move*, *grow fly* and *sing*. The particular items, relations, and attributes are based on the information contained Figure 1.2. When the network has learned to correctly complete all of the propositions, it has encoded the same information stored in the propositional hierachy.

We have adapted Rumelhart's model a little to suit our own purposes. We treat the network as a simplified model of experience with objects in the world and of spoken statements about these objects. The *Item* layer is construed as providing a simplified proxy for an input representation of an object as encounted in experience; the units stand for the occurrence of the actual items themselves, not for their names, and in our work they always correspond to the items at the bottom level of the taxonomic hierarchy in Figure 1.2. The relation layer is viewed as a simplified specification of the context in which the item is encountered. The attribute layer is thought of as representing the predicted consequences or sequelae following from the occurrence of the object in the given context. The completion of a simple three-term proposition about the object is an example of what we have in mind: given "canary can ," the possible sequelae include "grow," "move," "fly" and "sing." In this approach, the four propositional relations from Quillian's hierarchy are viewed as distinct contexts. The *ISA* relation corresponds to a naming or explicit categorization context, in which the object's name or category label is of relevance, as these might be indicated verbally (e.g., by a sentence such as "This is a bird" or "This is a canary"). The *can* relation corresponds to a context in which the behaviors of the object might be observed; the *is* relation corresponds to a context in which its appearance properties are highlighted; and the *has* relation corresponds to a context in which its parts are highlighted.

The network consists of a series of nonlinear processing units, organized into layers, and connected in a feed-forward manner as shown in the illustration. Patterns are presented by activating one unit in each of the *Item* and *Relation* layers (i.e. these activations are set to 1 and activations of all other input units are set to 0). Activation then feeds forward through the network, modulated by the connection weights. Activations are updated sequentially, layer by layer, so that first the representation layer is updated, then the hidden layer, then the attribute layer. To update the activation of a unit, first its net input is calculated: the sum, over each of the unit's incoming connections, of the activity of the unit sending its activation through the connection, multiplied by the value of the connection weight. The net input is then transformed into an activation according to the logistic activation function shown in Figure 2.3.

Initially, the connection weights in the network have small random values, so that the activations produced by a given input are weak and random in relation to the target output values. In order to perform correctly, the network must find a configuration of weights that will produce the correct target state across the output units for a given item-relation input. Each target state consists of a pattern of 1s and 0s like the one shown for the input *canary can* in Figure 2.2 — the target values for the black units are 1 and for all other units they are 0.

To find an appropriate set of weights, the model is trained with the backpropagation algorithm already mentioned (Rumelhart et al., 1986). Backpropagation is an inherently gradual learning process, which depends on the presentation of many training examples. When the environment is characterized (as in this case) by a relatively small and finite set of training examples, each example is presented multiple times, interleaved with presentations of the other examples. Training consists of a set of epochs, each encompassing the presentation of every item in the training set. There are several possible variants of the general training method. The approach we have used involves presenting each example once per epoch, in random order, and adjusting the weights after every training example. We use this *on-line* connection-weight adjustment procedure because we imagine that the brain uses

a similar procedure: small connection weight adjustments are made after every experience, and these gradually accumulate to form the basis of semantic knowledge and representation.

For each presentation of a training example, the process works as follows. First, the item and relation are presented to the network as inputs, and activation is propagated forward to the output units as described above. The observed output states are then compared to the target values, and the difference is treated as a measure of error at each output unit. Small adjustments to the connection weights throughout the network are made to reduce the error. A brief description of the procedure is given in the caption to Figure 2.3; fuller discussions including derivation of the algorithm are available in Rumelhart et al. (1986) and Hertz, Krogh, and Palmer (1991). What is important conceptually about the procedure is that it involves the propagation of the error information at the output backward through the (forward-projecting) connection weights in the network, so that the changes to each weight—including those from the input units to the representation layer—are dictated by the effect that they will ultimately have on the error at the output. The weight adjustments are scaled by a constant e, that must be small

to ensure progress (McClelland, McNaughton, & O'Reilly, 1995). Some adjustments cumulate across examples, while others cancel out. Overall the weights adapt slowly, yielding gradual evolution of the patterns of activation at each level of the network, and gradual reduction of error.

Although the model's inputs are localist (one unit for each item), each individual *Item* unit projects to all of the units in the layer labelled *Representation*. The activation of a single item in the model's input, then, generates a distributed pattern of activity across these units. The weights connecting the item and representation units evolve during learning, so the pattern of activity generated across the *Representation* units for a given item is a *learned internal representation* of the item. Though the model's input and target states are constrained to locally represent particular items, attributes, and relations, the learning process allows it to derive distributed internal representations that are governed by the structure of the domain.

For reasons we will elaborate in the next chapter, the learned representations in the Rumelhart network gradually come to capture the semantic similarity relations that exist among the items in the training environment. Indeed, as Rumelhart (1990; Rumelhart &

Todd, 1993) showed, the hierarchy explicitly represented by the ISA links in Quillian's network come to be represented in a latent form, in the similarity relations among the patterns the network gradually learns to assign to each of the items. Because each item is represented locally in the input, the model has no initial information about how the objects in its virtual world are related to one another. For example, the *pine* and *oak* inputs are no more similar to one another than each is to the *salmon* input. Due to the small, random values of the weights from these units to the *Representation* units, the patterns initially all have very similar internal representations, with only sight random differences. However, as the model learns to complete particular propositions, these representations gradually change, and as we will see later, items that differ on many attributes come to be differentiated, while items that share many attributes continue to be represented by similar patterns of activity across the *Representation* units. Thus, the connections that link *Item* and *Representation* units in the network can be viewed as encoding a set of semantic similarity relationships among a set of otherwise arbitrary markers. The connections in the rest of the network then reflect the learned mappings between these internal representations, in combination with input from the relation context, and the explicit properties coded in the output.

An important departure from other representational schemes (including those used in some other connectionist approaches) is that the internal representations acquired by the Rumelhart network are not lists of semantic features—in no way are they directly interpretable semantically. Individual units in the model's *Representation* layer do not encode the presence or absence of explicit, intuitive object properties. Rather, the representations are abstracted from the featural decomposition of objects represented in the output layer. The network's representations capture the similarities existing among different kinds of objects, not the actual semantic properties themselves. The individual semantic properties can only be recovered through the combined effects of the units in the distributed representation, working in concert with units in other parts of the network.

Progressive Differentiation

The progressive differentiation of semantic representations in the Rumelhart and Todd network was previously considered in a simulation reported in McClelland et al. (1995). Other investigators interested in differentiationof perceptual and conceptual representations in infancy have also presented similar simulations (Quinn & Johnson, 1997; Miikkulainen & Dyer, 1991; Schyns, 1991). Here we replicate the progressive differentiation simulation of McClelland et al. (1995), examine the process in detail and explore why it occurs, and then go on to consider deterioration of semantic representations within the same system.

We trained the network shown in Figure 2.2 with the same corpus of propositions used by Rumelhart and Todd (1993). These are all the true propositions derivable from the Quillian propositional hierarchy in Figure 1.2. A table listing the complete set of items and attributes is presented in Appendix B, Table B–1. The weights were initialized to small random values selected from a uniform distribution with a mean of zero and variance of 0.9.

With each pattern presentation, a single unit was activated in each of the *Item* and *Relation* layers, and activation was propagated through the network. The discrepancy between the pattern produced and the correct pattern was calculated, and weights were adjusted by a small amount (learning rate = 0.1) to reduce the sum-squared error across output units, using the backpropagation algorithm. The network processed each input-relation pair once in each training epoch, but the order of patterns within an epoch was randomized. In this simulation, the simplest version of backpropagation was used: the model was trained without noise, weight decay, or momentum. Weights were updated after each processing trial. To ensure that the model's output responses relied entirely on input from other units in the network, we assigned to all units in the model a fixed, untrainable bias of -2. Thus, in the absence of input, each unit's state would rest at approximately 0.19. We trained the network for 3500 epochs, at which point each output unit was within 0.1 of its target activation (0 or 1) on every pattern.

To see how internal representations develop in the network, we stopped training at different points during learning and stepped through the eight items, recording the states of the representation units for each. In Figure 3.1 we show the activations of the representation units for each of the eight item inputs at three points in learning. Each pattern of activation at each time point is shown using eight bars, with each representing the activation of one of the representation units. Initially, and even after 50 epochs of training as shown, the patterns representing the items are all very similar, with activations hovering around 0.5. At epoch 100, the patterns corresponding to various animal instances are similar to one another, but are distinct from the plants. At epoch 150, items from the same intermediate cluster, such as *rose* and *daisy*, have similar but distinguishable patterns, and are now easily differentiated from their nearest neighbors (e.g. *pine* and *oak*). Thus, each item has a unique representation, but semantic relations are preserved in the similarity structure across representations.

The arrangement and grouping of the representations shown in Figure 3.2 reflects the similarity structure among the internal representations, as determined by a hierarchical clustering analysis using Euclidean distance as the measure of similarity between patterns. At 50 epochs the tree is very flat and any similarity structure revealed in the plot is weak and random. By epoch 100 the clustering analysis reveals that the network has differentiated plants from animals: all the plants are grouped under one node, while all the animals are grouped under another. At this point, more fine-grained structure is not yet clear. For example, oak is grouped with rose, indicating that these representations are more similar to one another than is oak to pine. By epoch 150, the network has learned the correct similarities, and we can see that the hierarchical relations among concepts made explicit in Quillian's ISA hierarchy is fully captured in the similarities among the learned distributed representations.

In order to better visualize the process of conceptual differentiation that takes place in this model, we performed a multidimensional scaling of the internal representations for all items at 10 equally-spaced points during the first 1500 epochs of training. Specifically, the *Representation* layer activation vector for each item at each point in time was treated as a vector in an 8-dimensional space. The Euclidean distances between all vectors at all points over development were calculated. Each vector was then assigned a 2-d coordinate, such that the pairwise distances in the 2-d space were as similar as possible to the distances in the original 8-d space.

The solution is plotted in Figure 3.3. The lines trace the trajectory of each item throughout learning in the 2-dimensional compression of the representation state space. The labelled end points of the lines represent the internal representations learned after 1500 epochs of training, at which point the network is able to activate all of the correct output properties for all items above 0.6. The figure shows that the items, which initially are bunched together in the middle of the space, soon divide into two global clusters (plant or animal) based on animacy. Next, the global categories split into smaller intermediate clusters, and finally the individual items are pulled apart.

Discussion of Differentiation Results

Our simulation replicates the previous demonstration (McClelland et al., 1995) showing that when a backpropagation network is trained on a set of training patterns with a hierarchical similarity structure, it will exhibit a pattern of progressive differentiation. Similar demonstrations with other types of training sets have been made by others (e.g., Quinn & Johnson, 1997). One interesting aspect of this process is the tendency for the different levels to differentiate in relatively discrete stages, first completing differentiation at the most general level before progressing to successively more fine-grained levels of differentiation. This tendency to exhibit stage-like learning is a feature of connectionist models that has been considered extensively elsewhere (McClelland, 1989; Plunkett & Sinha, 1992; McClelland, 1994a). Our present task is to try to provide the reader with a mechanistic understanding of the progressive differentiation process, drawing on insights expressed in the papers just cited (see also McClelland & Rumelhart, 1988) to explain how stage-like progressive differentiation works in the present case.

With the training set used here, very early in learning, the network comes to represent all the animals as similar to one another, and as quite distinct from the plants. Only later does it come to learn to differentiate the patterns at an intermediate level, and only after that does it learn to differentiate the items from each other at the subordinate level. Why does this occur? To begin to gain an intuitive understanding of this, let us consider how the network learns about the following four objects: the oak, the pine, the daisy, and the salmon. Early in learning, when the weights are small and random, all of these inputs produce a similar meaningless pattern of activity throughout the network. Since oaks and pines share many output properties, this pattern results in a similar error signal for the two items, and the weights leaving the *oak* and *pine* units move in similar directions. Because the salmon shares few properties with the oak and pine, the same initial pattern of output activations produces a different error signal, and the weights leaving the salmon input unit move in a different direction. What about the daisy? It shares more properties with the oak and the pine than it does with the salmon or any of the other animals, and so it tends to move in a similar direction as the other plants. Similarly, the rose tends to be pushed in the same direction as all of the other plants, and the other animals tend to be pushed in the same direction as the salmon. As a consequence, on the next pass, the pattern of activity across the representation units will remain similar for all the plants, but will tend to differ between the plants and the animals.

This explanation captures part of what is going on in the early stages of learning in the model, but does not fully explain why there is such a strong tendency to learn the superordinate structure first. Why is it that so little intermediate level information is acquired until after the superordinate level information? Put another way, why don’t the points in similarity space for different items move in straight lines toward their final locations? Several factors appear to be at work, but one is key:

For items with similar representations, coherent covariation of properties across these items tends to move connections coherently in the same direction, while idiosyncratic variation tends to move weights in opposing directions that cancel each other out.

To see how this happens in the model, let us consider the fact that the animals all share some properties (e.g., they all can move, they all have skin, they are all called animals). Early in training, all the animals have the same representation. When this is so, if the weights going forward from the representation layer “work” to capture these shared properties for one of the animals, they must simultaneously work to capture them for all of the others. Similarly, any weight change that is made to capture the shared properties for one of the items will produce the same benefit in capturing these properties for all of the other items: If the representations of all of the items are the same, then changes applied to the forward-projecting weights for one of the items will affect all of the others items equally, and so the changes made when processing each individual item will tend to cumulate with those made in processing the others. On the other hand, weight changes made to capture a property of an item that is not shared by others with the same representation will tend to be detrimental for the other items, and when these other items are processed the changes will actually be reversed. For example, two of the animals (canary and robin) can fly but not swim, and the other two (the salmon and the sunfish) can swim but not fly. If the four animals all have the same representation, what is right for half of the animals is wrong for the other half, and the weight changes across different patterns will tend to cancel each other out. The consequence is that:

Properties shared by items with similar representations will be learned faster than the properties that differentiate such items.

The preceding paragraph considers the effects of coherent covariation in the weights forward from the representation later in the Rumelhart network. What about the weights from the input units to the representation layer? As previously stated, items with similar outputs will have their representations pushed in the same direction, while items with dissimilar outputs will have their representations pushed in different directions. The question remaining is why the dissimilarity between, say, the fish and the birds does not push the representations apart very much from the very beginning. The answer is somewhat complex, but understanding it is crucial, since it is fundamental to understanding the progressive nature of the differentiation process.

The key to this question lies in understanding that the magnitude of the changes made to the representation weights depends on the extent to which this will reduce the error at the output level. The extent to which change in the representation weights will reduce the error at the output in turn depends on whether the forward weights from the representation layer to the output are able to make use of any changes in the activations of the representation units. Their ability to make use of such changes depends on them already being at least partially organized to do so. Put in other words, we can point out a further very important aspect of the way the model learns:

Error back-propagates much more strongly through weights that are already structured to perform useful forward-mappings.

We can illustrate this by observing the error signal propagated back to the representation units for the *canary* item, from three different kinds of output units: those that reliably discriminate plants from animals (such as *can move* and *has roots*), those that reliably discriminate birds from fish (such as *can fly* and *has gills*), and those that differentiate the canary from the robin (such as *is red* and *can sing*). In Figure 3.4, we show the mean error reaching the *Representation* layer throughout training, across each of these types of output unit when the model is given the *canary* (middle plot). We graph this alongside measures of the distance between the two bird representations, between the birds and the fish, and between the animals and the plants (bottom plot); and also alongside of measures of activation of the output units for *sing*, *fly* and *move* (top plot). We can see that there comes a point at which the network is beginning to differentiate the plants and the animals, and is beginning to activate *move* correctly for all of the animals. At this time the average error information from output properties like *can move* is producing a much stronger signal than the average error information from properties like *can fly* or *can sing*. As a consequence, the information that the canary can move is contributing much more strongly to changing the representation weights than is the information that the canary can fly and sing. Put differently, the knowledge that the canary can move is more "important" for determining how it should be represented than the information that it can fly and sing, at this stage of learning. (The error signal for *move* eventually dies out as the correct activation reaches asymptote, since there is no longer any error signal to propagate once the model has learned to produce the correct activation).

Figure 3.4 also indicates that the rapid learning of coherently covarying properties is not solely a function of the frequency with which the property is experienced in the environment. In this training corpus, the property *is yellow* is true of three objects (the canary, the daisy, and the sunfish), whereas the property *has wings* is true of only two (the robin and the canary). Nevertheless, the network learns that the canary can fly more rapidly than it learns that the canary is yellow. The reason is that *can fly* varies coherently with several other properties, and this coherent covariation drives the model to represent the two birds as similar to one another and as different from the two fish. This structure in turn allows learning about the robin to generalise to the canary (and vice versa), and boosts the rapidity with which properties shared by the two birds are acquired. Comparing *is yellow* with *has wings* (also shown in the figure), we can see that in spite of its greater frequency, the time course of learning *is yellow* mirrors that of *can sing*. Learning about both properties depends on (and contributes to) the differentiation of the canary from the robin. The greater frequency of *is yellow* compared to *has wings* does exert a slight effect, but it only influences the relative degree of activation within a given level of differentiation—not the timing of when the information is mastered.

Properties that are shared by all items in the model's environment (e.g. *can grow*, *is living*, *isa living thing*) behave slightly differently than other properties in the training corpus. The network learns to activate these properties most rapidly, but they end up contributing almost nothing to the model's learned internal representations. Both phenomena may be understood with reference to the principles outlined above. Properties common to all objects are learned rapidly because all items have similar internal representations at the outset of training. Learning that one item can grow thus generalises to all other items; and because there are no items for which this is an incorrect attribution, there are no learning episodes that will "cancel" these weight changes out. The model comes very quickly to activate these properties, even before internal representations have differentiated to any great extent. Despite being rapidly learned, however, common properties exert very little influence on the way that representations change with learning. Because the target states for these properties are the same for all items, the error derivatives propagating back from them in early phases of learning will induce the internal representations to change in almost exactly the same direction—they will not serve to differentiate representations at all. The properties can be mastered quickly even without differentiation, and once they have been learned, they generate very little error and will not contribute strongly to any further representational change.

These phenomena suggest that there may be good computational reasons for the semantic system to begin with very similar, undifferentiated internal representations. Specifically, such an initial state would permit very rapid learning about properties common to all things— for example, that they are bounded, they fall if unsupported, they do not vanish and reappear spontaneously, etc. Certainly infants are known to show signs of such knowledge at very young ages (Spelke, Breinlinger, Macomber, & Jacobson, 1992; Baillargeon, 1995), a finding that has often been taken to indicate that infants are innately imbued with a “naive physics.” However the simulations suggest that, even if such knowledge was not innate, it could be acquired very rapidly if infants begin life with very similar and undifferentiated internal representations of objects.

The Importance of High Order Covariation

These developmental changes are driven by the model’s sensitivity to patterns of high-order covariation amongst the attributes of items in its environment. The properties to which the model first becomes sensitive, and which organise its nascent conceptual distinctions, are precisely those that consistently vary together across contexts. To begin to see this, consider Figure 3.5, which shows the covariances among all pairs of properties in the training corpus for simulation 3.1 (summarized in Appendix B, Table B–1). The properties are ordered according to the magnitude and direction of their mean covariance with other attributes, so that property pairs with strong positive covariances appear in the upper left and lower right quadrants, those with strong negative

covariances appear in the lower left and upper right quadrants, and those with covariances near zero appear toward the middle of the plot. The strongest covariances observed in the training corpus occur among properties that reliably discriminate broad conceptual domains (e.g. *has skin*, *has roots*). The covariance of properties shared by intermediate groupings of items is moderately strong and positive, whereas the covariance between properties characteristic of different clusters (e.g. between *has wings* and *has gills*) is moderately strong and negative. The weakest covariances are observed among properties common to all items in the training corpus (e.g. *can grow*), and among properties idiodyncratic to one or two items (e.g. *can sing*). From the Figure, it is clear that the order in which the network acquires sensitivity to different sets of properties is closely linked to the covariance structure of the items appearing in its training environment. Properties with the strongest covariances are precisely those to which the model first becomes sensitive during training.

While this inspection of the raw covariance matrix is informative, it only provides part of the story. The reason is that the covariance matrix only shows the relationships between individual pairs of properties; but the network's weights are shaped, not solely by *pairs* of properties that reliably covary together, but by *coherent sets* of multiple properties that all covary reliably together. The pairwise covariance of *has skin* and *can move* is certainly one factor that determines their contribution to learning and representation in the network; but more important is the fact that both of these properties also covary strongly and positively with *isa animal*, moderately strongly and positively with properties that typify birds or fish, and strongly and negatively with charactertistic plant properties. The model's stagelike mastery of conceptual structure at different levels, and its tendency to weight to object attributes differently at different stages, result from its sensitivity to the complete system of covariation across all sets of properties. Pairs of properties that covary strongly with one another, but which have no systematic relationship with other attributes, will not contribute strongly to the learning and representation that takes place in the network.

This is an important point, which we will take up again in considerable detail in later chapters. Here, we would simply like to introduce a second way of looking at the covariance structure of the training corpus in the current simulation, which more transparently reveals higher-order patterns of covariation. Figure 3.6 shows an eigenvector decomposition of the covariance matrix shown above, with the first eight eigenvectors represented.

Mathematically, an *eigenvector* of a matrix is any vector that does not change direction when multiplied by that matrix. For the present purposes it is most useful to understand each eigenvector as a set of weightings for the properties in the model, selected to account as closely as possible for patterns of property covariation shown in Figure 3.5. The first eigenvector contains one element for each property in the model, chosen in such a way that the product of any two elements comes as close as possible to the observed covariance of the corresponding two properties in Figure 3.5. For example, the element in the 1st eigenvector that corresponds to *can move* is large and positive, whereas the element that corresponds to *has roots* is large and negative. The product of these numbers is itself is large and negative, matching the observed large, negative covariance between the properties *can move* and *has roots* in Figure 3.5. In general, properties that, on average, tend to have large covariances with other properties will receive large weightings on this eigenvector. Idiosyncratic properties such as *is yellow* do not covary strongly with any other properties, and contribute little to the covariance matrix itself—hence they receive low weightings in the first eigenvector. Similarly, properties like *can grow* or *is living*, despite being very frequent, do not vary at all and hence do not covary strongly with other properties. Consequently, these properties do not receive strong weightings in the first (or subsequent) eigenvector. In this sense, the first eigenvector "picks out" those properties that account for as much of the property covariation matrix as possible. As Figure 3.6 illustrates, these tend to be the properties that reliably discriminate plants and animals in the current training corpus.

To derive the second eigenvector, the algorithm calculates the difference between the observed covariances, and those predicted from the 1st eigenvector alone (a *residual* matrix). The second eigenvector then consists of a set of weightings for each property, such that the product of any two weightings comes as close as possible to the elements of the residual matrix. Cells in the original covariance matrix that closely match the values predicted from the first eigenvector will yield residual values near zero, and thus will cease contributing to the calculation of further eigenvectors. Cells in the original matrix that were not well predicted from the first eigenvector will yield large positive or negative values in the residual matrix, and thus will contribute strongly to the calculation of values for the second eigenvector. Thus, the second eigenvector "explains" covariances that are not well accounted for by the first. When the second vector has been calculated, a new residual matrix is computed, and a third vector is derived from it, etc.

In Figure 3.6, the first 8 *eigenvectors* are shown. Each vector was normalised to span the same range, so that it is easy to compare the relative contributions of different features to each vector. As noted above, the first vector weights properties according to the degree to which they disciminate plant and animal items in the training corpus. The second vector strongly weights properties that discriminate birds and fish, whereas the third picks out properties that differentiate trees from flowers. Notably, attributes characteristic of animals or plants generally do not contribute to these vectors at all. Vector 4 differentiates red from yellow items, which is useful for discriminating robins and canaries as well as roses and daisies. Note that these properties, despite being more frequent overall in the training corpus, do not account for as much covariation as less frequent but more coherent properties (such as *has wings*), which are picked out by earlier *eigenvectors*. Vectors 5-7 weight those properties that serve to individuate pairs of items: vector 5 discriminates the pine from the oak; vector 6 separates both the rose and daisy and the robin and canary; and vector 7 differentiates the salmon and the sunfish. At this point, the residual matrix has effectively been reduced to very small random values, so the 8th vector (and all subsequent vectors) do not contain systematic weightings.

It is clear from the Figure that there is a close correspondence between the eigenvector decomposition of the property covariance matrix, and the model's discovery of representational structure. Features that are strongly weighted by the first eigenvector are those that are first emphasised by the model as it learns. Their values determine on which side of the network's first conceptual boundary each concept will fall. Subsequent vectors, as ordered, match the further discrimination of finer-grained conceptual categories in the network—the properties that receive strong weightings on these vectors determine on which side of further conceptual boundaries a given concept will fall. The reason for this close correspondence is simply that the models' weights are shaped predominantly by patterns of higher-order covariation among sets of attributes in the environment, as discussed above. The eigenvector decomposition provides an explicit representation of these patterns.

The overall situation can be summarised as follows. Initially the network is assigning virtually the same representation to all of the items. At this point, the properties that vary systematically with the network's internal representations are only those that are shared across everything — the *is living*, *can grow*, and *isa living thing* outputs. All other output properties have their effects on the forward weights almost completely cancelled out. However, the plants have several properties that none of the animals have, and vice-versa. Weak error signals from these properties begin to cumulate, eventually driving the representations of plants and animals apart. At this point, the shared animal representation can begin to drive the activation of output units that are common to all animals, and the shared plant representation can begin to drive activation of outputs common to all plants. The weights so structured in turn allow these coherently-varying properties to exert much stronger influences on the representation units than those exerted by the properties that differ between the birds and the fish. The result is that the individual animal representations stay similar to one another, and are rapidly propelled away from the individual plant representations. Very gradually, however, the weak signals back-propagated from properties that reliably discriminate birds from fish begin to cumulate, and cause the representations of these sub-groups to differentiate slightly, thereby providing a basis for exploiting this coherent covariation in the forward weights. This process eventually propagates all the way down to the subordinate level, so that idiosyncratic properties of individual items are eventually mastered by the net. In short, there is a kind of symbiosis of the weights into and out of the representation units, such that both sets are sensitive to successive waves of coherent covariation among output properties. Each "wave" of properties only maps systematically to the system's internal representations after the prior stage of differentiation has occurred. The timing of different waves of differentiation, and the particular groupings of internal representations that result, are governed by high-order patterns of property covariation in the training environment.

From this analysis, we can see that the network's tendency to differentiate its internal representations in a coarse-to-fine manner does not arise from some general bias toward discovering general or superordinate category structure per se. The network will not be strongly pressured to differentiate superordinate categories that do not have very cohesive structure, even if these are typically considered to form fairly broad superordinate classes (e.g. toys vs. tools). Its first conceptual distinctions will always correspond to those that capture the strongest patterns of coherent covariation across all of the items it encounters. The model thus suggests that not all superordinate categories will be discriminated with equal facility early in life—distinctions between items that share coherentlycovarying sets of properties are mostly likely to be acquired by young infants.

We believe that the dynamics of learning we have reviewed in this section can help us to understand several different aspects of semantic representation in development and in adulthood.

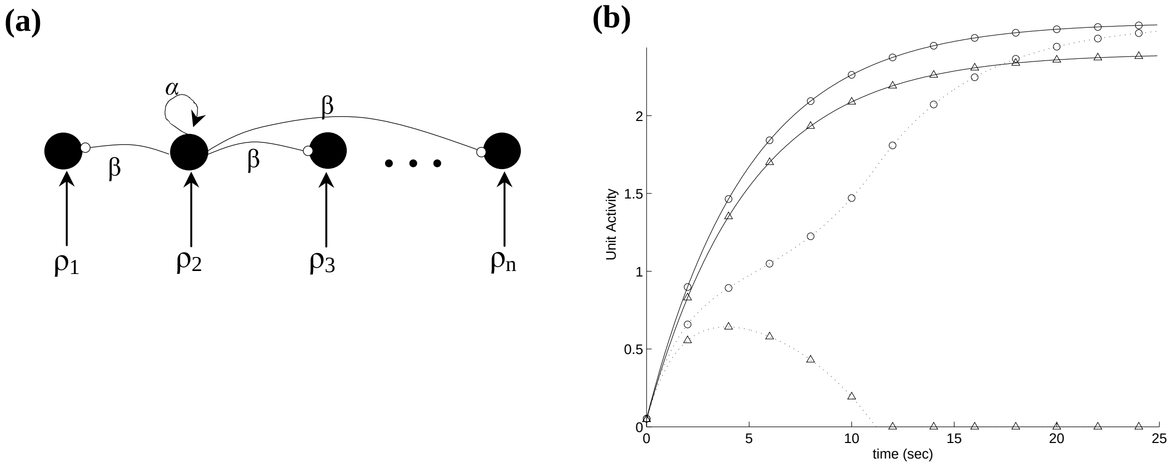


Figure 1. (a) The architecture of the Usher-McClelland network. α and β are the self-excitation and inhibition weights, and ρ_i is the external input to unit *i*. (b) The unit activities x_1 and x_2 (○ and Δ, respectively) when β=0.0 (solid lines) and β=0.4 (dotted line). In all cases, k=0.2, ρ_1=0.52, $\xi_1=\xi_2$=0, and dt=0.1.

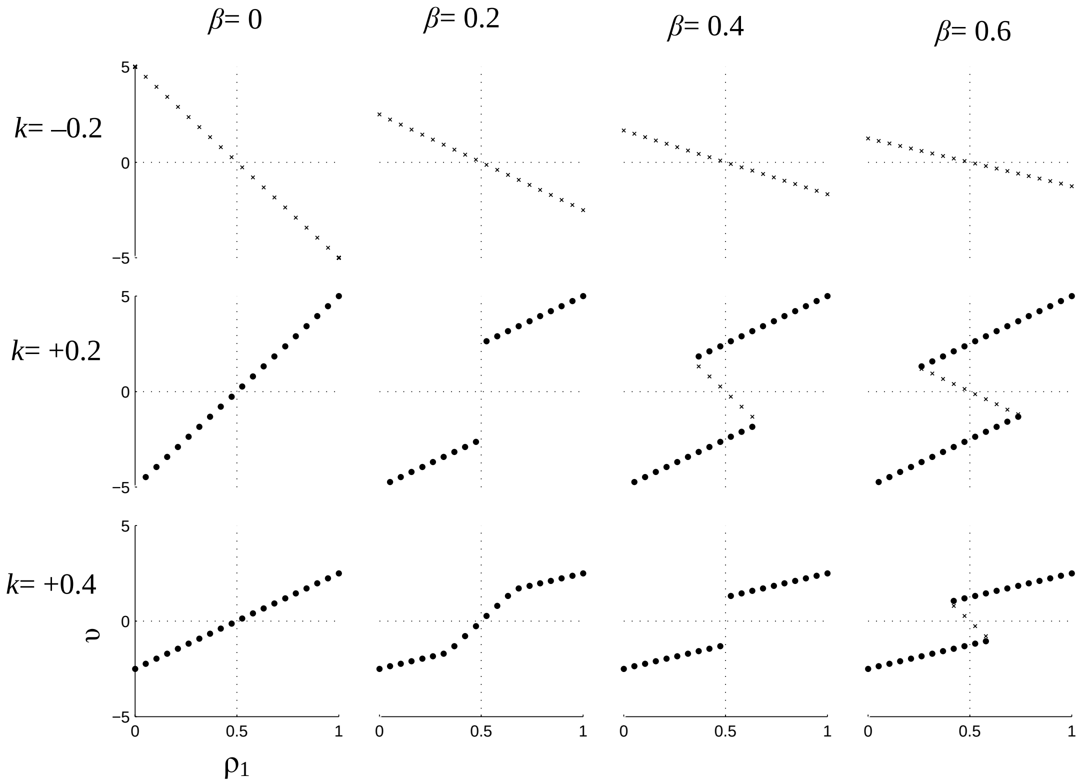


Figure 2. Bifurcation diagrams showing the stable and unstable fixed points (● and x respectively) for the dynamics of υ (≡ $x_1 - x_2$) in the Usher-McClelland network. In all cases ξ_1=ξ_2=0, and *dt*=0.1. When there are two stable fixed points for a parameter setting, they are referred to as υ^- and υ^+.

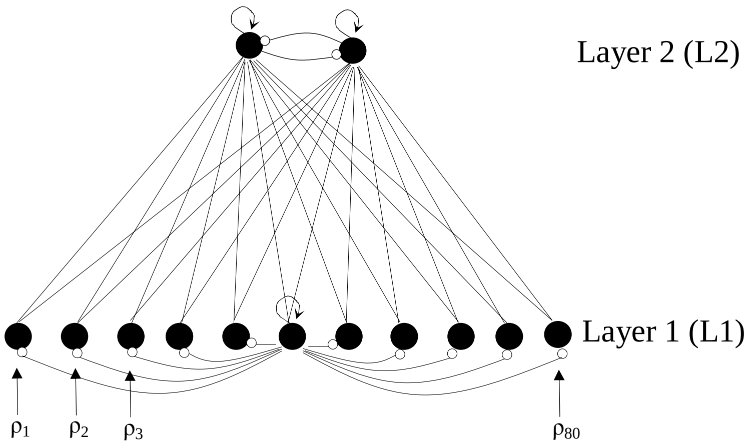


Figure 3. Architecture of Vallabha-McClelland model. L1 and L2 are fully connected to each other, with the L1→L2 weights initialized from *Uniform*(0, 0.03) and the L2→L3 weights from *Uniform*(0, 0.0005). Each unit has a self-connection with a weight of +1.0. The within-layer inhibitory connections have a weight of –0.2 for L2 units and –2.0 for L3 units.

Development as Change of System Dynamics: Stability, Instabilty, and Emergence.

G. Schöner
Institut für Neuroinformatik
Ruhr-Universität Bochum
Germany

May 1, 2006

1 Introduction

The conference, out of which this book grew, was dedicated to Esther Thelen. Three themes, that permeated Esther Thelen's work, are central also to Dynamical Systems thinking (Thelen, Smith, 1994; Smith, Thelen, 1993; Smith, Thelen, 2003). First, the conviction that cognition cannot be separated from is sensory and motor embodiment in a rich and changing environment; Second, the idea that cognition and other complex forms of behavior may emerge in suitable behavioral and environmental contexts. The soft-assembly of such skills may be multi-causally linked to diverse processes, none of which need be the single cause for the emergence of the skill. Third, all behavior and development unfold in time. The behavioral history in a task context and the developmental history matter for when and how a skill emerges. Development occurs therefore through individual developmental trajectories.

These themes structure Dynamical Systems thinking at different levels. Esther Thelen liked to emphasize how Dynamical Systems thinking can provide metaphors that help ask new questions, generate new experimental paradigms and new measures and lead to new kinds of explanations. In this chapter I will review Dynamical Systems Theory (DST) as a set of concepts that formalize such metaphors and thus turn these into a scientific theory of considerable rigor. While Dynamical Systems Theory is sometimes also viewed as a collection of certain kinds of models, this chapter will not focus on any specific model nor review modelling work in general.

In fact, DST is *not* the idea differential equations or dynamic iterative maps are good models of the nervous system. Instead, DST is a much more specific set of concepts, of which I shall emphasize five in this chapter: (1) Behavioral patterns resist change, that is, are stable. This may be mathematically characterized by considering behavioral patterns attractor states of a dynamical system. (2) Behavioral change is brought about by a loss of stability. (3) Representations possess stability properties as well and can be understood as the attractor states of dynamic fields, that is, of continuous distributions of neural activation. (4) Cognitive processes emerge from instabilities of dynamic fiels. (5) Learning consists of changes of the behavioral or field dynamics, which shift the behavioral and environmental context in which instabilities occur.

What I shall do below is walk the reader through each of these conceptual steps to explain and illustrate the key ideas, evoke exemplary model systems in which these ideas have been brought to fruitition, and finally link the ideas to other approaches and problems.

2 Dynamical Systems Theory (DST)

The central nervous system is tightly interconnected. As a result, any given state of the central nervous system is exposed to a range of influences some of which may be pushing to change the neural state. The very flexibility of the central nervous system, the interconnectedness and multi-functionality of many of its components make such perturbations the norm rather than the exception. Only neural states that resist such perturbations will persist long enough to be observable and to influence down-stream processes enough to have behavioral and long-term effects. Stability, the capacity to resist perturbations, is thus a key property of the functional states of nervous systems.

Mathematically, stability is the constitutive property of attractors. Illustrated in Figure 1, an attractor is an invariant (unchanging in time) solution of a dynamical system, toward which solutions converge if they start nearby. If perturbations push the state away from an attractor, the dynamical system restores the attractor state. Attractors emerge from the dynamics of a system as points at which the forces pushing in opposite directions converge and balance. Neurons and neural networks are naturally described as dynamical systems (Wilson, 1999) and provide through their intrinsic dynamics the mechanisms that stabilize attractors (see Hock, Schöner, Giese, 2003, for a discussion). In that sense, stability comes for free from the neuronal dynamics prevalent in the central nervous system, although the sensory-motor periphery (such as from muscle viscosity and elasticity) may also contribute.

Once a stabilization mechanism is in place, there is no need for other computational mechanisms to determine the output of a dynamic neural network. In fact, the very forces that restore an attractor state following perturbations also helps the system track any changes to the attractor state incurred as inputs to the network change (Figure 2).

The picture up to this point, a unique stable state which may track changes, is not

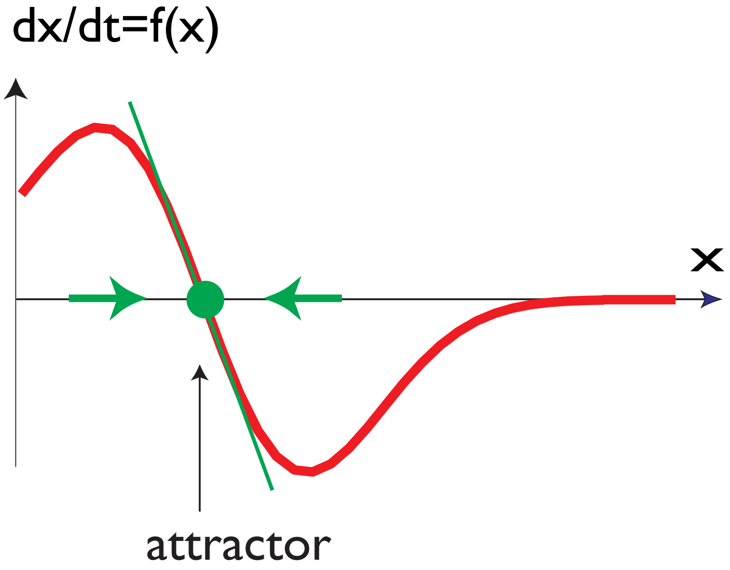


Figure 1: A differential equation model of a dynamical system is defined by how the rate of change, dx/dt, of the state variable, x, depends on the current state: $dx/dt = f(x)$. The present state thus determines the future evolution of system. In the presence of an attractor the system evolves by converging to the attractor as indicated by the arrows: A negative rate of change for values larger than the attractor state leads to a decrease in time toward the attractor, a positive rate of change for values smaller than the attractor leads to increase in time toward the attractor. The time-invariant attractor thus structures the temporal evolution of the state of the dynamical system in its vicinity.

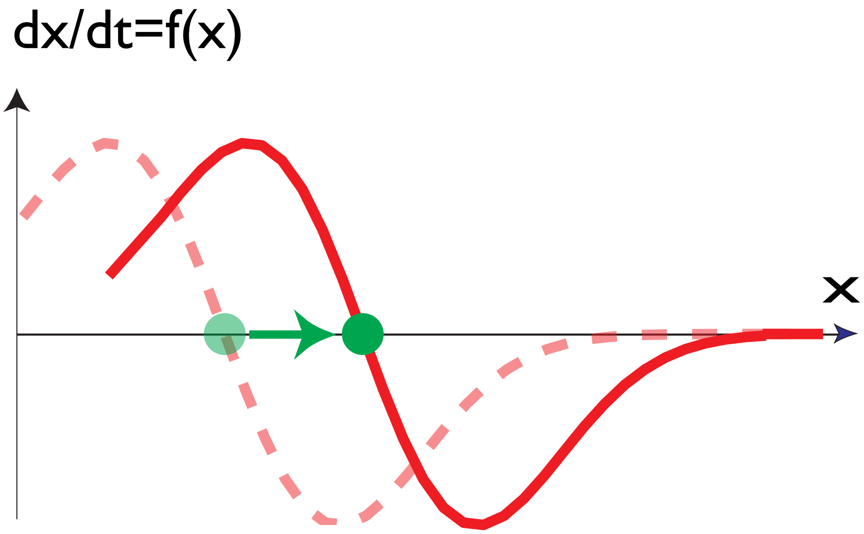


Figure 2: When a dynamical system changes (from the form shown as a dashed line to the form shown as a solid line), the stability of an attractor state leads to automatic update of the state through relaxation to the new attractor.

different from the basic framework of cybernetics. Something new happens, however, when multiple attractors co-exist. The simplest such case, bistability, is illustrated in Figure 3. Which of the two attractor states is realized depends on the prior history of the system. As long as the state of the system lies within the basin of attraction of attractor one (to the left of the repellor), the system tracks that state as the dynamics change. In a sense, selecting one over another attractor is a simple form of decision making and the dynamics stabilizes such decisions. On the other hand, on rare occasions stochastic perturbations may shift the system sufficently far so that a cross-over into the alternative basin of attraction occurs and a stochastic switch of state is induced.

Stochastic switches become more likely when a change of the dynamics reduces the stability of one attractor. Figure 4 illustrates what happens when changes of the dynamical system reach a critical point at which an attractor (number 1 on the left) loses stability. At the (tangent) bifurcation the attractor 1 collides with the repellor and the two annihilate. If the system was originally in attractor 1, it may track this attractor until it becomes unstable, at which point the system will switch to attractor 2.

Instabilies may thus lead to qualitative change, as contrasted to the mere tracking of a continuously changing state. The change is preceeded by tell-tale signatures of instability such as increasing variability and increasing time needed to recover from

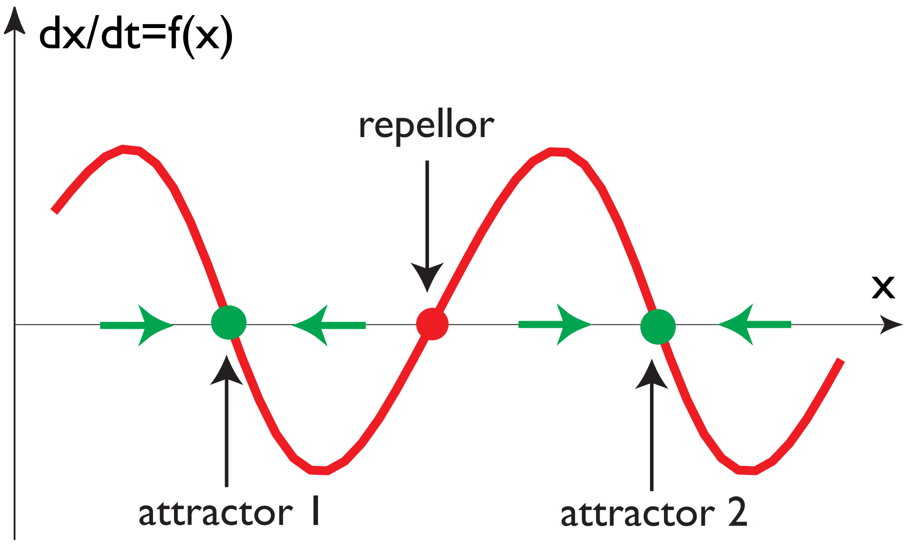


Figure 3: A non-linear dynamical system with two attractors separated by a repellor, a time-invarant solution from which the system diverges.

perturbations. These can be exploited to detect instabilities and thus to distinguish qualitative from quantitative change (review, Schöner, Kelso, 1988). In non-linear dynamical systems, instabilities arise naturally in response to even relatively unspecific changes of the dynamics. In the illustration of Figure 3, for instance, the changes to the dynamics are not specifically localized around the attractor "one" that is losing stability, but rather amount essentially to an increasing bias toward larger values of the state variable "x".

Attractors are thus not fixed entities. When they disappear, they are not stored somewhere, or simply "deactivated". That attractors may emerge out of nowhere when the conditions (the dynamics) are right can be visualized by looking at the scenario of Figure 4 in the reverse order: As the bias to larger values of "x" is reduced, a new attractor may be formed spontaneously, coming "out of nowhere" and branching off an associated repellor that forms a new boundary between two basins of attraction.

So far, I have talked about the "state" of a system in the abstract. What kind of variables "x" would describe behaviors, patterns, decisions etc.? Many readers may be familiar with now classical examples of DST, reviewed for instance in Scott Kelso's book (1995), in which variables such as the relative phase between two rhythmically moving limbs are called "order parameters" and considered sufficient to characterize patterns of movement coordination. Relative phase may be the only obvious example of a single variable that clearly describes a pattern (of relative timing) and that is independent of

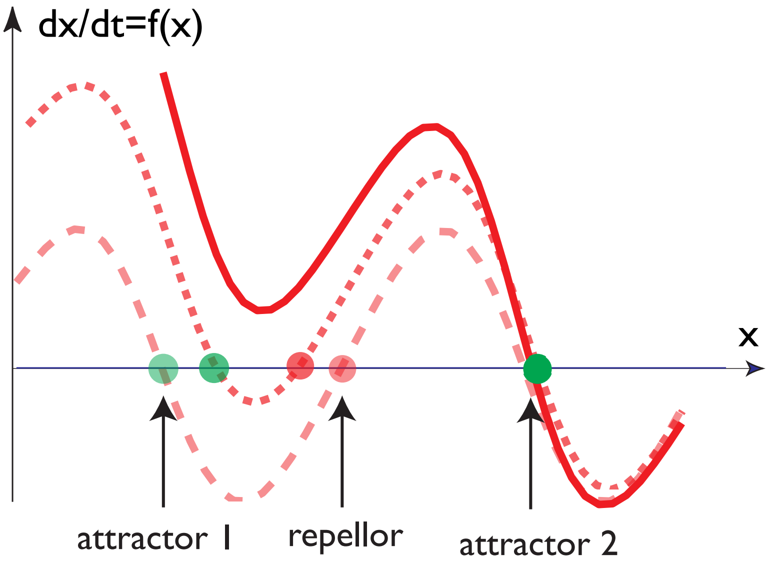


Figure 4: Changes of the bistable dynamical system (from dashed via dotted to solidl line) lead to an instability, in which the attractor (1) collides with the repellor, leaving a single attractor (2) behind.

other details of how the pattern is generated (e.g., the movement amplitude or the exact trajectory shape). Other examples from the literature are less obvious. For instance, the many oscillator models in the literature of coordination of rhythmic movement are formulated in terms of variables that describe the spatial position and velocity of the moving effector, but are not identical to these physical variables. When the limb is perturbed, for instance, the physical position and velocity is changed, but the oscillator driving the movement is not necessarily affected (Kay et al., 1987). On the other hand, the oscillator variables are not directly neural activations either (see, e.g., Grossberg, Pribe, Cohen, 1997).

That Dynamical Systems models may require only a small number of variables, may be low-dimensional, is a central assumption of DST. What is it based on? Why should it be possible to describe what happens when a nervous system generates behavior, when millions of neurons engage sensory and motor processes, coupled through feedback in the outer world, by a simple differential equation of one or two dynamical variables? There is a mathematical answer to this question, which I will briefly sketch, even if a full explanation goes beyond what can be achieved in a survey chapter. That the ensemble of neural processes involved in the generation of behavior forms a high-dimensional dynamical system is plausible on the basis of the physiology and physics of neural function (Wilson, 1999). Stable states in such a high-dimensional dynamcis are the persistent macroscopic states which are observable at the behavioral level. Stability means that in all directions of the high-dimensional space, restoring forces secure the state against perturbations. An attractor becomes unstable when the restoring forces in one particular direction begin to fail (Figure 5). Only under exceptional circumstances caused by symmetries would stability fail in multiple directions at the same time. The direction along which the restoring forces become weak defines the low-dimensional Center-Manifold (see, e.g., Perko, 1991, for a textbook treatment). The temporal evolution of the state of the system along the Center-Manifold is slower than in any other direction of the high-dimensional state. To predict the long-term evolution of the system one may therefore assume that the system has already relaxed to the Center-Manifold and describe the dynamics only within that manifold. This intuition is formalzied in the Center-Manifold-Theorem, which says that knowing how the system evolves along the Center-Manifold uniquely determines how the system evolves in the original high-dimensional space. Thus, to capture the macroscopic states of the high-dimensional dynamics and their long-term evolution, it is sufficient to model the dynamics along directions along which instabilities occur. Although this is mathematically true only at a bifurcation, in practice the low-dimensional description provides a fair representation of the fuller dynamics whenever the system is near an instability, as is almost always the case in the nervous system (see also Haken, 1983, for an introduction).

Does this mean that the low-dimensional dynamical systems are purely descriptive while a full mechanistic account must take place in the original, high-dimensional space? The answer depends on what is meant by “full mechanistic account”. If that means, literally, an account that captures the state of all neural processes, then, by definition, only extensive high-dimensional computational modelling will be satisfactory. If this means, however, that an account is sufficient to actually generate a behavior in a real system, then capturing the macroscopic, low-dimensional dynamics qualifies. A prove of sufficiency in that sense has been provided, for instance, by generating simple robotic behaviors such as target acquisition, target selection, obstacle avoidance, etc. from low-dimensional attractor dynamics with appropriate bifurcations interfaced with very simple sensory and motor systems (see Bicho, Schöner, 1997, for an example).

A related question is how abstract the low-dimensional dynamical descriptions of behavior end up being. The Center-Manifold argument suggests that fairly abstract models may result, models that cut through the high-dimensional space describing the neural systems supporting behavior in ways that depend on the task, on the state studied, and on the particular parameteric and input conditions under which an instability is observed. On the other hand, over the last few years a closer alliance of Dynamical Systems models with neurophysiological principles has contributed much to reducing the gap between the low-dimensional dynamical descriptions and the neural networks that implement them. This will a theme in the next section.

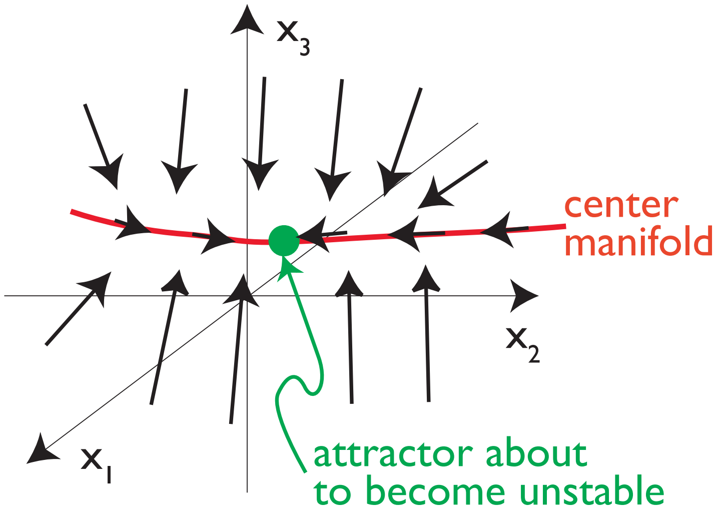


Figure 5: When an attractor of high-dimensional dynamical system (of which 3 dimensions, x_1, x_2, and x_3 are sketched here), becomes unstable, there is typically one direction in the high-dimensional space along which the restoring forces begin to fade (shorter arrows) while in other directions the stabilization mechanism still works (longer arrows). That first direction spans the Center-Manifold.

Finally, is DST primarily about motor behavior? The exemplary model systems that influenced the development of Dynamical Systems ideas did come from the motor domain (but see Hock, Schöner, Giese, 2003). On the other hand, much work has since shown that the ideas are not intrinsically tied to motor behavior. A conceptual problem must be confronted, however, when stepping beyond the domain of motor control in the narrower sense. In motor control one may without difficulty talk about a state of the motor system, e.g., the position of my arm or the frequency or phase of my arm's rhythmical movement. That state exists at all times and changes continuously in time. The dynamical variables we have postulated capture that continous change. What value, however, does the phase of my arm's rhythmical movement have when I stop moving? With which phase does it start up again? Obviously, as we move beyond motor control to include the selection of motor acts, the initiation of action, but also the generation of perceptual patterns, the comission to memory of a perceptual or motor state, we must find some form in which the values of state variables represent these more abstract states of affairs. We must examine, to which extent such state variables may change continuously as the seemingly discrete acts take place.

3 Dynamic Field Theory (DFT)

The classical concept of activation can do this work for us. As a neural concept, activation is invoked in much of cognitive psychology and in all connectionist models. Activation may be mapped onto observable behavioral states by postulating that high levels of activation impact on down-stream systems, including ultimately on the motor system, while low levels of activation do not. This captures the fundamental sigmoidal nonlinearity of neural function: Only activated neurons transmit to their projection targets, while insufficiently activated neurons do not. There are multiple neuronal mechanism through which activation may be realized neuronally (e.g.,. through the intra-cellular electrical potential in neurons, through the firing rate of neurons, through the firing patterns of neural populations, etc., see, e.g., Dayan and Abbott, 2001). But the concept of activation does its work for DST independently of the details of its neural implementation.

The second concept we'll need is that of an activation field, that is, of a set of continuously many activation variables, $u(x)$, defined over a continuous dimension, x, that spans a range of behaviors, percepts, plans, et cetera. Different states of affairs can be represented by such activation fields (Figure 6). Localized peaks of activation (top) indicate two things: the presence of large values of activation means that the activation field is capable of influencing down-stream structures and behavior; the location of the peak represents the value along the dimension that this peak specifies. By contrast, flat patterns of low level activation (bottom) represent the absence of specific information

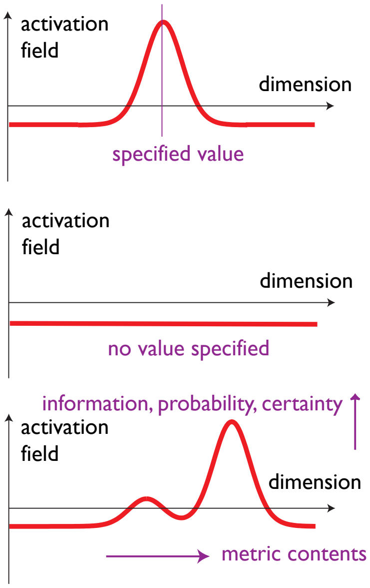


Figure 6: An activation field defined over a continuous dimension, x, may represent through a localized peak of activation both the presence of information and an estimate of the specified value along the dimension (top), while flat, low-level distributions of activation indicate the absence of information about the dimension (middle). More generally, activation fields may represented in a graded way information about the metric dimension over which they are defined (bottom).

about the dimension. Conceptually, localized peaks are the units of representations in DFT. When peaks shift continuously along the field dimension we are back to the simpler picture of DST, in which the value of the state variable, x, represented by the peak location, changes continuously. Graded patterns of activation may represent varying amounts of information, probabilities of a response or an input, or how close the field is to bringing about an effect (Figure 6, bottom panel). The dimensional axis encodes the metrics of the representation, what is being prepared, perceived, or memorized and how different the various possible values along the dimension are.

This form of representation is essentially the space code principle of neurophysiology (e.g., Dayan and Abbott, 2001), according to which the location of a neuron in the neural network determines *what* the neuron encodes, while its level of activation represents how certain or important or imminent the information is that the neuron transmits. Feature maps are a possible neuronal realizations of such activation fields, consisting of ensembles of neurons that spatially code for the feature dimensions. Such maps conserve topology, so that neighboring neurons represent neighboring feature values. The activation field is the approximation of such a network by a continuum of

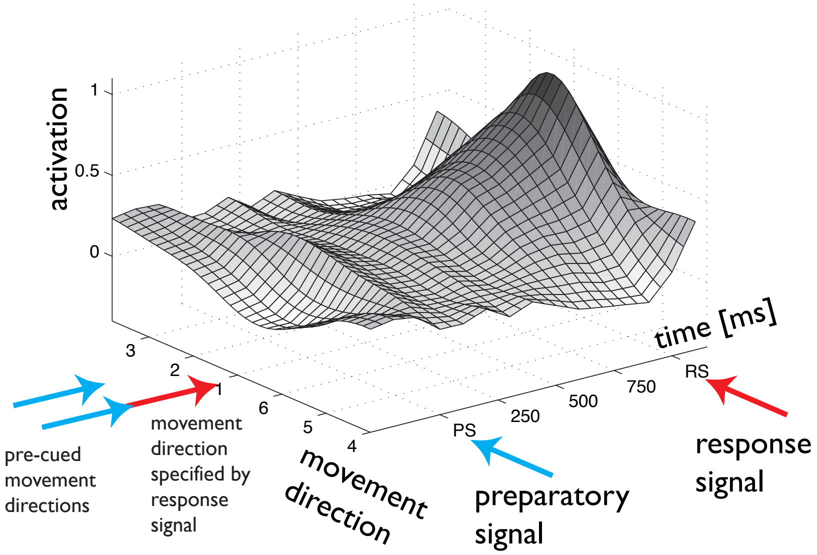


Figure 7: A distribution of population activation over the dimension "movement direction" evolves in time under the influence of a preparatory signal, which specifies two possible upcoming movement directions, and a response signal, which selects one of the two. The distribution is estimated from the tuning curves of about 100 neurons in motor and pre-motor cortex and their firing rate in 10 ms time intervals.

neurons, justified because neural tuning curves overlap strongly.

More generally, there is no need for feature representations to remain spatially aligned with the feature spaces represented. Distributions of population activation may be constructed from the neural activation levels of many neurons, irrespective of where these neurons are located in a cortical or subcortical area. Each neuron is localized in the field based on its specificity with respect to the represented dimension (Erlhagen et al., 1999). Figure 7 shows an example, in which the firing rates of about 100 neurons in motor and premotor cortex are used to construct a population distribution of activation over the space of movement directions, the spatial directions into which a monkey moves its hand toward visual targets (the data are from Bastian, Schöner, and Riehle, 2003, the representation is generated based on the optimal linear estimator method as described in Erlhagen et al., 1999). The initiation of a voluntary movement is linked to the generation of a peak, the direction of the movement being specified by the location of the peak.

How may activation fields be endowed with stability and attractors? Naturally, the fields themselves are assumed to form dynamical systems, consistent with the physiology and physics of the corresponding neural networks. The temporal evolution

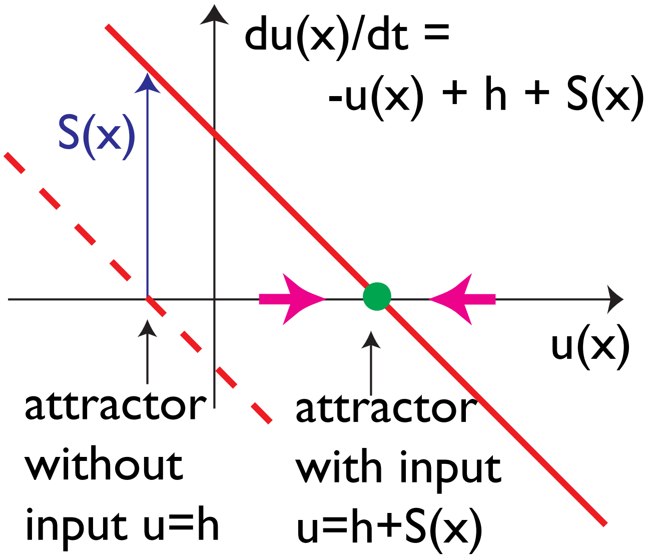


Figure 8: The generic dynamics of an activation variable, $u(x)$, is illustrated in the absence of interaction. The fundamental stabilization mechanism ("$-u$") generates an attractor state at the (negative) resting level h in the absence on input (dashed line). Input, $S(x)$, shifts the rate of change upwards, generating a new attractor at $u = h + S(x)$ (solid line).

of these Dynamic Fields is thus generated by forces that determine the rate of change of activation at each field site. Two factors contribute to the field dynamics, inputs and interactions. Inputs are those contributions to the field dynamics which do not depend on the current state of the field. The role of inputs may thus be understood separately for every field site as illustrated in Figure 8. The propensity of the field dynamics to generate attractors is an instrinsic property of neural dynamics and by assuming that the rate of change of activation is inversely proportional to the level of activation (any monotonically decreasing function would do). Inputs set attractor states for each activation variable while, in the absence of input, the activation variable has an attractor at a resting level (assumed negative here by convention although the units are arbitrary)

Different field sites, $u(x)$, receive different inputs, $S(x)$. For sensory representations, the way inputs vary along the field dimension, x, defines the meaning of that dimension. The typical idea is that inputs derive from the sensory surfaces, so that the input function varies continuously as a function of x and neighboring sites receive overlapping inputs from any item on the sensory surface (Figure 9). For motor fields, the meaning of the dimension is established by how the field projects onto the mo-

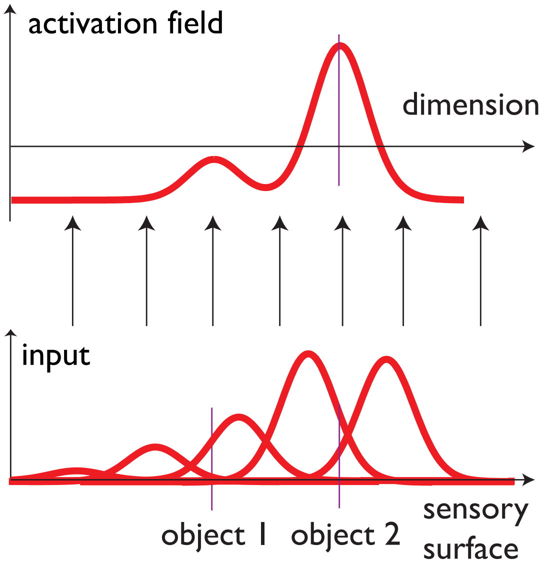


Figure 9: Input to a dynamic activation field may derive from a sensory surface. The map from the surface to the field may involve feature extraction or complex transformation, although it is represented by a homogeous one-to-one mapping here. Sensory cells with broad tuning generate inputs to a range of field sites are illustrated here for a sample of only five sensor cells. In this example, the input is generated by two objects on the sensory surface. The dynamic field selects object number 2 through interaction.

tor surface. Connectionist networks sometimes make the same assumption (e.g., when modelling cortical feature maps), although other connectionist networks have much more complex input functions which do not necessarily lead to localized patterns of input, but instead to distributed respresentations.

Interactions are all contributions to the field dynamics which depend on the current activation level at any location of the field. Connectionist refer to networks with interactions as "recurrent" networks, although it is not possible in a general form to characterize any single connection as being either recurrent or input related. Dynamic Field Theory is based on the assumption that there is a universal principle of interaction, local excitation – global inhibition (Figure 10). First, neighboring field sites which represent similar values of the dimension are assumed to interact excitatorily, that is, activation at either site generates positive rates of change at the other site. This form of interaction stabilizes peaks of activation against decay and thus contributes to the stability of peaks. Second, field sites at any distance are assumed to interact inhibitorily, that is, activation at either site generates negative rates of change at the other site. This form of interaction prevents peaks from broadening through lateral dif-

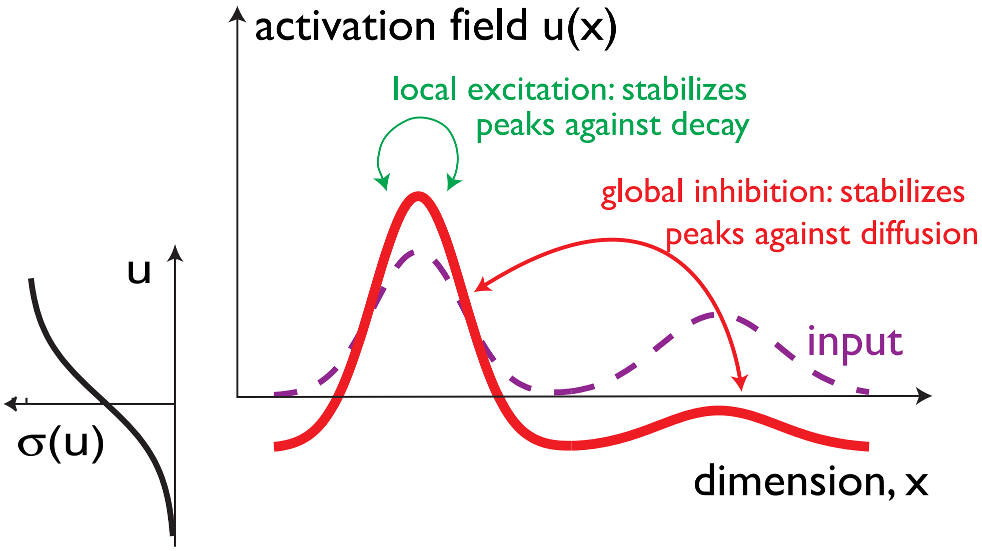


Figure 10: Locally excitatory and globally inhibitory interaction enable dynamic fields (solid line) to generate localized peaks of activation (left), centered on locations of maximal input (dashed line), as well as to suppress activation at competing locations (right). Interaction is mediated by a sigmoidal function sketched on the left, so that only activated sites contribute. The activated site on the left, for instance, suppresses activation at the site on the right, but not vice versa.

fusion and thus also contributes to the stability of peaks. This pattern of interaction is ubiquitous in cortex and many subcortical structures. Only field sites with a sufficient level of activation contribute to interaction, a principle of neural function described by sigmoidal transmission functions and source of the fundamental non-linearity of neural dynamics (e.g., Grossberg, 1973).

When interaction is strong it may dominate over inputs, so that the attractor states of the dynamic field are no longer dictated by input. In other words, strongly interacting dynamic fields are no longer described by input-output relationships. Instead, such fields make decisions. This may be demonstrated even for the simplest response of a dynamic field, the detection of a single localized object on the sensory surface (Figure 11). For weak input strength, the field simply reproduces the input pattern as its stable activation pattern. This is the input driven regime, in which most neural networks are operated in connectionist modelling. When input becomes stronger, this state becomes unstable because local excitatory interaction begins to amplify the stimulated field site. The field relaxes to the other stable state available, in which the activated peak is self-stabilized by the combination of local excitation and global

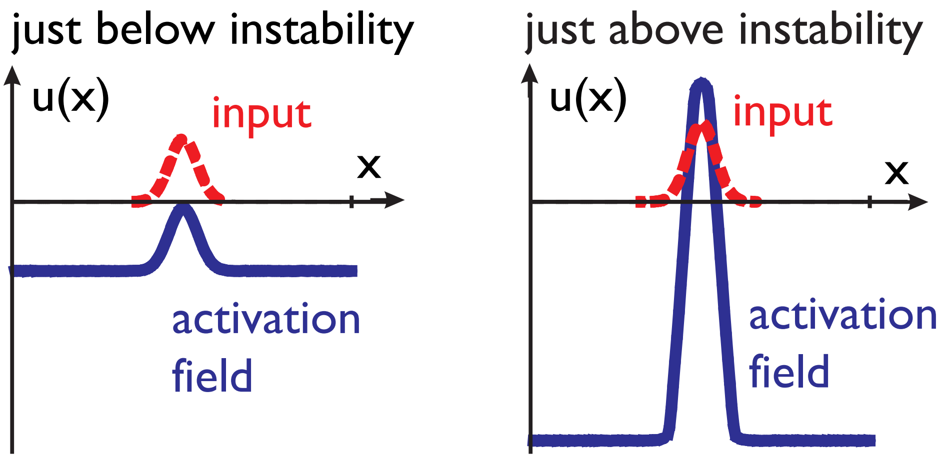


Figure 11: In response to a single localized input (dashed line) a dynamic field generates a input-defined pattern of activation while the input is below the detection instability (left). When the input drives activation beyond a critical level, the field goes through an instability, in which the input-defined pattern becomes unstable and the field relaxes to an interaction-dominated peak (right). This peak is stabilized by local excitatory interaction, which pulls the activation peak higher than specified by input, and global inhibition, which strongly suppresses all locations outside the peak.

inhibition (see Amari, 1977, for seminal mathematical analysis).

Under appropriate conditions, a self-stablized peak of this kind may persist as an attractor of the field dynamics in the absence of input. It may then serve as a form of sustained activation, supporting working memory (Thelen, et al., 2001; Schutter, Spencer, Schöner, 2003; Spencer, Schöner, 2003; for neurocomputational models see, e.g, Durstewitz, Seamans, Sejnowski, 2000; Wang, 1999). Such sustained peaks may coexist with input-defined patterns of activation, forming a bistable dynamical system (Figure 12). By lowering the resting level, the self-sustained solution may be made unstable, returning the system to a mono-stable input-driven state and "resetting" working memory. The detection instability enables the "setting" of memory by making the input-driven state unstable. Together, these two instabilities establish a form of neural "flip-flop".

The capability of dynamic fields to make decisions by selecting one of multiple sources of input emerges similarly from an instability (Figure 13). When such sources are metrically close, a peak positioned over an averaged location is monostable. When

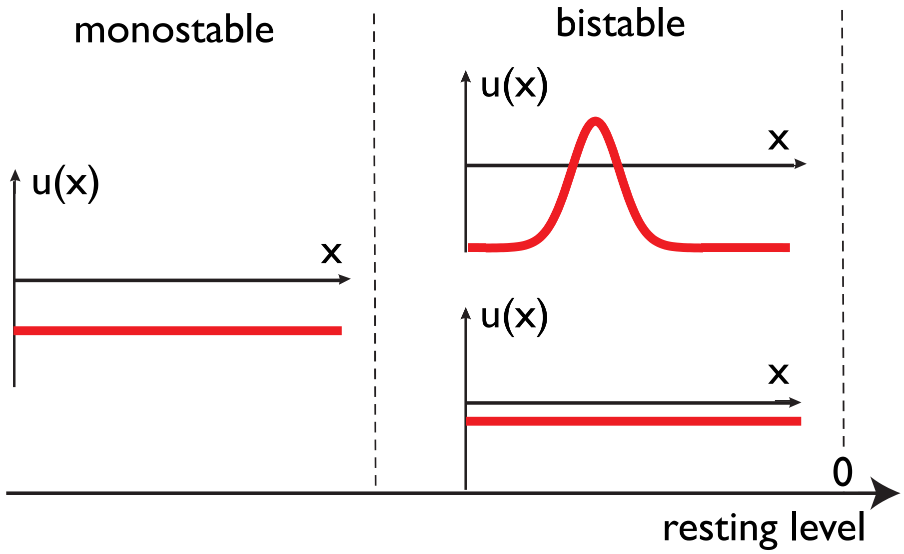


Figure 12: Self-sustained activation peaks (top right) and flat activation patterns at resting level (bottom right) are bistable in an appropriate parameter regime. When activation is globally lowered (e.g., by lowering the resting level), the sustained activation mode becomes unstable and a mono-stable regime results (left).

such sources are metrically far, the field selects one of the sources and suppresses activation at the other locations. An account for the transition from averaging to selection in visually-guided saccades based on this instability has successfully described behavioral and neural features of saccade initiation in considerable detail (Kopecz, Schöner, 1995; Trappenberg et al., 2001; Wilimzig, Schneider, Schöner, 2006).

Why is it important that representations are endowed with stability? And what role do instabilities play then? The computational metaphor fundamentally is timeless, conceiving of cognition as the computation of responses to inputs. The inputs are given at some point in time, the response is generated after a latency which reflects the amount of computation involved. An embodied and situated cognitive system, by contrast, acts under the continuous influence of inputs. New processes are started on a background of ongoing processes. To provide for any kind of behavioral coherence, cognitive processes must be stabilized against the continuous onslaught of new sensory information, of sensory feedback from ongoing action, and of internal interactions from parallel, possibly competing processes. The localized activation peaks of DFT are instances of stable states that support representation and resist change through the self-stabilization induced by interaction. Resistence to change induces the reverse problem, of course, that of achieving flexibility, being capable to releasing one processes to give

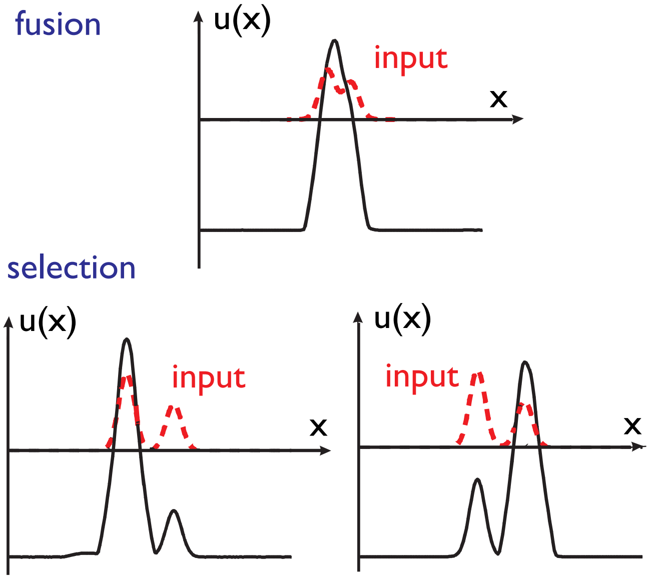


Figure 13: Two closely spaced peaks in the input (dashed line) may generate a monostable fused activation peak (solid) positioned over an averaged location (top). This is due to local excitatory interaction. When the two input peaks are increasingly more separate (bottom), this fused solution becomes unstable and the fused attractor bifurcates into two attractors. In each, one input peak is selected while activation at the other is suppressed.

way to a new process. In DFT, stable peaks of activation must be destablized to allow for the generation of new peaks. Instabilities play this crucial role. Through instabilities, moreover, cognitive properties emerge such as sustained activation, the capacity to select among inputs, to fuse inputs, to couple to time-varying inputs, or to suppress such coupling.

Stabilities and instabilities thus form the foundation for embodied cognition because they enable the emergence of cognitive states while retaining close ties to the sensory and motor surfaces, which are situated in rich, structured, and time-varying environments. Within DFT, the mechanisms for creating stability and instability are neurally plausible. Finally, the Dyanamical Systems account both in the classical form of DST and in the expanded form of DFT is open to an understanding of learning and development.

4 Learning and development

In DST learning means changing the dynamics of the system. Altough easily stated and seemingly obvious, this insight has far-reaching consequences. Work on how people learn new patterns of interlimb coordination may serve to illustrate the ideas (Schöner, 1989; Zanone, Schöner, Kelso, 1992; Zanone, Kelso, 1992).

Participants practiced a new pattern of bimanual coordination, a 90 deg phase relationship over 5 sessions in so many days. They were provided with knowledge of results after each trial. Before and after each session, they performed other phase relationships sampling the range from in-phase (0 deg) to phase alternation (180 deg) in 10 steps. These required relative phases were made invoked by presenting two metronomes with the requested phase relationship. Figure 14 highlights the main results. During learning, both the constant and the variable error error decrease, consistent with increasing stability of the practiced pattern. Before learning, performance of other patterns of relative timing is systematically biased toward phase alternation, one of the intrinsically stable patterns of coordination. After learning, performance has not only improved at the practiced 90 deg pattern, but also for all other patterns. There is now a systematic bias toward 90 degrees , reflecting the changed dynamics of coordination which have acquired a new force attracting to 90 degrees relative phase.

Learning a motor skill thus amounts to generating new forces which stabilize the new patterns. To see what such learning processes look like in neural language we examine the simplest learning processes in DFT. Patterns of activation may be stabilized simply by leaving a *memory trace* of ongoing activity, which acts to preshape the activation fields when new inputs arrive (Schöner, Kopecz, Erlhagen, 1997; Erlhagen, Schöner, 2002; Thelen et al., 2001). Hebbian strengthening of inputs to those field sites at which a peak is generated would be a simple neural mechanism for what functionally

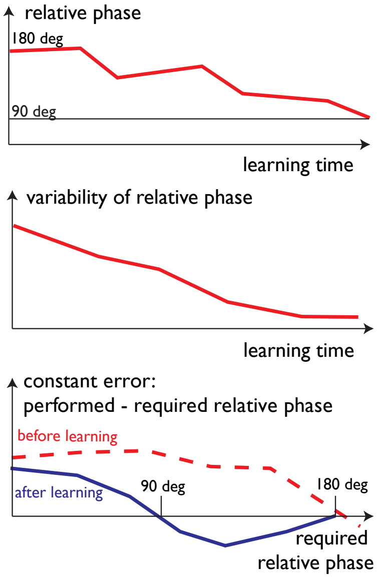


Figure 14: Schematic summary of the results of the Zanone-Schöner-Kelso (1992) experiment on the learning a bimanual coordination pattern of 90 degrees relative phase. Over learning, the mean relative phase (top) approached correct performance at 90 degrees from an initial bias toward phase alternation (180 degrees), while the variability of relative phase (middle) decreased. The bottom graph compares the constant error of relative phase at a range of required relative phases before (dashed) and after learning (solid). Note bias toward phase alternation before learning (constant error is positive for for required relative phases below 180 degrees) and bias to 90 degrees after learning the improvement not only at the practiced 90 degree pattern, but also in the vicinity of this pattern (constant error is positive below, negative above 90 degrees).

the memory trace does. We refer to the combined mechanism of a memory trace and its role of preshaping the field around previously activated locations as the *preshape dynamics* of a dynamic field.

Support for the concept of preshaping and a preshape dynamics comes from accounts for a wide range of phenomena, including how the probability of choices influences reaction times (Erlhagen, Schöner, 2002), how motor habits are formed and then influence motor decisions in the A not B paradigm (Thelen et al., 2001), and how spatial memory is biased toward previously memorized locations (Schutte, Spencer, Schöner, 2003). As an illustration consider perseverative reaching in the A not B paradigm. In this task setting, infants must make a motoric decision by either reaching to the A or to the B location. Thelen and colleagues (2001) have shown how a memory trace built during the first few reaches to the A location, generated in response to an attractive stimulus at A, biases the decision on the B trial in favor of A. Recent work by Evelina Dineva (2005; Dineva, Schöner, Thelen, in preparation) has elaborated how the behavioral history in the task predicts performance. Figure 15 illustrates the evolution of a dynamic field representing reaching directions and its preshape over six A- and two B-trials. Reaches to A leave a memory trace, which preshapes the field at A and thus promotes further reaches to A, inducing the A not B error in the B trials. Spontaneous errors, that is, reaches to B on A trials, leave a memory trace at B, which promotes further reaches to B and supports correct reaching to B on B trials.

Does learning always have a stabilizing effect? We already saw that learning may destabilize patterns that were stable before learning (such as the tendency to phase alternation in bimanual coordination, see also Schöner, 1989, and Schöner, Zanone, Kelso, 1992 for a dicussion of instabilities during learning). In DFT, a destabilizing form of learning is implied in habituation (Schöner, Thelen, 2006). A perceptual dimension is used to span the different stimuli probed in a typical habituation experiment, expressing similarities between stimuli through metric distances within such a dimension. Figure 16 illustrates this idea for the influential drawbridge paradigm (Baillargeon, Spelke, Wasserman, 1985). Test stimuli that overlap more with the habituation simulus along this dimension are considered "familiar" stimuli, test stimuli that overlap less are "novel" stimuli.

Along the perceptual dimension, two dynamic fields are defined. Looking is controlled by the activation field that receives direct input from the visual array (Figure 17) and drives a second field which projects as inhibition onto the first field. Looking away is thus controlled by the slow build-up of inhibition, which destabilizes positive peaks of activation.

In the Dynamic Field model, the (feature)space and time structure of perceptual experience determines looking. Thus, for instance, the amount of dishabituation is determined by how strongly habituation and test stimuli overlap (Figure 18). The temporal order may also matter such as when dishabituation is stronger to the familiar

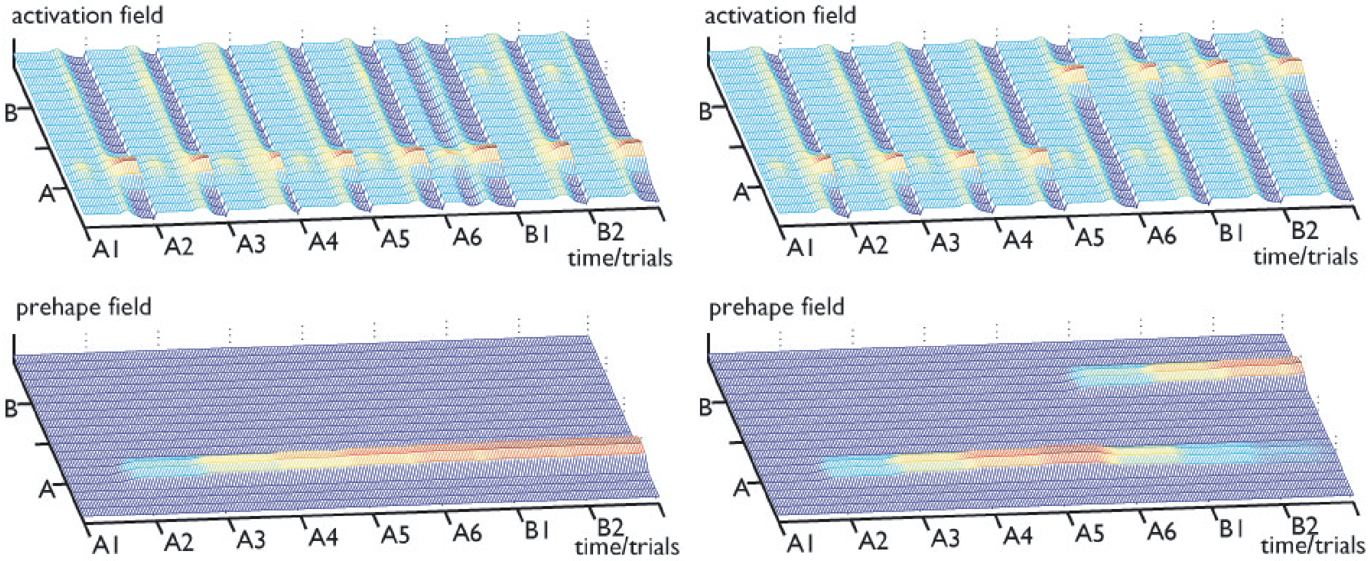


Figure 15: Dynamic fields representing the reaching direction in the A not B task are shown here (top) together with the associated preshape fields (bottom). Along the time axis, the time structure of the task is reflected by a sequence of 6 A trials (stimulation at A) and 2 B trials (stimulation at B). On each trial a peak is generated either at A or at B, and leaves a matching memory trace, which in turn preshapes the field on the next trial. This generates a tendency for perseveration both in the conventional sense, perseveration to A after a number of reaches to A on A trials (left) as well as in a new sense, correctly reaching to B after a number of reaches to B on A trials (right).

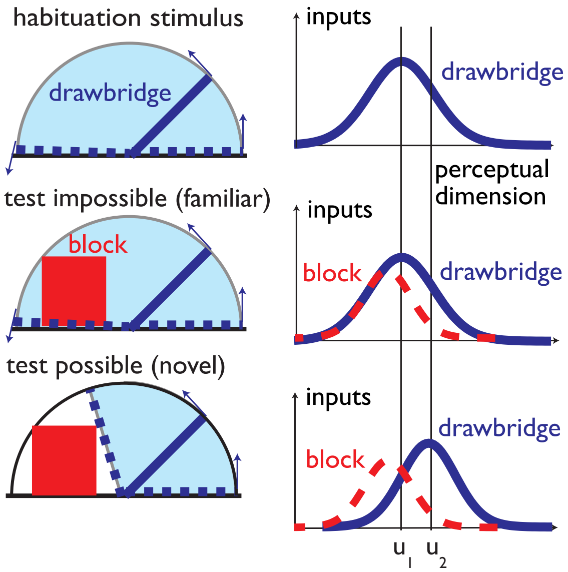


Figure 16: In the drawbridge paradigm, infants habituate to a flap moving 180 degrees (top left in the figure, the infant would observe this stimulus from the right). They are then tested with two different stimuli: the "impossible" (middle left) involves the same 180 degree flap motion (hence familiar), although initially a block is visible that would appear to block the path of the flap; and the "possible" (bottom left), in which flap motion is stopped at an angle consistent with the block remaining in place (hence novel). A perceptual dimension spanning these stimuli could be the spatial location in depth at which movement or perceptual structure is seen. The moving flap provides broad input (solid) across this spatial dimension, but centered for a larger mean depth for the 180 degree motion (top two panels on the right) than for the stopped motion (bottom right). The block provides added input (dashed) at a larger depth (bottom to panels on the right). The two marked locations (vertical bars) are represented by two activation variables, u_1 and u_2, in the dynamic model of Schöner and Thelen (2006).

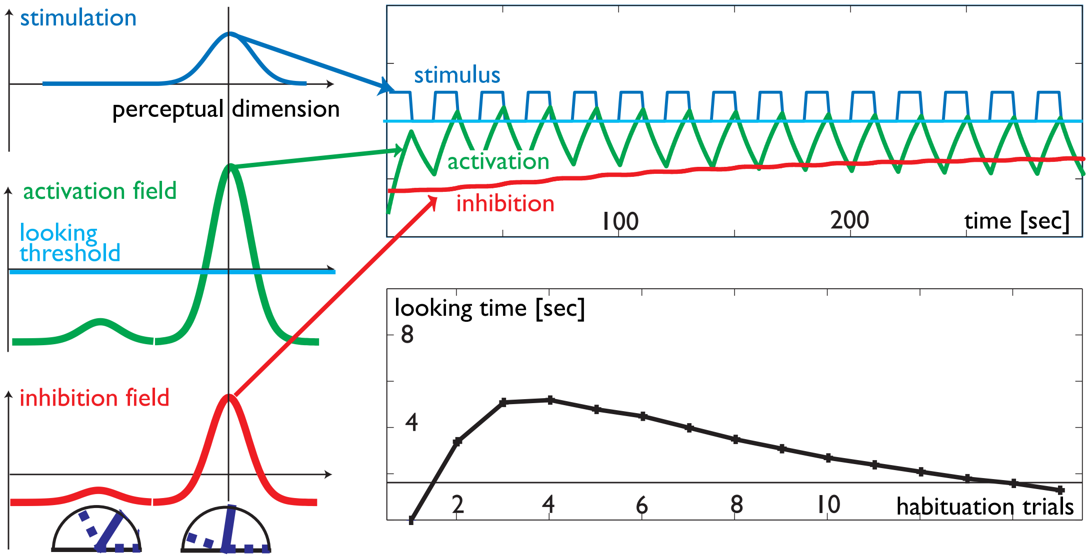


Figure 17: A dynamic field model of habituation consists of two layers, an activation field (left middle) that receives sensory input (left top) and represents the propensity to look at the stimulus, and an inhibition field (left bottom) which is driven by the activation field which it in turn inhibits. These fields span perceptual dimensions such as the spatial location of perceived visual motion in the drawbridge paradigm hinted at here (sketches at bottom). When a particular stimulus (here the 180 degree drawbridge motion) is presented repeatedly (stimulus trace top right), activation at the corresponding location in the field (activation trace top right) is driven upward, which promotes looking (looking time shown on bottom, computed as amount of time during which activation is above zero). This is counteracted on a slower time scale by inhibition (inhibition trace top right), which reduces activation and thus looking time.

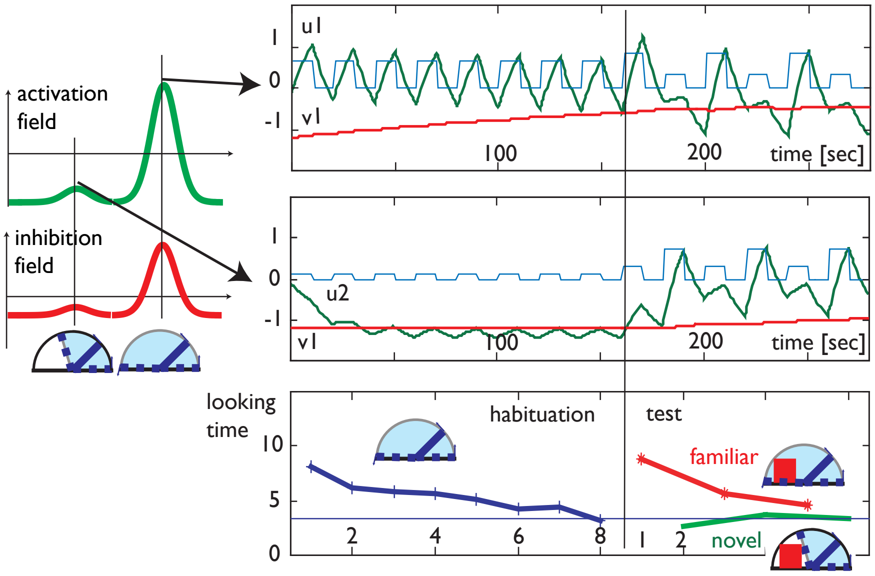


Figure 18: Modelling the evolution of activation at two locations in the field that represent the two drawbridge test stimuli (left column), an account for dishabituation is given. During the habituation phase, activation and inhibition build for u_1 (top), leading to decrease of looking (bottom) to criterion (horizontal line). During test, the block provides a boost of input to both variables, but dishabituation is stronger to the familiar than to the novel stimulus because the familiar stimulus overlaps more with the habituation stimulus and, when the familiar stimulus is tested first, activation starts from a higher level.

when it comes first, as it may build on a larger level of prior activation. On this basis, the model accounts for the full scope of effects in the habituation literature including age differences, fast vs. slow habituators, and graded effects when the overlap between stimuli is varied (Schöner, Thelen, 2006).

What about development? The DFT account for perseverative reaching outlined above postulates that what changes during the development is the effective neural dynamics so that older infants reach the dynamic regime that supports sustained activation more easily than do younger infants. Similarly, the DFT model of habitation accounts for age differences through differences in the dynamics. Similar as for the analogous postulate for learning, this link of developmment to dynamics has important consequences. Development as change of dynamics provides, for instance, for a form of automatic generalization across a range of behaviors that involve related elements of neural dynamics. The spatial precision hypothesis, for instance, predicts a gradual shift to sharper neural interaction functions which increasingly take on a stronger role compared to input (Schutte, Spencer, Schöner, 2003; Spencer this volume). This has consequences for a range of different behaviors including spatial memory, spatial discrimination, and the capacity to maintain object-centered reference frames. In all of these behaviors, correlated changes are predicted to occur during development.

Another implication comes from the emergence idea. If it is the dynamics that develop, then developmental processes must be analyzed by investigating the causes of behavioral patterns, not just the patterns themselves. Thus, at a particular stage of a developmental process the system may have a certain propensity to generate patterns of sustained activation supporting working memory, but whether or not such patterns are actually stable may depend on the perceptual context, the short-term behavioral history, or the total amount of stimulation provided.

But how about the developmental processes through which the dynamics change? To date, the DFT approach has not addressed these processes at a level of specificity and detail comparable to that achieved in capturing the changes themselves. This is clearly a task for the future. In contrast, Paul van Geert (1991; 1998) and colleages have used dynamical systems thinking primarily at the time scale of development. The idea is to think of development as a growth process, which can be described abstractly as the temporal evolution of a dynamical system. Qualitatively different patterns of temporal evolution can then be accounted for: continuous growth, characteristic variation of growth rate from slow to fast to slow (S-shaped), and oscillatory evolutions. Moreover, coupling among various growth processes may induce complex patterns of growth, such as stage-like alternations between slow and fast rates of change.

A dynamic description of the process of development in such global terms requires the characterization of the state of the developing system that is global enough to capture all that is involved in the development of a competence and all that is affected by such development. This is much more difficult to make explicit than the task-specific state variables and dimensions used in DFT. Typical empirical variables that the growth models have linked up to are, for example, sizes of vocabularies or the frequency of use of certain competences. These variables are relatively disembodied and not easily linked to the real-time behavior at any particular moment in time. Thus, it is not easy to conceive of specific mechanisms that would account for how real-time behavior drives development so that it follows the dynamical systems laws. Moreover, the question arises, how a developmental state is established on the developmental time scale, while on the time scale of behavior, the system is exposed to many different stimuli and is engaged in many different behaviors and behavioral changes. The mathematical models are typically used primarily to demonstrate how different qualitative features of developmental processes may emerge from simple dynamical laws, not so much to account in quantitative detail for specific experiments. Van der Maas and colleagues (1992) have used catastrophy theory to characterize the emergence of stage in development in this spirit.

Consistent with this kind of use of models, Paul van Geert and colleagues emphasize that dynamical systems thinking may be powerful at the level of metaphor (see van Geert, Steenbeek, 2005, where the relationship between the "Bloomington" and the "Groningen" approaches is discussed in some detail, including a new example of using the growth law approach on the faster time scale of behavior). Dynamical Systems as metaphor has been an important source of new ideas, new methods of analysis, new questions, in both flavors of the approach. An example is the emphasis on variability as a measure of underlying stabilization mechanisms, important both on the time scale of behavior (Thelen, Smith, 1994) and on the developmental time scale (van Geert, 1998).

5 Conclusions

In this overview, I have shown how four concepts, attractors and their instabilities, dynamic activation fields, and the simple learning mechanism of a memory trace together may enable an embodied theory of cognition, that is, a theory that takes seriously the continuous link of cognitive processes to the sensory and motor surfaces, that takes into account that cognition emerges when systems are situated in richly structued environments, that is open to an understanding of learning, and that is based on neuronal principles.

Connectionist thinking shares many of the same ambitions and core assumptions, in particular, the committment to using neuronal language to provide accounts for function and learning. In some instances, connectionist models have used the exact same mathematics as dynamical systems models (e.g., Usher, McClelland, 2001) and many connectionist networks are formally dynamical systems. Connectionism is really about change, about how under the influence of a universe of stimuli, statistical regularities may be extracted by learning systems, which may reflect these properties in the structure of its representations (Elman et al., 1997). Signatures of that process include accounts for learning curves, so that connectionist models speak to the evolution of systems on the time scale of development. On the other hand, some connectionist models also speak to the flow of behavior on its own time scale, the generation of responses to inputs provided in a given task setting. Connectionism thus straddles the two time scales that the two flavors of dynamical systems ideas have studied, for the most part, separately.

The detailed comparison of Dynamical Systems thinking and connectionism is no the goal of this overview (see Smith and Samuelson, 2003, and much of this book). Maybe I can say just this much: The problem of bringing Dynamical Systems thinking to bear on the process of development itself may be the most constructive point of convergence. While connectionist networks have shown how the statistics of stimulation drive the development of representations, the models to date have not taken into account how infants and children control their own stimulation through their own behavior, how individually different developmental trajectories may emerge from that fact, and how the space-time constraints of behavior determine what can be learned when. DST provides the framework within which online linkage to the environment can be modelled and how cognitive functions emerge from instabilities. By combining the structured account for real-time behavior of DST with the neural substrate for learning provided by connectionism, maybe we will be able to understand not only what develops, but how individuals through their own behavior generate the experiences on which development is based.

References

[1] S Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] R Baillargeon, E Spelke, and S Wassermann. Object permanence in 5-month-old infants. *Cognition*, 20:191–208, 1985.

[3] A Bastian, G Schöner, and A Riehle. Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience*, 18:2047–2058, 2003.

[4] E Bicho and G Schöner. The dynamic approach to autonomous robotics demonstrated on a low-level vehicle platform. *Robotics and autonomous systems*, 21:23–35, 1997.

[5] P Dayan and L F Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* MIT Press, 2001.

[6] Evelina Dineva. *Dynamical Field Theory of Infants Reaching and its Dependence on Behavioral History and Context.* PhD thesis, Ruhr–Universität Bochum, 2005.

[7] Daniel Durstewitz, Jeremy K. Seamans, and Terrence J. Sejnowski. Neurocomputational models of working memory. *Nature Neuroscience Supplement*, 3:1184–1191, 2000.

[8] Jeffrey L Elman, Elizabeth A Bates, Mark H Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett. *Rethinking Innateness — A Connectionist Perspective on Development.* The MIT Press, Cambride, MA, 1997.

[9] W Erlhagen, A Bastian, D Jancke, A Riehle, and G Schöner. The distribution of neuronal population activation (DPA) as a tool to study interaction and integration in cortical representations. *Journal of Neuroscience Methods*, 94:53–66, 1999.

[10] W Erlhagen and G Schöner. Dynamic field theory of movement preparation. *Psychological Review*, 109:545–572, 2002.

[11] S Grossberg. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52:217–257, 1973.

[12] S Grossberg, C Pribe, and M A Cohen. Neural control of interlimb oscillations. I. Human bimanual coordination. *Biological Cybernetics*, 77:131–140, 1997.

[13] H Haken. *Synergetics–An Introduction*. Springer Verlag, Berlin, 3 edition, 1983.

[14] H S Hock, G Schöner, and M A Giese. The dynamical foundations of motion pattern formation: Stability, selective adaptation, and perceptual continuity. *Perception & Psychophysics*, 65:429–457, 2003.

[15] B A Kay, J A S Kelso, E Saltzman, and G Schöner. The space-time behavior of single and bimanual rhythmical movements: data and model. *Journal of Experimental Psychology: Human Perception and Performance*, 13:178–192, 1987.

[16] J A S Kelso. *Dynamic Patterns: The Self-Organization of Brain and Behavior*. The MIT Press, 1995.

[17] K Kopecz and G Schöner. Saccadic motor planning by integrating visual information and pre-information on neural, dynamic fields. *Biological Cybernetics*, 73:49–60, 1995.

[18] L Perko. *Differential Equations and Dynamical Systems*. Springer Verlag, Berlin, 1991.

[19] G Schöner. Learning and recall in a dynamic theory of coordination patterns. *Biological Cybernetics*, 62:39–54, 1989.

[20] G Schöner and J A S Kelso. Dynamic pattern generation in behavioral and neural systems. *Science*, 239:1513–1520, 1988.

[21] G Schöner, K Kopecz, and W Erlhagen. The dynamic neural field theory of motor programming: Arm and eye movements. In P G Morasso and V Sanguineti, editors, *Self-Organization, Computational Maps and Motor Control, Psychology Series, Vol. 119*, pages 271–310. Elsevier-North Holland, 1997.

[22] G Schöner and E Thelen. Using dynamic field theory to rethink infant habituation. *Psychological Review*, in press.

[23] G Schöner, P G Zanone, and J A S Kelso. Learning as change of coordination dynamics: Theory and experiment. *Journal of Motor Behavior*, 24:29–48, 1992.

[24] A R Schutte, J P Spencer, and G Schöner. Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development*, 74:1393–1417, 2003.

[25] L B Smith and E Thelen, editors. *A dynamic systems approach to development: applications*, Boston, U.S.A., 1993. The MIT Press, Bradford Books.

[26] L B Smith and Esther Thelen. Development as a dynamical system. *Trends in Cognitive Sciences*, 7(8):343–348, 2003.

[27] Linda B Smith and Larissa K Samuelson. Different is good: Connectionism and dynamic systems theory are complementary emergentist approaches to development. *Developmental Science*, 2003.

[28] J P Spencer and G Schöner. Bridging the representational gap in the dynamical systems approach to development. *Developmental Science*, 6:392–412, 2003.

[29] E Thelen, G Schöner, C Scheier, and L Smith. The dynamics of embodiment: A field theory of infant perseverative reaching. *Brain and Behavioral Sciences*, 24:1–33, 2001.

[30] E Thelen and L B Smith. *A Dynamic Systems Approach to the Development of Cognition and Action*. The MIT Press, A Bradford Book, Cambridge, Massachusetts, 1994.

[31] T P Trappenberg, M C Dorris, D P Munoz, and R M Klein. A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, 13(2):256–271, 2001.

[32] Marius Usher and James L McClelland. On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, 108:550–592, 2001.

[33] van der Maas H L and Molenaar P C. Stagewise cognitive development: an application of catastrophe theory. *Psychological Review*, 99(3):395–417, 1992.

[34] P van Geert. A dynamic systems model of cognitive and language growth. *Psychological Review*, 98:3–53, 1991.

[35] P van Geert. A dynamic systems model of basic developmental mechanisms: Piaget, vygotsky, and beyond. *Psychological review*, 105(4):634–677, 1998.

[36] P van Geert and H Steenbeek. Explaining after by before — Basic aspects of a dynamic systems approach to the study of development. *Developmental Review*, 25(3-4):408–442, 2005.

[37] Xiao-Jing Wang. Synaptic basis of cortical persistent activity: the importance of nmda receptors to working memory. *Journal of Neuroscience*, 19(21):9587–9603, 1999.

[38] Claudia Wilimzig, Stefan Schneider, and Gregor Sch öner. The time course of saccadic decision making: Dynamic field theory. *Neural Networks*, 2006.

[39] H R Wilson. *Spikes, Decisions, and Actions: Dynamical Foundations of Neurosciences.* Oxford University Press, 1999.

[40] P G Zanone and J A S Kelso. The evolution of behavioral attractors with learning: Nonequilibrium phase transitions. *Journal of Experimental Psychology: Human Perception and Performance*, 1992.

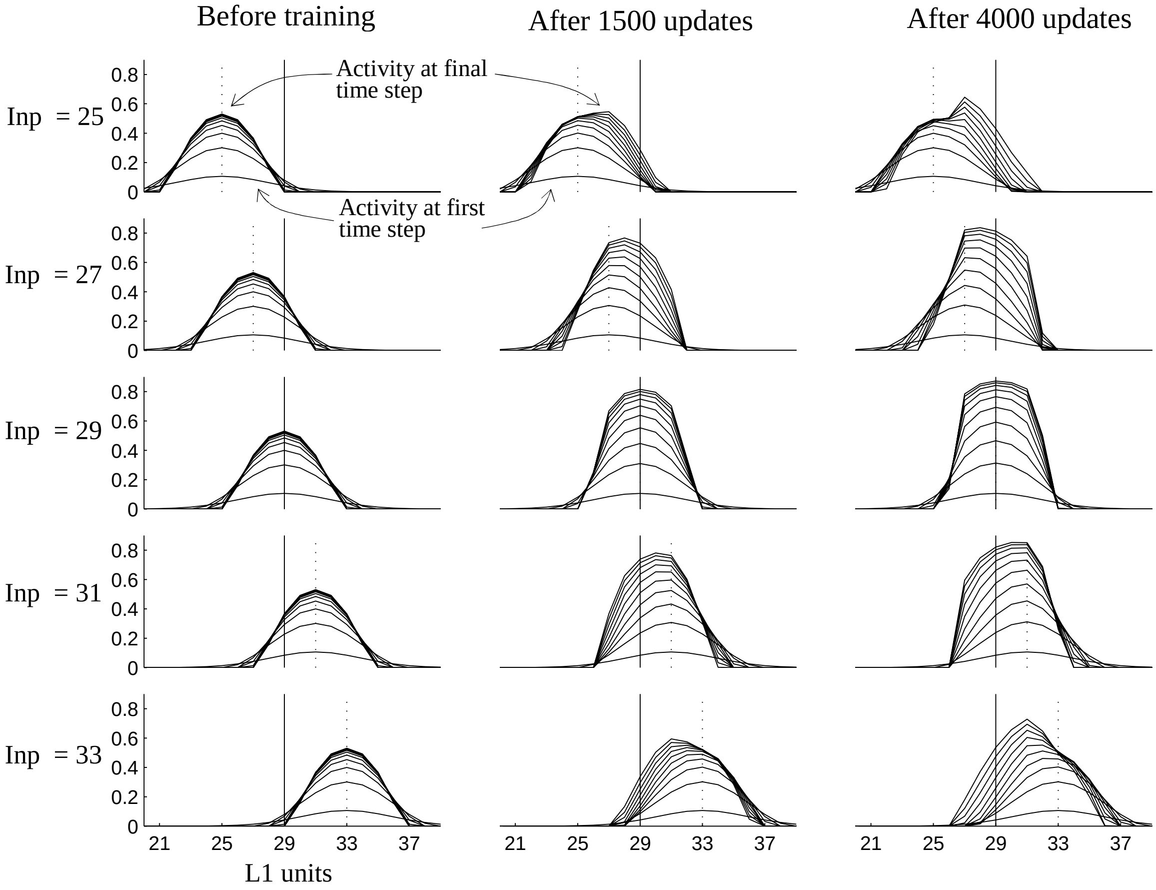


Figure 4. L1 activity at different stages of training in the Vallabha-McClelland model. Each pane shows the evolution of L1 activity for one input location. "Inp = *x*" indicates that the input is centered over L1 unit *x* (the vertical dotted line). The vertical solid line indicates the center of the category (input location 29).

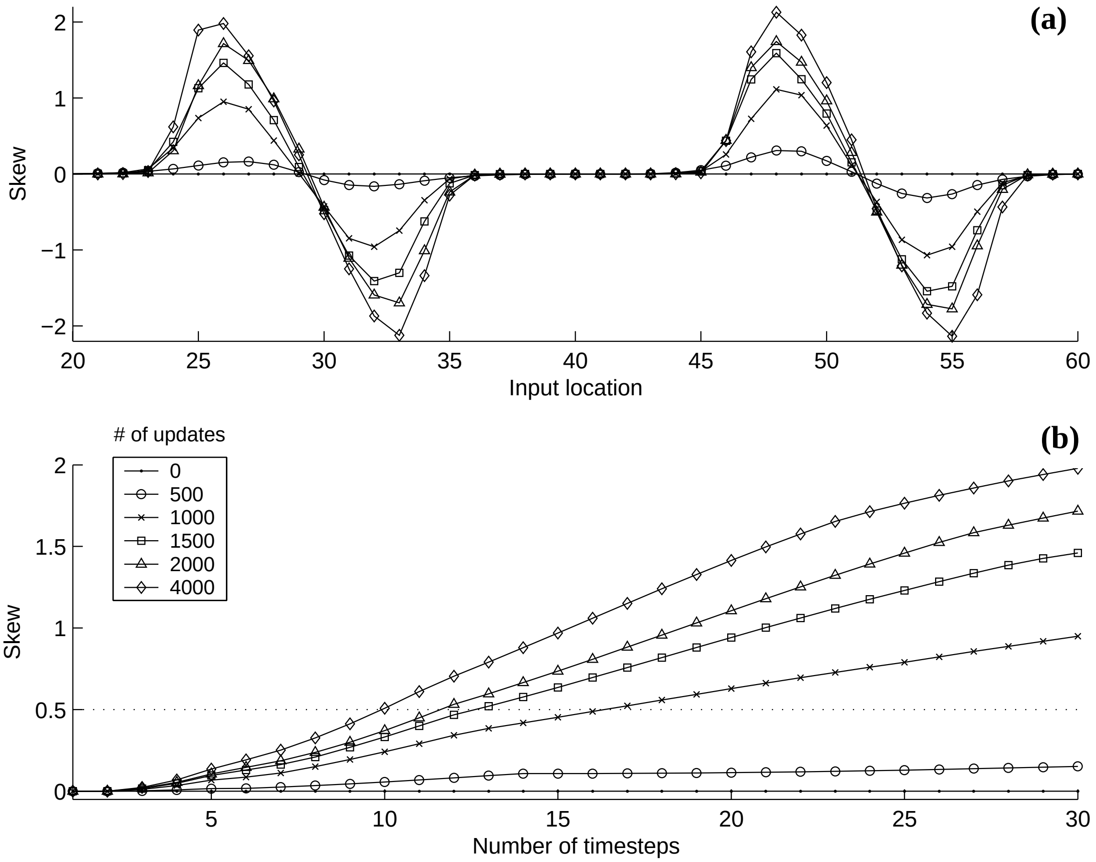


Figure 5. (a) The developmental "phase plot" for the Vallabha-McClelland model. For each input location *k*, the skew (≡ *x* – center of mass at final timestep c_k) is shown at different stages of training. See text for details. (b) The development of skew as a function of processing time and training for input location 26. The horizontal dotted line shows that a particular level of skew is acheived more rapidly after training.

Time Scales in Connectionist and Dynamical Systems Approaches to Learning and Development

Karl M. Newell[1], Yeou-Teh Liu[2], and Gottfried Mayer-Kress[1]

[1]Department of Kinesiology

The Pennsylvania State University

University Park, PA 16802, USA

[2]Department of Physical Education

Taiwan Normal University

Taipei, Taiwan

Submitted 02/08/06

Key Words: Time Scales, Dynamical Systems, Connectionism, Learning, Development

Mailing Address:

K. M. Newell
College of Health and Human Development
Pennsylvania State University
201 Henderson
University Park, PA 16802, USA

Tel: (814) 863-2426
Fax: (814) 865-3282
kmn1@psu.edu

Abstract

Epigenetic landscapes have been proposed in both connectionist and dynamical systems frameworks to capture the hallmark feature of learning and development: namely, persistent change over time. Here we focus on the dynamical origins of the time scales of change within our epigenetic landscape models for motor learning and development. We examine the role of constraints in channeling the time scales and the contrasting approaches of dynamical systems and connectionist models to stability and the distance to the goal as elevation function of behavioral dimensions in the landscape. Three case studies of landscapes and times scales in motor learning and development illustrate key aspects of the approach.

Time Scales in Connectionist and Dynamical Systems Approaches to Learning and Development

Connectionist and dynamical systems frameworks have both invoked the construct of time scales in theorizing about the central problem of the laws of change over time in learning and development. Indeed, the construct of time scales is fundamental in a number of contemporary theories and models from both metaphorical camps (e.g., Elman et al., 1996; Thelen & Smith, 1994). Nevertheless, while the phrase 'multiple time scales of change' has been used with increasing propensity in contemporary accounts of learning and development there is actually a paucity of direct theoretical or empirical investigation by either the connectionist or dynamical frameworks into the core construct of time scales.

In this chapter we examine the origins to the construct of time scales as expressed in Waddington's (1957) metaphor for development of the epigenetic landscape, and its subsequent use in connectionist (e.g., Elman et al., 1996) and dynamical (Kugler, 1986; Muchisky, Gershkoff-Stowe, Cole, & Thelen, 1996; Thelen & Smith, 1994; van Geert, 1994) system frameworks as a metaphor for change in learning and development. From this background we outline the dynamical origins to time scales in behavioral change through a consideration of the time scales arising from growth/decay and oscillatory processes in living and nonliving systems. A related theme is how the constraints to action (evolutionary, ontogenetic, physiological and task goals) channel the dynamics of behavior, and the rates of change of learning and development in systems at all levels of the organism.

The theoretical framework for the ideas on time scales presented here is our epigenetic landscape model for motor learning and development that is constructed as a low dimensional approximation of a high or infinitely dimensional complex dynamical system that possess a score, error, or fitness function (Mayer-Kress, Liu, & Newell, 2006; Newell, Liu, & Mayer-Kress, 2001, 2003). This approach has been driven by the overarching principles of self-organizing systems (Haken, 1983; Iberall, 1972; Prigogine & Nicolis, 1971) together with the operational developments of nonlinear dynamical systems in physical nonliving systems (Boccara, 2003). Indeed, the fields of motor control (Haken, Kelso, & Bunz, 1985; Kelso, 1995; Kugler & Turvey, 1987); motor learning (Schöner, 1989), and motor development (Thelen, Fogel, & Kelso, 1987) were influenced by the self organization metaphor and the principles and methods of nonlinear dynamics earlier than many areas of psychology. Interestingly, and in contrast, the movement domain has been relatively uninfluenced by the ideas of connectionism (although see Simms, 1994).

In our dynamical framework to the evolution of behavioral change in motor learning and development we share the views of connectionist approaches (Elman et al., 1996) in regard to the important role of the task goal (target) in organizing the landscape. Here we contrast the role of stability as emphasized in some dynamical systems approaches and that of task goals (zero error) in the connectionist frameworks to the determination of the time scales of change. We conclude by illustrating briefly three case studies of the time scales of change in our epigenetic landscape framework to motor learning and development.

What is a Time Scale?

The construct of time scales is being increasingly used in neuroscience and psychology, and related subdomains of the study of human behavior. And as noted above, the phrase multiple time scales is now common in dynamical systems and connectionist approaches to behavior and physiology. The concept of multiple time scales in the extreme situation of infinitely many time scales in physiology and behavior is most evident in phenomena that scale as a 1/f like fractal process (Bassingthwaighte, Liebovitch, & West, 1994; Ward, 200). Nevertheless, the origin and meaning of the term time scale is rarely articulated even in the life science fields of human behavior where it is used, perhaps because its long historical use in the development of the concept of time is seen to make such an elaboration unnecessary.

The concept of time has been refined throughout history in part due to the development of new time scales. The circadian rhythm of earth rotation is the most obvious indicator of the passage of time and the time scale of the day is the fundamental element of all calendars. Not all frameworks of time, however, are based upon the rotation of the earth. Indeed, the forms of time known as dynamical, atomic and coordinate time are not based on the idea of the day in the standard sense (Halliday, Resnick, & Walker, 2005), but they do relate to properties of dynamical systems.

In the traditions of dynamical systems there are two types of idealized motions that naturally introduce the construct of time scales (Haken, 1983; Kaplan & Glass, 1995). These motions have either periodic oscillations (such as the rotation of the earth) or growth-decay at a constant rate (such as in radio active decay used for instance in carbon dating of fossils). In oscillatory systems, the period (inverse of frequency) determines the intrinsic time scale of the system. In growth-decay systems the intrinsic time scale is the inverse of the growth/decay rate sometimes also expressed as “doubling” or “half life” times.

For linear dynamical systems these two classes of behavior and combinations thereof are the only forms of movement observed apart from the trivial dynamics of a non-moving fixed point. They are also the basis of almost all known clocks: earth rotation, a swinging pendulum, the vibrations of a quartz crystal, and cesium atoms represent clocks based on periodic movements. Water clocks and carbon dating as well as the red-shift of distant galaxies are examples of methods based on constant rates of change that are used to measure time at a vast range of scales. For both types of exponential trajectory a variable $x(t)$ at time t can be expressed as a complex exponential function:

$$x(t) = A\, e^{(\gamma + i\omega)t} \qquad (1)$$

Where A is the position at time $t = 0$, the real part γ of the exponent is the growth/decay rate, and the imaginary part ω is the frequency of the oscillation. The reciprocal of the growth or decay rate is the intrinsic or characteristic time scale

The two types of idealized motion, oscillatory and growth/decay are fundamental to describing behavior close to a fixed point, although nonlinear dynamical systems have a much richer repertoire of trajectories. The fixed point is the mathematical concept that is associated with the equilibrium state of the system (Kaplan & Glass, 1995; Strogatz, 1994). Fixed points correspond to the absence of motion as in a pendulum at rest. Furthermore, close to fixed points, a nonlinear system can be approximated to be linear, so that the motion of the trajectory can be characterized by the exponential function of

Equation 1. Thus, the time scale within the growth or decay process to a fixed point is characterized by the time it takes for the dynamics to double/half the distance to the fixed point. In this sense then, a time scale is not the time of an event as it is usually interpreted, but a particular duration (time) of a *dynamical* growth/decay process (Note that reduction of error or distance to the goal is formally a decay process.)

This set of basic assumptions can be developed so that what is called in learning and development the multiple time scales of change can be interpreted as originating from the system's trajectory on an *evolving* attractor landscape. Different attractor landscapes and their unique evolution of the pathway of change over time in terms of bifurcations and transient phenomena can lead to different patterns of change at the behavioral level, including multiple time scales as expressed in learning and development. For example, the laws of learning in exponential, power law and s-shaped change can be mapped to the landscape interpretations of attractor dynamics (Newell et al., 2001) as can the development of the fundamental movement patterns in infancy (Newell et al., 2003).

Thus, the different functions of change observed at the behavioral level in learning and development map to different dynamical phenomena and hold a particular meaning in the determination of the time scales of change. For example, the power law when interpreted formally has an infinite number of time scales within the range established for the phenomena at hand (Schroeder, 1991). The exponential has a single time scale given its single proportional rate of change. Furthermore, the function for the change in behavior over time might reflect a small number of characteristic exponential time scales but not actually reflect a power law. Thus, the phrase multiple time scales of change has multiple potential instantiations that are not well understood in learning and development.

The traditions of curve fitting and the evaluation of the nature of change though percent of variance criteria are prone to false interpretations of the time scales of change (Liu, Mayer-Kress, & Newell, 2003; Newell et al., 2001; Pitt & Myung, 2002). First, the universality of the power law in learning (Newell & Rosenbloom, 1981) is not as apparent as has long been assumed (see Gallistel, Fairhurst, & Balsam, 2004; Heathcote, Brown, & Mewhort, 2000; Newell et al., 2001). Second, the traditional practice of averaging data over subjects and trials can also lead to a false interpretation of the dynamical model of change (Newell et al., 2001). Third, even the information criterion of Akaike (Burnham & Anderson, 2002) treats the time scales of change in a neutral fashion in establishing criteria to evaluate functions and model selection with a different number of parameters. We claim that time-scales play a more significant role in characterizing a system than other system parameters such as some initial amplitude of a signal. Therefore, it is not appropriate to treat all statistical fit parameters in an indiscriminate way and thus view the data analysis tool as a black box. In summary, there appears not to be a single function of learning and that the veridical estimate of the persistent change over time has been biased by theoretical shallowness of taking at face value the output from fitting packaged programs to data sets.

It is our position that the time scales of change in learning and development need to be established from a theoretical background *prior to* curve fitting procedures not on a *post hoc* determination of model function from percent of variance decision making (Liu et al., 2003; Newell et al., 2001). This general theoretical orientation should be applied to properties of the change over time of individual behavior if a veridical estimate of the time scale(s) of change is to be made. Our approach to this is an a priori dynamical landscape model that treats seriously the role of time scales in determining the learning model (Newell et al., 2001). From this approach the modeling of individual learning data is emphasized but averaging of aspects of the data can be done judiciously from properties of the individual time scale(s) of change.

Constraints and Time Scales in Learning and Development

There have been a number of approaches to categorizing the primary constraints to action and their respective time scales. The general view, although expressed in somewhat different labels and terms in each approach, is that there are 4 classes of constraint to action that reflect orders of magnitude differences in their time scales of influence. These categories of constraint are evolutionary, ontogenetic (developmental), physiological (epigenetic biological) and task (Eckhardt, 2000; Mittenthal & Baskin, 1992; Waddington, 1957).

The time scale of evolutionary constraints is over the generations of the species and, although recognized as a fundamental category of influence to the learning and performance of individuals, it is not typically pursued theoretically or experimentally in studies of learning and development in context. Indeed, most emphasis is given to either the time scales of ontogeny or those arising from the local dynamics of physiology and biology (epigenetics) more broadly. In learning and developmental theorizing the role of the task in organizing the time scales is typically either minimized or ignored.

The time scales of the dynamics of biological systems are most usually approached (directly or indirectly) by a consideration of the rhythms of the system output. It is well established that the time scales of change in the rhythms at the multiple levels of biological analysis can occur over a wide range of scales in ontogeny from years to months, weeks, days, hours, minutes and even fractions of a second (Foster & Kreitzman, 2004; Luce, 1971). The time scales are usually shorter the more micro the level of analysis of the system but there are counter examples to this general rule of thumb. For example, synaptic density in the brain, telomere lengths – the end regions of DNA that shorten with each cell division and, therefore, constitute a molecular, biological clock with long time scales of years.

The rhythms of biological systems provide an intrinsic metric for the time scales of many basic biological phenomena across a range of levels of analysis of the system. The premise of our attractor landscape framework to learning and development is that the link between time scales and oscillatory and growth/decay phenomena can be elaborated to essentially all aspects of human behavior. This includes the range of learned behaviors that are evident throughout the life span and that have been the focus of different subdomains of learning and development.

In our framework as noted above the term time scale refers to the particular time duration of a process that is supported by the dynamical principles arising from oscillatory and growth/decay changes in the system. However, in learning and development the idea of time scales has also been used less restrictively as in the context of just referring to the so-called real time of an event, without any link to oscillatory or growth/decay processes. Given the context of the intrinsic dynamics of oscillatory and growth/decay processes it is intuitive how the real time of a behavioral action or episode can be viewed as an emergent property rather than a specified feature or property of an action.

The phrase multiple time scales of action has been introduced principally through dynamical accounts of learning and development (Thelen & Smith, 1984). However, the specific development of the construct of time scales to the problems of the persistent and transitory change in learning and development has been less forthcoming. The most prevalent orientation is to link in a metaphorical way the notion of time scales to the epigenetic landscape of Waddington (1957). We will develop the fundamental assumptions of this approach in the next section of the chapter but before doing that we outline the essence of two approaches to time scales in learning and development to provide some practical implications of the possibilities.

Van Geert (1994) has presented the case for the major transitions in development to reflect a fractal process. The fundamental rule that he examines is the fractal relation in which the frequency of behavioral change is a power function of their magnitude. In this view the developmental changes that have the greatest magnitude are those identified as stages and these are also small (very few) in frequency of occurrence. He outlined the basis of time scales for the major transitions or stages of development and uses the framework of Fischer and Rose (1993) to provide developmental examples of these changes in time scale. In van Geert's (1994) view the underlying principle for the fractal pattern of transitions is the logistic function.

In more of a life span perspective Lemke (2002) has examined the multiple time scales over many orders of magnitude in the social organization of learning. A key feature of this framework is the attempt to explain why some learning experiences contribute to long term patterns of action and personal development while some experiences are short term and their influence transitory. The center piece of this analysis is a table of time scales (in seconds) for human activity that goes from the very short time scales of chemical synthesis (10^{-5}), membrane process (10^{-4}), neural firings (10^{-3}), neuronal patterns (10^{-2}) to the very long time scales of world system change as in new cultures languages (10^{11}), ecosystem climate change (10^{12}) to evolutionary changes and the scale of human evolution (10^{14}). Embedded in the middle range of these extreme fast and slow time scales are human activities of vocal articulation (10^{-1}), utterance (1-10 s), exchange (2-10^{2}), episode (10^{3} , 15 min), lesson (10^{3}-10^{4}, hour), lesson sequence (10^{4}, 2.75 hours), school day (10^{5} , day), Unit (10^{6} , 11.5 days), semester/yr curriculum (10^{7} , 4 months), multiyear curriculum (10^{8} , 3.2 years), and lifespan educational development (10^{9} , 32 years). Lemke observes that the learning in social organization systems that has longer time scales tends also to operate in more spatially embedded systems. There does not appear to be any experimental evidence for the intrinsic dynamics of the time scales arising from these levels of social organization.

Thus, and in summary, we see that the term time scale refers to a period of time but it has been used in two different ways in learning and development. In the most generic and loose way it refers to the duration of a process or event. In the more formal interpretation from dynamical systems is relates to the time periods that emerge from the intrinsic oscillatory and growth/decay processes of the system. In either approach the time scales are assessed as a first pass in regard to their order of magnitude difference for different processes, such as features of learning and development.

Evolving landscapes as Metaphor in Motor Learning and Development

The origins of the dynamical interpretation of change in learning and development can be traced to the idea of an evolving epigenetic landscape (Waddington, 1957). Waddington was interested in embryogenesis and the landscape model was a metaphor for considering the dynamics of developmental growth and change. The idea of the epigenetic landscape has since been used to exemplify the emergence of a range of developmental phenomena (e.g., Connolly, 1986; Meinhardt, 1982) and also changes in gene expression triggered by behavioral changes such as exercise and practice in learning a task.

Insert Figure 1 about here

The landscape in Figure 1 illustrates the pathways of developmental change and the motion of the ball on the landscape reflects the developing phenotype. The fore to aft dimension is time or developmental age, the horizontal axis instantiates the emergence and dissolution of particular activities that hold dynamic equilibriua, and the slope of the surface captures the rate of developmental change. The stability of the system is inferred from the depth of the landscape wells. This 2 dimensional landscape in a 3D space (two geographical dimensions + one elevation dimension) is a low dimensional schematic of the basic dynamic processes of development that is derived from a high dimensional biological system. Waddington's theoretical approach had the properties of the landscape emerging from genetic control but the landscape metaphor for the dynamic processes can be elaborated to include the confluence of constraints from the organism, environment and task (Connolly, 1986).

The general features of the landscape have been used to capture the dynamic processes of change in motor development (Connolly, 1986; Kugler, 1986; Muchisky et al., 1996). Muchisky et al. (1996) have elaborated the Waddington schematic to that shown in Figure 2 to reflect a broader range of the dynamic changes evident in the development of the fundamental movement sequence in infants. This figure in essence shows a longer period of developmental time than implied in the original Waddington schematic and the potential for the evolution and dissolution of a broader range of stable states with the passage of time. Indeed, the image here could reflect the learning of actions and tasks throughout the lifespan and the spiral of stability and instability that reflects the evolving nature of the dynamics of learning and development.

Insert Figure 2 about here

Figure 3 shows a hypothetical error surface from a neural network (Elman et al., 1996). These authors pointed up the similarity in the features of the artificial network landscape to that of the original Waddington epigenetic landscape (contrast to Figure 1). In the artificial network image of Figure 3 the phenotype is the evolving set of connection strengths between artificial synapses. And, as in the Waddington landscape, the network seeks the “low ground” of error in its weight space. There are often many pathways to equivalent solutions – an instance of, equifinality (Bertalanffy, 1975). Elman et al. (1996) claimed that the similarity between the natural and artificial landscapes is more that a coincidence in that they reflect fundamentally the same process. More importantly, for the purposes of this chapter, it follows that there is common ground by the dynamical systems and connectionist frameworks in considering the role of time scales in learning and development.

Insert Figure 3 about here

The images of the different versions of the epigenetic landscape give a common view to the metaphor of dynamical processes in learning and development. But, like many metaphors, when one pushes beyond the surface interpretations and analogies, difficulties and problems begin to emerge. But a more formal examination of the dynamical processes is important if we are to go beyond the metaphor (van der Maas, 1996) and develop formal models of the processes of learning and development.

In Newell et al. (2003) we discussed several limitations to the theoretical and experimental messages inherent in the landscape images of the above examples. Here we re-emphasize briefly some of these limitations to the extant landscape metaphors with emphasis by way of example to motor learning and development, although we would propose that the points are general to the landscape modeling of change in all behavior contexts. This section also provides some background to the subsequent consideration to the contrasting roles of distance from the goal ("error" in neural net works) and stability (dynamical systems) in landscape models.

1. The standard landscape image has the most stable state at the beginning of phenotypical development dissolving with the passage of time. Thus, in motor

development, for example, it implies that lying down dissolves as a stable activity. Yet, the postural state of lying down remains stable through the life span.

2. The horizontal dimension captures the qualitative stable states as reflected in posture, locomotion and manipulation. However, the horizontal or activity dimension is multidimensional. Without considering the stability of the activities in a multidimensional framework it is difficult to asses the dynamical relations between the stable states and how these relations change over developmental time.

3. The landscape models do not reflect that the dynamical landscape supporting physical activities can change due to the shorter time scales at which goals and intentions can emerge and change. Thus, the standard landscape metaphor does not capture sufficiently the wide range of time scales that influence behavioral change.

4. The landscape metaphor shows the collection of states that are possible at a particular age but it does not capture the relative dynamic properties of these states at any moment in time. Another way of saying this is that the landscape shows the potential scope of stable states at a point in time of development but not the actual nature of the landscape as it evolves in real time in the implementation of a given activity, in spite of the apparent concerns and claims for this.

Thus, the landscape metaphor is just that – a metaphor. It has limitations for the development of formal models be they connectionist or dynamical. These limitations also act as background features in mediating the important difference in the approach of the two frameworks with respect to the role of distance from goal and stability in the landscape models.

Distance-from-Goal and Stability in Landscape Models

The original use of landscape models in physics described how the potential or "free" energy of a system changes with the state of the system. In the simple case of a ball in geographical landscape, the state of the system is given by the position of the ball. In our context we assume the surface of the landscape being slippery but with high friction ("covered with molasses") so that the speed of the ball is always determined by the local slope of the landscape and, therefore, momentum of the ball need not be treated as an independent degree of freedom. That implies specifically that there is no "over-shooting", i.e. if the ball reaches a minimum, it will come asymptotically to a stop at the minimum.

A more realistic ball in a smooth landscape would have a finite speed and kinetic energy as it reaches the minimum. Therefore, it would continue moving uphill until friction losses and build-up of potential energy add up to the kinetic energy and the ball comes to a stop before it returns back towards the lowest point. In this case the approach to the minimum would be described by a damped oscillation ("bracketing") instead of the "creeping" approach in the landscape models of the context of epigenetic landscapes. At any point in the landscape we can calculate the free-energy of the system namely the maximum mechanical work that the system can do (without friction i.e. at infinitesimally small speed) while sliding down the landscape to its lowest point. It is proportional to the elevation of that point in the landscape: points at higher elevations can do more work on the way down than points close to the bottom. In an anthropomorphic sense one can also say that the system "wants to go to the lowest point".

In the context of learning we, therefore, identify the lowest point in the landscape under consideration the “goal“ of the learning task. The position in the landscape represents a participant’s (potential) coordinated movement pattern whereas the elevation corresponds to a score function that becomes zero when the goal is reached. Therefore, we refer to the elevation in the landscape as “distance to the goal (d2g)” in the sense of score difference. The distance in the behavioral space, corresponding to the geographic distance in a horizontal direction describes how much the behavior must be changed in order to reach the goal performance.

As is clear from our experience in geographical landscapes, this distance depends on the slope of the landscape: If the slope is very steep close to the goal point, small changes in behavior yield big improvements in performance score. In physical terms a large slope, therefore, corresponds to a strong attractive force of the goal point. In the context of movement a shallow landscape makes it more difficult to recognize which (random) change of a movement pattern leads closer to the goal, because the change in the score value will be very small.

Thus, it is clear that the stability of a goal point (minimum in a landscape) is related to the slope of the neighboring part of the landscape. If we take a snapshot of a random collection of balls close to the goal or look at the trajectory of a ball close to the minimum under the influence of a random force, then we will observe that in a minimum with steep surroundings the spread of the distribution of probable locations of the system is smaller than in the case of a shallower minimum. Thus we can characterize the (linear) stability of a minimum by the width of a random distribution of points under the influence of fluctuations.

In some discussion of landscapes we see a different notion of stability, namely the depth of the potential well: The deeper the well the more stable the system. This notion is related to the amount of work it would take to move the system out of the well (for now we call it "integrated stability" in the context of chemical reaction, it is also known as "activation energy") whereas linear stability or the steepness of the slope indicates how much maximum force is necessary to leave the well. Close to a fixed point linear stability is calculated from how much the restoring force increases with the distance from the fixed point. In mathematical terms in the context of a potential landscape it is equivalent to the eigenvalues of the Jacobian matrix of the force field **F(x*)** for a fixed point **x***. If the eigenvalues are both (in a two-dimensional landscape) negative, then the fixed point is stable and direction corresponding to the eigenvalue with larger magnitude has a steeper increase in slope. Figure 4 shows a special example of a Gaussian peak superimposed by a Gaussian valley with smaller range. The result resembles a volcano with an asymmetric crater.

Insert Figure 4 about here

The explicit equation for the potential landscape shown in Figure 4 is given by:

$$V(x, y) := e^{-\left(\frac{x^2}{\sigma_x{}^2} + \frac{y^2}{\sigma_y{}^2}\right)} - r\,e^{-\left(\frac{x^2}{\tilde{\sigma}_x{}^2} + \frac{y^2}{\tilde{\sigma}_y{}^2}\right)}$$

Where $r = 0.9$, $\sigma_x = 1$, $\sigma_y = 1.5$, $\tilde{\sigma}_x = 0.2$, $\tilde{\sigma}_y = 0.5$

Both local and integrated stability of a fixed point contribute to the attractiveness of the goal.

Note, however, that both notions of stability are a function of and can be calculated from the elevation function. Knowing the V or d2g for every point in the landscape, therefore, provides us with the stability information of a goal state but many different landscapes can have local minima (goals) that have identical stability indicators both for local and integrated stability. This is one manifestation of the concept of degeneracy in landscapes reflecting the many-to-one nature of mappings from locations to elevations.

Case Studies of Landscapes and Time Scales in Learning and Development

In this last section of the chapter we present 3 case studies of epigenetic landscapes from our work in motor learning and development. The examples focus on the acquisition of different motor skills and at different ontogenetic stages. The emphasis is time scales and their role in determining the nature of the dynamical processes that captures the persistent change in behavior over time. Each example captures one of the principles about the role of time scales in learning and development outlined previously.

Case Study 1: The Development of Prone Progression in Infancy (Newell et al., 2003).

The emergence of the infant motor development sequence represents one of the most fundamental and pronounced sequence of changes in the dynamics of behavior over the life span. The motor development sequence was the focus of the early classic studies of development (Gesell, 1929; McGraw, 1943; Shirley, 1931) and interest in this developmental process has been revived by dynamical approaches to development (Kugler, 1986; Kugler, Kelso, & Turvey, 1982; Thelen & Smith, 1994).

Developmentalists were attracted by the intuitiveness of the dynamical constructs of stability and change and their face validity relevance to the study of the emergence of new movement forms.

Here we consider the landscape for the development of prone progression in the postural states of the infant (Gesell, 1929, 1946). Prone progression includes the progression of the initial posture of lying down, the subsequent emergence of chin up, sitting, crawling, creeping and standing. Gesell showed that the developmental progression of the fundamental activities of prone progression is nonlinear. In Newell et al. (2003) we developed a developmental landscape of attractor dynamics of this postural progression. In isolating the development of postural activities does not imply that we assume no relation to other activity classes of the motor development sequence. We also assume that our strategy has some general implications in principle to the consideration of all action subcategories. The phenomenological model is developed for a land-based surface of support such as the room of a house and is built on the empirical findings of Gesell (1946).

The landscape is developed from the consideration of head height and locomotion speed as being the "order parameters" for the emergence of prone progression. We use the term "order parameter" more loosely than in the traditions of synergetics but the nature of infant multidimensional system presents challenges to modeling that have not been faced to date in developments of the HKB model. We also recognize the redundancy that allows for many micro level states to realize a configuration in the order parameter space. The contrasting interpretation to be drawn from the double (Figure 5) and single (Figure 6) order parameter confirms the point made earlier of the importance of considering the multidimensional nature of the activity dimensions.

Insert Figures 5 and 6 about here

The dynamical landscape in Figure 5 shows several basic features about the emergence of prone progression. First, the landscape is similar to that of Waddington (1957) because it represents a prototypical landscape of the potential stable states that exist over developmental time. It does not capture the changing landscape that relates to intention and the implementation of action. Second, the shape of the potential well for each activity is consistent with the behavioral properties shown by Gesell (1946) in regard to the emergence and preservation of the activity over time. Third, the landscape reveals that there is little equivalence for the infant to reach a certain head height and locomotion speed. In most cases, the infant at the given developmental time has only one stable state to achieve a given point in the coordinate space.

The single order parameter of head height organizes the same behavioral data for the landscape in Figure 6. The time axis corresponds to developmental time. It is clear that the single order parameter is insufficient to capture the developmental pathway of change as the transition from standing to lying down would have to pass through the other activities of sitting, creeping and chin up in the single dimension case. Thus, the difference in interpretation of Figures 5 and 6 confirms the need to consider the multiple dimensions of activities in the organization of the time scales of the landscape.

Case Study 2: Landscape of Learning with Warm-Up Decrement

The most simple landscapes can be associated with a single target that is the only movement pattern that is attractive (local minimum in landscape) and it can be reached without any transitions to new movement patterns that involves changes in the landscape itself or external forces (teacher or coach instructions etc.). Let us assume for the moment that two order parameters contribute to the learning and its resulting performance improvement. Genericity arguments would predict that these dimensions are not degenerate and lead to two different time-scales. The situation is, therefore, locally that of Figure 7 but now the fixed point would possess unlimited integrated stability or activation energy, i.e. the rim of the crater in Figure 7 would have infinite elevation.

Insert Figure 7 about here

The theoretical movement trajectory, therefore, can be easily predicted: With the high descent rate associated with the fast time scale the system would quickly approach the bottom of the valley (slow stable manifold of the fixed point) and then continue the approach to the fixed point at a constant modest rate given by the slow timescale. This pattern of fast->slow transitions in the learning curve seems to be universal and also applies to higher dimensional landscapes where it can easily be mistaken by powerlaw curves that show the same fast->slow ordering of (infinitely many) timescales.

This theoretical case becomes slightly more complicated when we consider actual human behavior. As can be easily seen a constant rate of approach for any time-scale will only bring one exponentially close to but never reach the fixed point. Realistic timescales also make it obvious that there are very few if any coordination skills where a participant starts as a novice and reaches expert performance level within a single practice session.

This raises the issue of practice schedules that has been discussed in the motor learning literature for a long time. In contrast to a mechanical system where we can reasonably expect that we can continue a measurement curve more or less in the same condition if we start a new session, this is obviously not the case with human motor learning or even learning in general. Activities between sessions will influence the state of the participant upon resumption of practice. Moreover, practice activity itself will initiate epigenetic physiological processes that will continue long after the end of the practice session.

Thus, it is not surprising that if we reconstruct a trajectory in a learning landscape that the end point of a practice segment of one session will be significantly different from the starting point of the next segment of practice trajectory. Observations - going back all the way to Snoddy's (1926) ground breaking work - indicate that there is a class of learning situation, where the performance at the beginning of a practice session is considerably depressed ("warm-up decrement") but then recovers to a level that would be expected from a second, slower learning rate that reflects more the persistent changes due to practice. In Newell, Liu and Mayer-Kress (2005) we were able to implement a quantitative reconstruction of a four segment learning trajectory from Snoddy's original data. We showed that warm-up decrement is primarily caused by a change in initial conditions away from the fixed point in the fast dimension. Surprisingly the reconstruction also shows that during the rest period the state of the system has moved closer to the fixed point in the direction of the slow dimension.

These patterns and resulting quantitative results would be difficult to obtain without the framework of a landscape model associated with the different time scales. This perspective now leads to new questions and predictions about change in learning and development that will be tested in forthcoming publications.

Case 3: Landscape Model of Learning a Coodination Task

In the previous case we have seen that an ***n*** dimensional landscape generically (i.e. without degeneracy of the fixed point) will have ***n*** different time-scales given by the different eigenvalues of the fixed point associated with the attracting minimum in the landscape. This will describe the complete dynamics of the system under the condition that the parameters describing the landscape (including the eigenvalues of the fixed points) will be constant. This assumption is only approximately true to a degree that depends on the context of the specific application. In some situations learning might be modeled including a new time-scale (typically much longer) that is given by the dynamics of the landscape itself. This phenomenon has a true analogue in geographic landscapes where e.g. the timescales associated with movement of water in streams is much shorter than the geological landscapes that can alter the location of river-beds etc.

In the following we want to discuss an example that describes the learning of a specific motor-task related to the celebrated Haken-Kelso-Bunz (1985) model of synchronized finger oscillations. The order parameter is given here by the relative phase $\Delta\phi$ of the finger oscillators. We know that for high oscillation frequencies only the "in-phase” ($\Delta\phi=0$) oscillator is stable. At lower frequencies a second attractor with “anti-phase” oscillations ($\Delta\phi=\pi$) emerges. With special practice one can learn other phase relationships such as $\Delta\phi=\pi/2$. It is a common phenomenon in non-linear systems that close to transitions to new patterns of behavior, the dynamics of the system will depend on other dimensions such as the history of the system or previous states (“hysteresis”).

In our approach we model this situation by embedding the one-dimensional HKB landscape (x coordinate) into a landscape $V_{x0}(x,y)$ with an extra dimension (y coordinate) reflecting this influence. A new bifurcation parameter c measures how well a participant has learned the new task of finger oscillations at a 90 degree phase difference ($\Delta\phi=\pi/2$. Note that for $c=0$ our model contains the HKB model as a special case. The parameter σ is related to the range of attraction of the new fixed point. The dynamics in the new y direction is chosen such that the system is slowly damped and approaches the $y=0$ hyper-plane. The x variable is related to the phase difference $\Delta\phi$ such that a phase difference $\Delta\phi=\pi/2$ corresponds to a target of $x_0=1/4$. The explicit form of our generalized HKB system for 90 degree phase learning is given by:

$$V_{x_0}(x, y) = y^2 - \cos(2\pi x) - \frac{1}{4}\cos(4\pi x) - c\left(1 - \frac{y}{4}\right) e^{-\frac{(x-x_0)^2}{\sigma^2}} \tag{4}$$

In this interpretation of the landscape, which is different from the interpretation in the different context of Newell et al. (2001), we only have the in-phase attractor stable at the beginning of practice. Independent of the initial condition the system will converge to an in-phase oscillation (Figure 8a). After a certain amount of practice the landscape has deformed and a stable fixed point emerges at the desired $x_0=90$

degree attractor. Note that still the majority of initial conditions will converge to the in-phase attractor and special preparations are required to reach the new fixed point (Figure 8b). Finally, after further practice a large area of initial conditions will be attracted to the new fixed point and the chance of performing the task is significantly increased (Figure 8c). This landscape model would predict that initial conditions close to x=270 degrees would all be attracted to the in-phase fixed point.

So far we have treated the change in the landscape as a non-autonomous external modification of a control parameter. It is straight-forward, however, to extend the dimension of the autonomous state space by assigning a new dynamical equation to the evolution of the bifurcation parameter c.

The complexity of this case, however, is also enhanced compared to the previous cases because it includes a bifurcation or non-equilibrium phase-transition which is known to exhibit the phenomenon of "critical slowing down" i.e. the divergence of timescales at the critical point of the transition. Therefore, it might be difficult to extract quantitative estimates of the time-scales of change of the landscape dynamics from limited experimental data.

Insert Figure 8a,b,c about here

Concluding Comments

A central premise arising in our epigenetic landscape models of learning and development is that the construct of time scales provides a fundamental way to assess the core processes underlying the persistent change in behavior over time. The time scales provide a window into the dimensions of behavioral change and how these evolve over time. A consideration of time scales has allowed us to distinguish dimensions that are often confounded in other modeling approaches.

A point of distinction or at least emphasis between the dynamical and connectionist approaches to learning and development is in their respective use of the term time scale. Dynamical approaches give importance to the intrinsic dynamics of the system in determining time scales but this formality is less evident in the use of time scales in connectionist approaches. This distinction gives rise also to the stronger emphasis on stability in the dynamical systems modeling than in connectionist modeling. But, these theoretical frameworks may just be on different time scales of development and it will not be long before it is difficult to tell one framework from the other (Spencer & Schöner, 2003).

In connectionist networks literature little attention seems to be given to the mathematical shape of the learning/error curve. We claim that the landscape modeling and associated structure of the learning curve especially the presence of multiple time scales is a universal feature of natural and artificial learning systems. We would, therefore, predict that in carefully designed learning simulations it should be possible to extract different time-scales and use them to quantitatively reconstruct a connectionist landscape that goes well beyond the metaphor. At least issues of stationarity and limitations in data collection time should be easier to control in connectionist simulations than in the wet-ware of biological learners. In future publications we want to address the issue of epigenetic landscapes that describe the change in the learning system between practice sessions. Here issues of unlearning and memory consolidation appear to be of central importance in driving the time scales of learning and development. Again, a parallel investigation in connectionist simulations could prove most fruitful.

Author Notes

This work was supported in part by NSF #0114568.

References

Bassingthwaighte, J. B., Liebovitch, L. S. West, B. J. (1994). Fractal physiology. Oxford: Oxford University Press.

Boccara, N. (2003). Modeling complex systems. New York: Springer.

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodal inference: A practical information-theoretic approach. New York: Springer.

Connolly, K. J. (1986). A perspective on motor control. In M. G. Wade & H. T. A. Whiting (Eds.), Motor development in children: Aspects of coordination and control (pp. 3-22). Dordrecht: Martinus Nijhoff.

Eckhardt, R. B. (2000). Human paleobiology. Cambridge: Cambridge University Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking innateness: A connectionist perspective on development. Cambridge, MA: The MIT Press.

Foster, R. G., & Kreitzman, L. (2004). Rhythms of life: The biological clocks that control the daily lives of every living thing. New Haven: Yale University Press.

Gallistel, C. R., Fairhurst, S., Balsam, P. (2004). The learning curve: implications of a quantitative analysis. Proceedings of the National Academy of Science USA, 101, 13124-13131.

Gesell, A. (1929). Maturation and infant behavior pattern. Psychological Review, 36, 307-319.

Gesell, A. (1946). The ontogenesis of behavior. In L. Carmichael (Ed.), Manual of child psychology (pp. 295-331). New York: Wiley.

Haken, H. (1983). Synergetics: An introduction (3rd Ed.). Berlin: Springer-Verlag.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51, 347-356.

Haliday, D., Resnick, R., & Walker, J. (2005). Fundamentals of Physics (6th Ed.). New York: Wiley.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. Psychonomic Bulletin & Review, 7, 185-207

Iberall, A. S. (1972). Toward a general science of viable systems. New York: McGraw-Hill.

Kaplan, D., & Glass, L. (1995). Understanding nonlinear dynamics. New York: Springer-Verlag.

Kelso, J. A. S. (1995). Dynamic patterns: The self-organization of brain and behavior. Cambridge, MA: MIT Press.

Kugler, P. N. (1986). A morphological perspective on the origin and evolution of movement patterns. In M. G. Wade & H. T. A. Whiting (Eds.), Motor development in children: Aspects of coordination and control (pp. 459-525). Dordrecht: Martinus Nijhoff.

Kugler, P. N., & Turvey, M. T. (1987). Information, natural law, and the self-assembly of rhythmic movement: Theoretical and experimental investigations. Hillsdale, NJ: Erlbaum.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of movement control and co-ordination (pp. 5-78). New York: Wiley.

Lemke, J. L. (2002). Keeping learning alive: Multiple time scales in the social organization of learning. International Conference of the Learning Sciences.

Liu, Y-T., Mayer-Kress, G., & Newell, K. M. (2003). Beyond curve fitting to inferences about learning. Journal of Motor Behavior, 35, 197-207.

Luce, C. G. (1971). Biological rhythms in human & animal physiology. New York: Dover.

Mayer-Kress, G., Liu, Y-T., & Newell, K. M. (2006). Complex systems and human movement. Manuscript under review.

McGraw, M. B. (1943). The neuromuscular maturation of the human infant. New York: Columbia University Press.

Meinhardt, H. (1982). Models of biological pattern formation. New York: Academic Press.

Mittenthal, J. E., & Baskin, A. B. (Eds.). (1992). The principles of organization in organisms. Reading, MA: Addison-Wesley.

Muchisky, M., Gershkoff-Stowe, L., Cole, E., & Thelen, E. (1996). The epigenetic landscape revisited: A dynamic interpretation. In C. Rovee-Collier (Ed.), Advances in infancy research (Vol 10, pp 121-159). Norwood. NJ: Ablex.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed), Cognitive skills and their acquisition (pp. 1-55). Hillsdale:, NJ: Erlbaum.

Newell, K. M. (1986). Constraints on the development of coordination. In M. G. Wade & H. T. A. Whiting (Eds.), Motor skill acquisition in children: Aspects of coordination and control (pp. 341-360). Amsterdam: Martinies NIJHOS.

Newell, K. M., Liu, Y-T., & Mayer-Kress, G. (2001). Time scales in motor learning and development. Psychological Review, 108, 57-82.

Newell, K. M., Liu, Y-T., & Mayer-Kress, G. (2003). A dynamical systems interpretation of epigenetic landscapes for infant motor development. Infant Development and Behavior, 26, 449-472.

Newell, K. M., Liu, Y-T., & Mayer-Kress, G. (2005). Learning in the brain-computer interface: Insights about degrees of freedom and degeneracy in a landscape model of motor learning. Cognitive Processing, 6, 37-47.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. Trends in Cognitive Neuroscience, 5, 421-425.

Prigogine, I., & Nicolis, G. (1971). Biological order, structure and instabilities. Quarterly Reviews of Biophysics, 4, 107-148.

Schöner, G. (1989). Learning and recall in a dynamic theory of coordination patterns. Biological Cybernetics, 62, 39-54.

Schroeder, M. (1991). Fractals, chaos, power laws: Minutes from an infinite paradise. New York: Freeman.

Shirley, M. M. (1931). The first two years: A study of twenty-five babies (Vol. I) Locomotor Development. Minneapolis: The University of Minnesota Press.

Simms, K. (1994). Evolving 3d morphology and behavior by competition. Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems. 28-39.

Snoddy, G. S. (1926). Learning and stability. Journal of Applied Psychology, 10, 1-36.

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. Developmental Science, 6, 392-412

Strogatz, S. H. (1994). Nonlinear dynamics and chaos. New York: Addison-Wesley.

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

Thelen, E., Kelso, J.A. S., & Fogel, A. (1987). Self-organizing systems and infant motor development. Developmental Review, 7, 39-65.

Van der Maas, H. L. J. (1996). Beyond the metaphor? Cognitive Development, 10, 621-642.

van Geert, P. (1994). Dynamic systems of development: Change between complexity and chaos. New York: Harvester Wheatsheaf.

Von Bertalanffy, L. (1975). Perspectives on general system theory. New York: Braziller.

Waddington, C. H. (1957). The strategy of the genes. London: George Unwin & Unwin.

Ward, L. M. (2002). Dynamical cognitive science. Cambridge, MA: MIT Press.

Figure Captions

1. Waddington's (1957) schematic of the epigenetic landscape.

2. The multiple time scales of development from Muchisky et al. (1996).

3. A hypothetical error surface from a neural network (adapted from Elman et al., 1996, Figure 1.2 with permission).

4. Example of a meta stable fixed point at **(x*,y*)=(0,0)**. The eigenvalue related to the x-direction is –43.0, the one associated with the y direction is –6.3, indicating that the valley is steeper in x-direction than in y direction. That implies that here x would correspond to the fast timescale and y to the slow timescale.

5. Upper Figure. The attractor landscape dynamics of the development of infant prone progression (see text for details of simulation). Lower Figure. The figure shows the individual contributions of the attractors for each activity corresponding to the different behavioral attractors. If the contributions of these wells are superimposed, the landscape in the upper figure is the result. (Adapted with permission from Newell et al., 2003).

6. One-dimensional version of the development of prone progression (omit speed). Onset times are 1, 2,…,6 time-units. (Adapted with permission from Newell et al., 2003).

7. Figure 4: Simple landscape of Snoddy (1926) with two time-scales associated with the x (slow) and y (fast) dimensions of a learning process. The elliptical curves in the landscape correspond to contour lines with constant elevation. The black dots correspond to empirical data that appear to follow the gradient lines in the landscape, two of which have been extrapolated from the data. (Reproduced from Newell et al., 2005).

8a. Landscape of learning the 90 degree phase task of the HKB model: At the beginning of practice (c=0.4) only temporary stabilization of the target phase x_0=0.25 can be achieved when starting from special initial conditions close to (c). (Reproduced from Newell et al., 2001).

8b. Right at the transition (c=0.425) the target phase $x_0=0.25$ shows one-sided stability: Initial conditions close to (c) will be attracted to the new attractor. Note that in this situation the system is very sensitive to noise perturbations. (Reproduced from Newell et al., 2001).

8c. After sufficient practice (c=0.525), all initial conditions close to the target attractor $x_0=0.25$ will converge to the fixed point. (Reproduced from Newell et al., 2001).

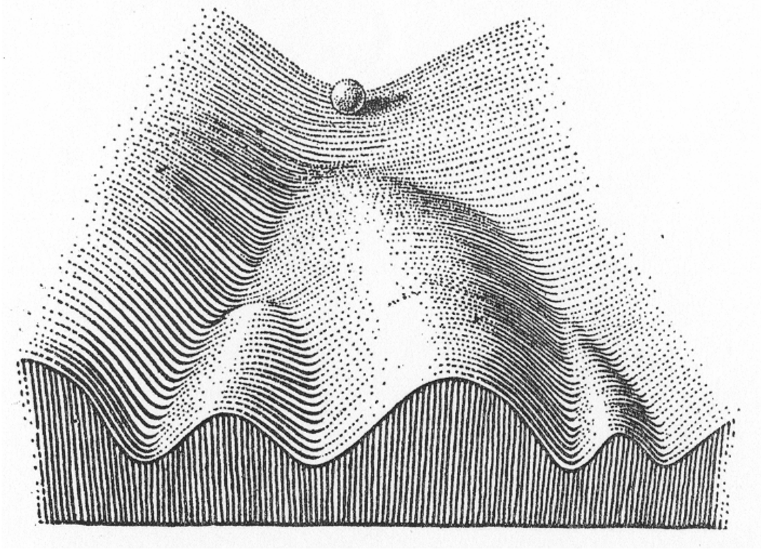


Figure 1

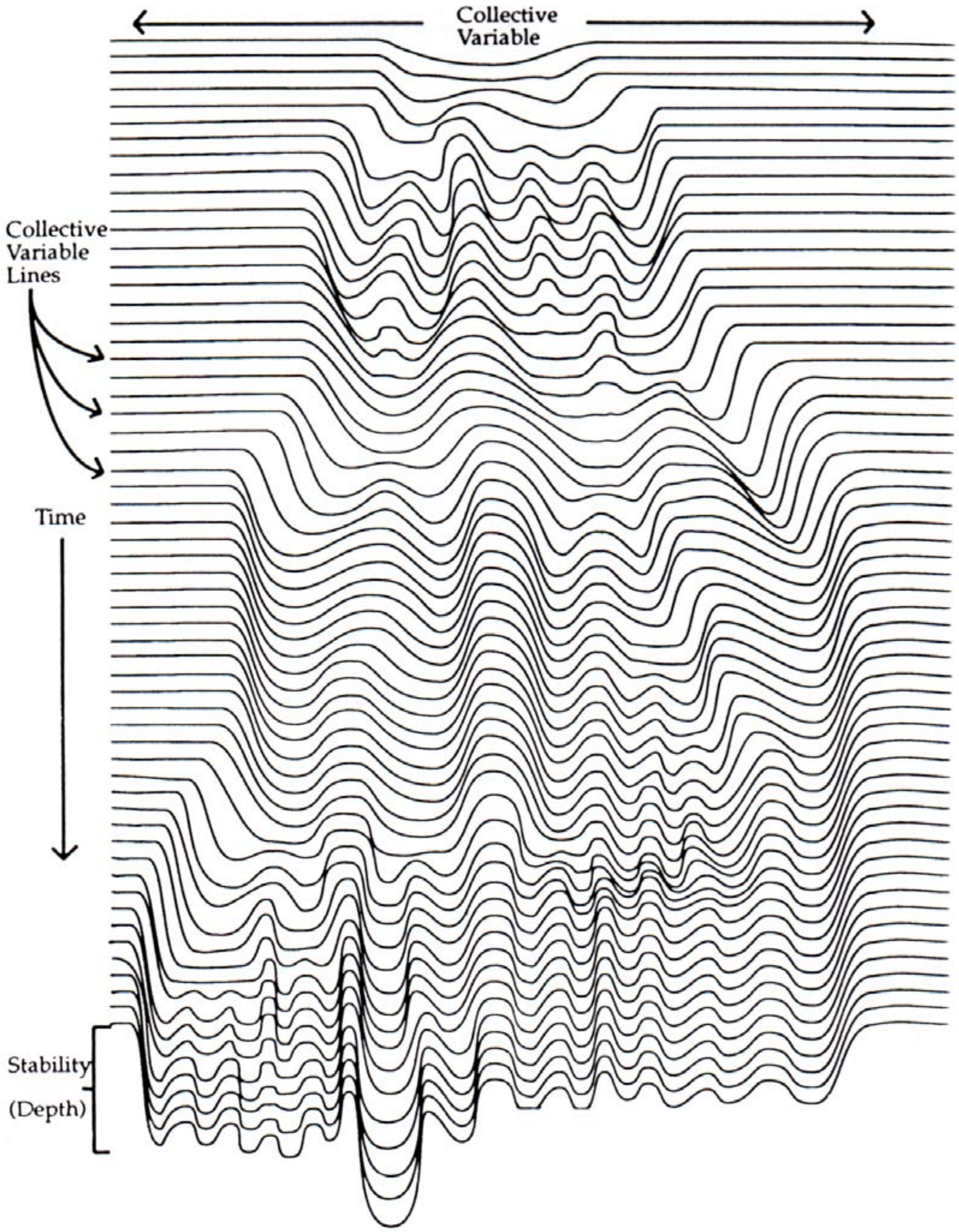

Collective Variable
Collective Variable Lines
Time
Stability
(Depth)

Figure 2

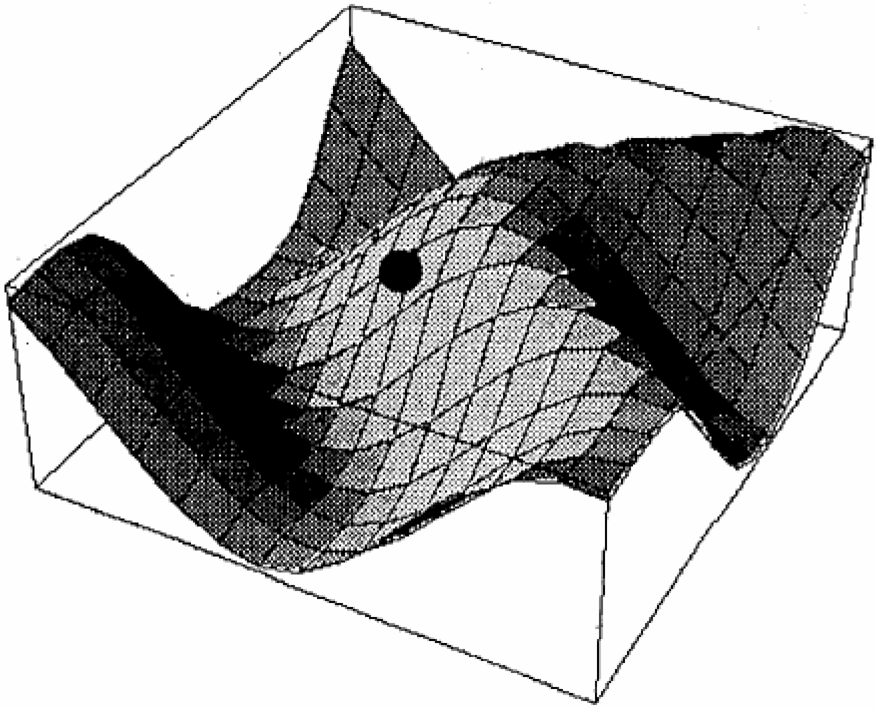


Figure 3

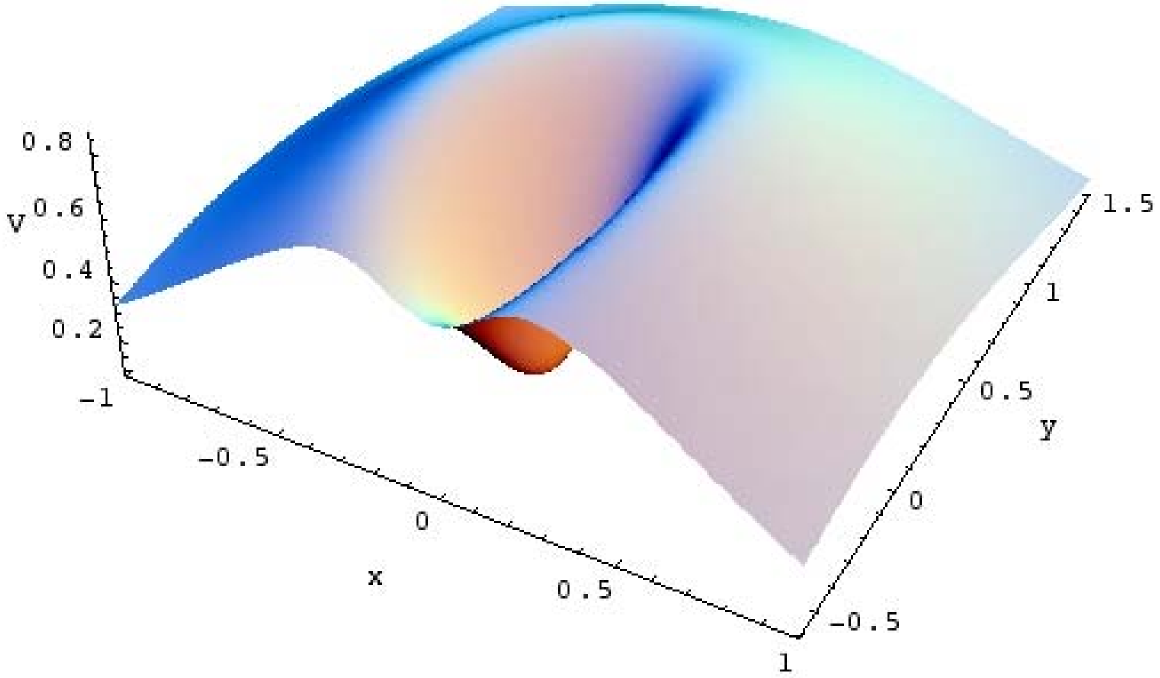


Figure 4

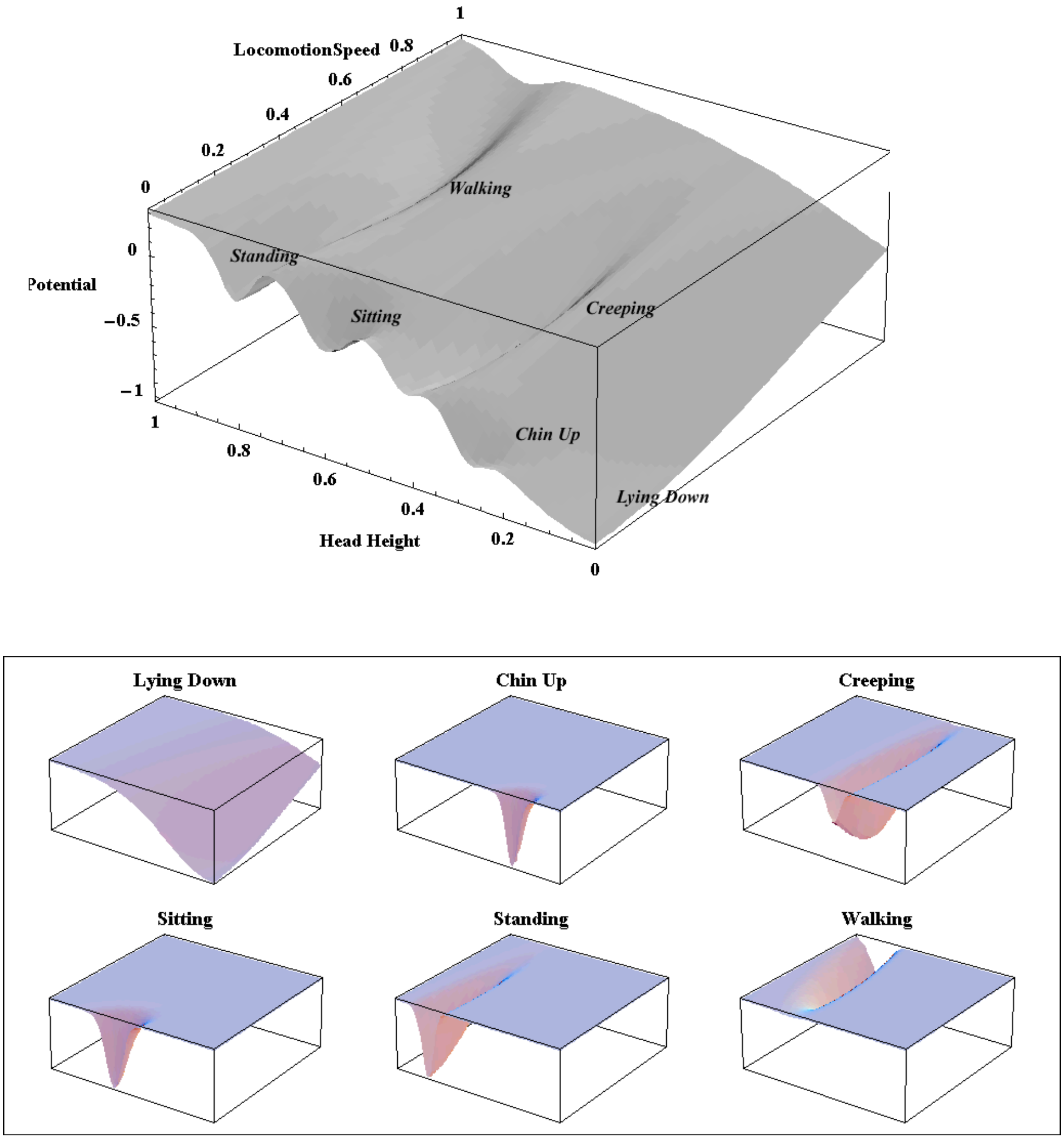

1
LocomotionSpeed
0.8
0.6
0.4
0.2
0
Walking
0
Standing
Potential
−0.5
Sitting
Creeping
−1
1
0.8
0.6
0.4
Head Height
0.2
0
Chin Up
Lying Down
Lying Down
Chin Up
Creeping
Sitting
Standing
Walking

Figure 5

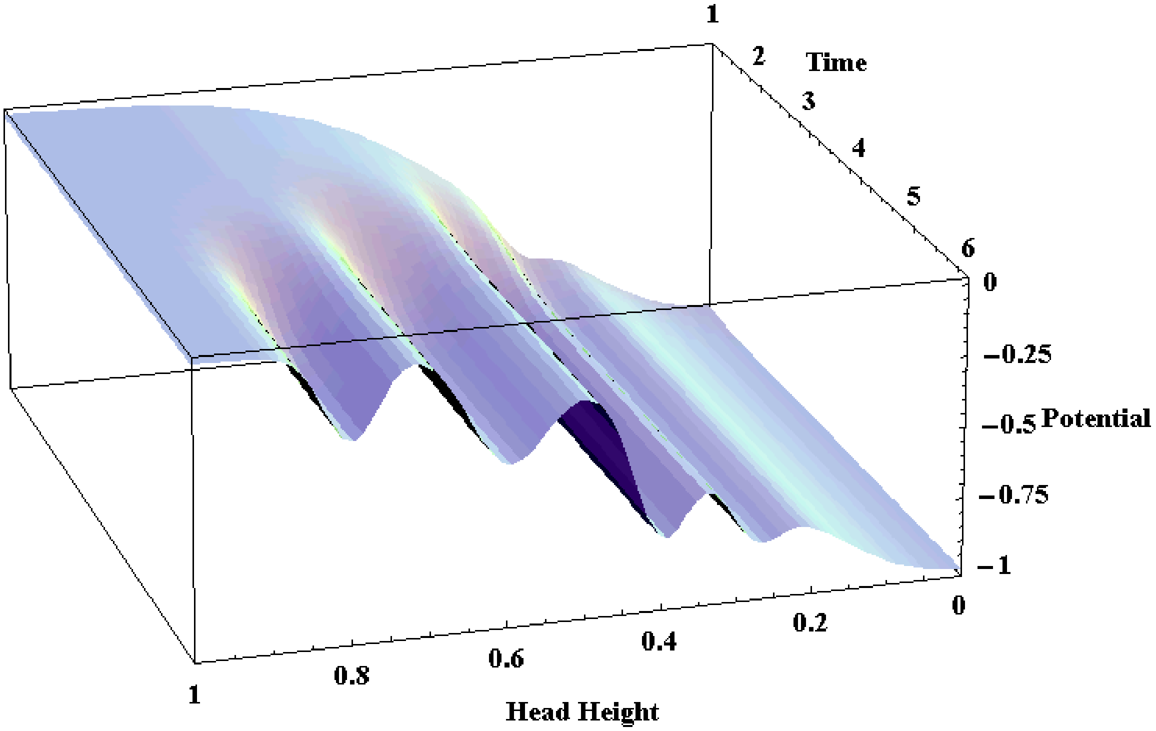


Figure 6

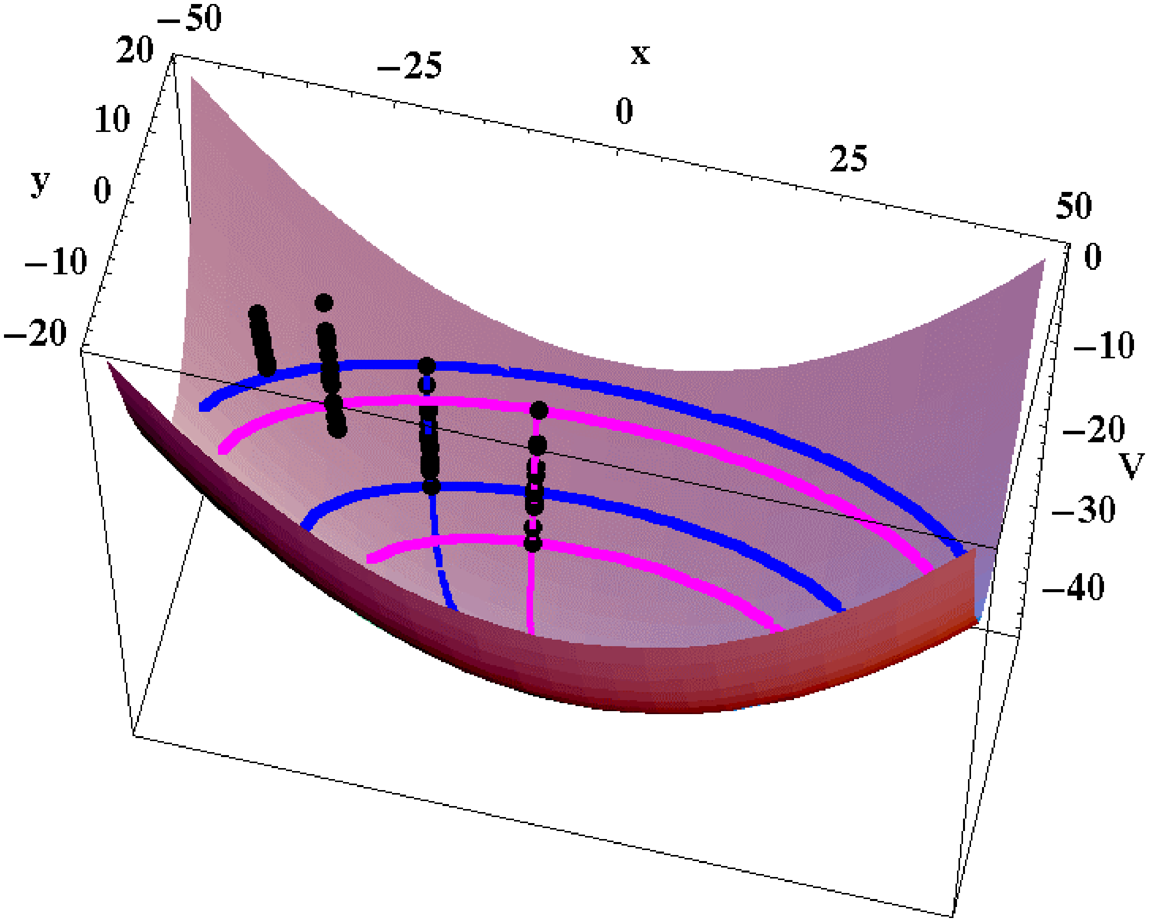


Figure 7

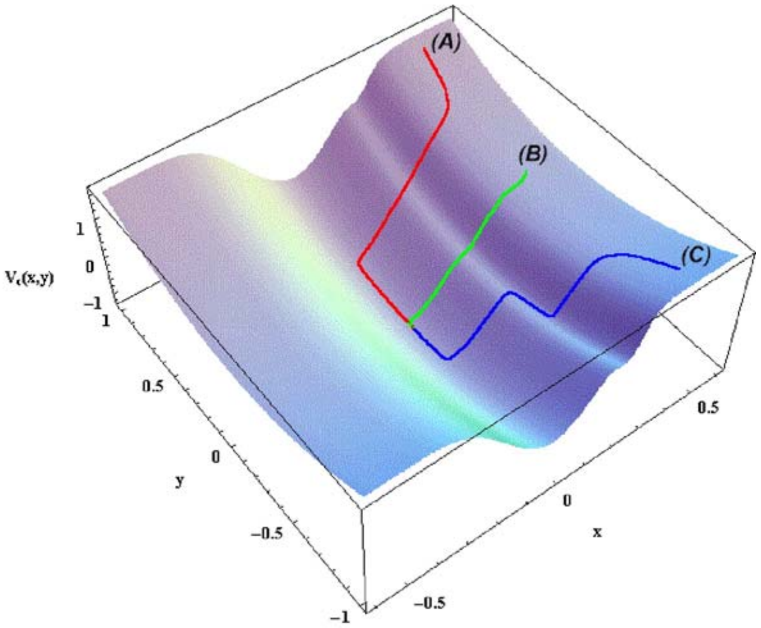


Figure 8a

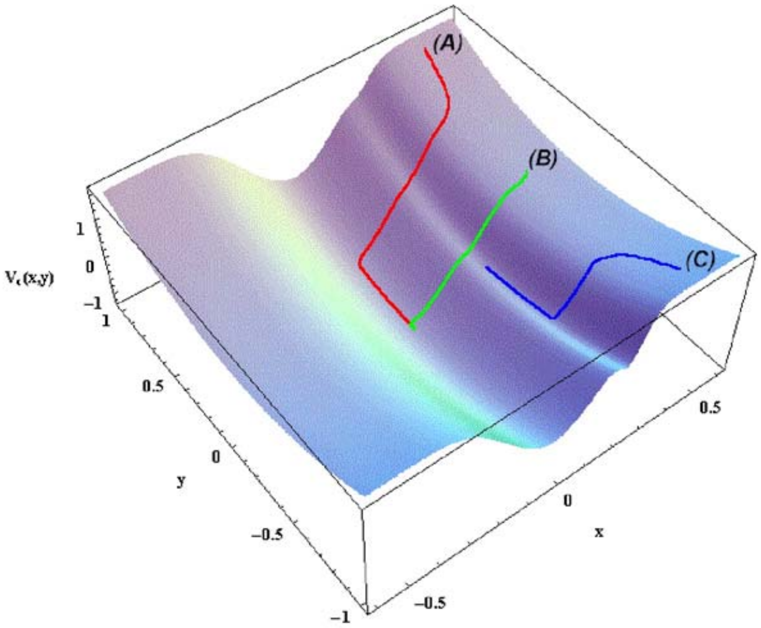


Figure 8b.

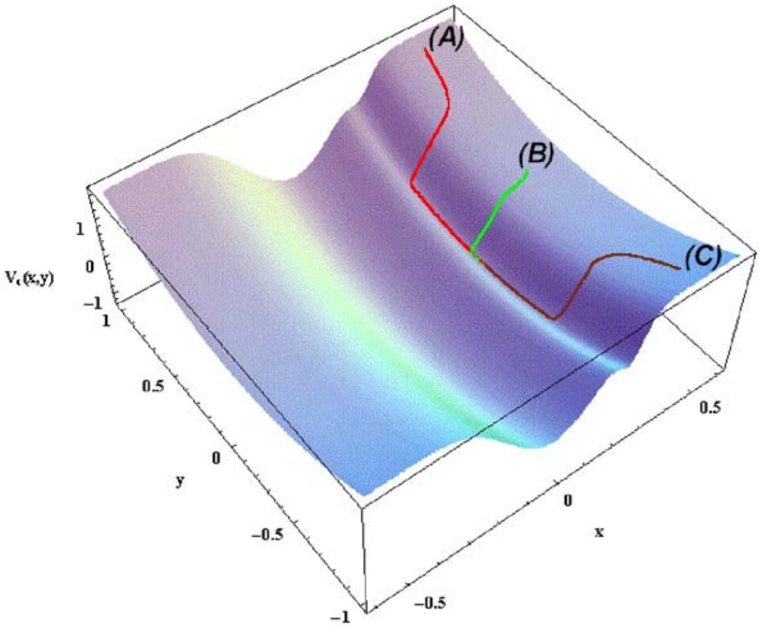


Figure 8c.

Dynamical Insight into Structure in Connectionist Models

Whitney Tabor
University of Connecticut

Please address correspondence to: Whitney Tabor
Department of Psychology, U-1020
University of Connecticut
Storrs, CT 06269

860 486 4910 (phone)
860 486 2760 (fax)

whitney.tabor@uconn.edu

Abstract

I discuss a connectionist model, based on Elman's (1990, 1991) Simple Recurrent Network, of the acquisition of complex syntactic structure. While not intended as a detailed model of the process children go through in acquiring natural languages, the model helps clarify concepts that may be useful for understanding the development of complex abilities. It provides evidence that connectionist learning can produce stage-wise development. It is consistent with prior work on connectionist models emphasizing their capability of computing in ways that are not possible within the symbolic paradigm (Siegelmann, 1999). On the other hand, it suggests that one mechanism of the symbolic paradigm (a pushdown automaton) may be identified with an attractor of the learning process, thus providing a useful bearing point in the complex landscape of dynamical computation. Thus, there is a menage-a-trois: a correspondence between dynamical systems theory and symbolic computation theory is embodied in a connectionist network.

Introduction

This paper adds its voice to those that have pointed out that connectionist models are a subclass of dynamical systems. They are a particularly interesting subclass because of their psychological plausibility. Moreover, they have a valuable trait: one can make connectionist models that behave in psychologically plausible ways but which are, at least

---

In addition to the very helpful interactions at the conference that prompted this book, I had enlightening discussions with James Dixon, James Magnuson, Jay Rueckl, and Dalia Terhesiu.

initially, inscrutable. Inscrutability might seem like a negative point. But one can sometimes learn things by striving to see into it. Dynamical systems theory is helpful in this regard: it provides conceptual constructs (e.g., trajectories, basins, attractors, bifurcations) and analytical methods (e.g., phase-space portraits, linearization) that can help one understand relationships and causes in these psychologically relevant models.

At this point, it is helpful to mention a third element: symbolic computation theory. By symbolic computation theory, I mean theories which describe behavior using discrete symbols and well-formedness rules. Observed behaviors are modeled as allowable symbol patterns under the well-formedness rules. The grammatical theories of natural language proposed by Generative Grammar are examples of this kind of theory (e.g., Chomsky, 1957). The symbolic approach has generally been treated as an enemy by both connectionists and ecological psychologists, including many dynamical-systems oriented ecological psychologists. While I agree with the rejection of symbolic models as a sufficient foundation for the science of mind-world systems, I argue here that symbolic models are nevertheless particularly relevant to cognition, and in fact, to the relationship between the connectionist and dynamical approaches. In a nutshell, symbolic models correspond to certain idealized behaviors which connectionist models can exhibit and which influence learning the way an attractor can be said to influence the trajectory of a dynamical system, even though realistic connectionist systems rarely, if ever, spend time at these attractors.

I'll make this more concrete by talking about one major domain: development. Several people have suggested that the dynamical notion of a phase transition is particularly relevant to development (Thelen & Smith, 1994; van der Maas & Molenaar, 1992; van Geert, 1998). In the sense intended here, a phase transition is a qualitative change in the behavior of a system that is associated with a monotonic and continuous change of a parameter, called a *control parameter*. Phase transitions seem potentially relevant to modeling developmental stages because they correspond to situations in which behavior changes subtly and quantitatively for a long period of time and then suddenly changes dramatically and qualitatively. Indeed, the researchers mentioned above have pointed out plausible resemblances between developmental data and various mathematical models of phase transitions. Here, I go along this path and then make an additional suggestion: developmental phase transitions may involve a shift from a phase in which well-formedness rules fail to characterize the structure of behavior to a phase in which these rules constitute a good approximation of the behavior. Thus, the account treats the symbolic models as relevant to characterizing stages, but not sufficient for characterizing all the states people go through in the process of growing up.

What is the evidence for these claims? It takes the form of a connectionist model called a Fractal Learning Neural Network (FLNN). I want to make clear at the outset that I offer no measured developmental data to motivate this model at all. Instead, one might think of this work as an exploration of abstract information-structure. For two reasons, I believe that the results may be of interest to developmentalists: 1. The model is a connectionist learning model and thus has relatives which have made plausible empirical predictions about many domains, including development; 2. The model helps a person think clearly about an appealing alternative to some current theories of development (e.g., maturational explanations for stages, superficial nativist accounts). The model's account is appealing because (a) it uses a simple learning mechanism and thus corresponds to a theory that links development and learning, (b) it predicts stage-wise behavior with progressive quantitative changes occurring within the stages, a phenomenon also observed in development, (c) it derives the phase transitions from the learning, thus predicting, rather than stipulating, the stages, (d) it shows highly structured, domain-specific behavior arising from a mechanism which is capable of a wide variety of behaviors, not all of which are highly structured, thus, in some sense, deriving the structure, (e) it generalizes beyond its input in a way that is at least partially consistent with human generalization.

I will describe the model and then return to these points.

Description of the Model

Background

Elman (1990, 1991) trained a network, which he dubbed the "Simple Recurrent Network", to process sequences of symbols, including sequences of words. The network (Figure 1) has three layers of units (input, hidden, and output) as well as a context layer, which stores a copy of the hidden unit activations from the previous time step. The connections feed forward from input to hidden to output and from context to hidden. The copying operation associated with the context layer is a kind of recurrence, and it permits the network to encode information about past inputs, allowing it to keep track of the temporal dependencies which are crucial to syntactic processing.

In one experiment, Elman trained his network on a corpus of words approximating the syntactic properties of English. The corpus included singular and plural nouns, as well as center-embedded relative clause structures (1) (articles like "the" and *a* were left out for simplicity).

(1) dogs/dog who girls chase run/runs.

In an example like (1), the network must store the number of the head noun of the subject of the sentence ("dog" or "dogs") during the time that it processes the relative clause ("who girls chase") in order to properly predict the number of the main verb ("run" vs. "runs"). Elman used a single-on-bit ("one-hot") encoding for each word on the input layer and on the output layer. The codes of the words of the corpus were presented in order, one at a time, on the input layer of the network. The network was trained (using the sum-squared error cost function and backpropagation of the error signal through time Rumelhart, Hinton, & Williams, 1986; Pearlmutter & MacDonald, 1995) on the task of predicting the next word at each point in time. Because the sequence of sentences in the corpus was approximately random, the network was not able to guess specifically what word was going to appear at any point in time. However, it could pick up on stable syntactic properties of the language. In particular, once the network was fully trained, the activation of each output unit approximated the probability that the word corresponding to that output unit would occur next in the corpus. For example, after having encountered the word "dog" at the beginning of a sentence, output activation was concentrated mostly on the singular verbs and the word "who". While such local co-occurrence tendencies are fairly easy to model (e.g., with Markov processes), some kinds of longer distance dependencies, including those arising in center-embedded structures like (1), are more difficult. Thus, it is significant that the network was also successful at predicting the number on the verbs in relative clause structures like (1).

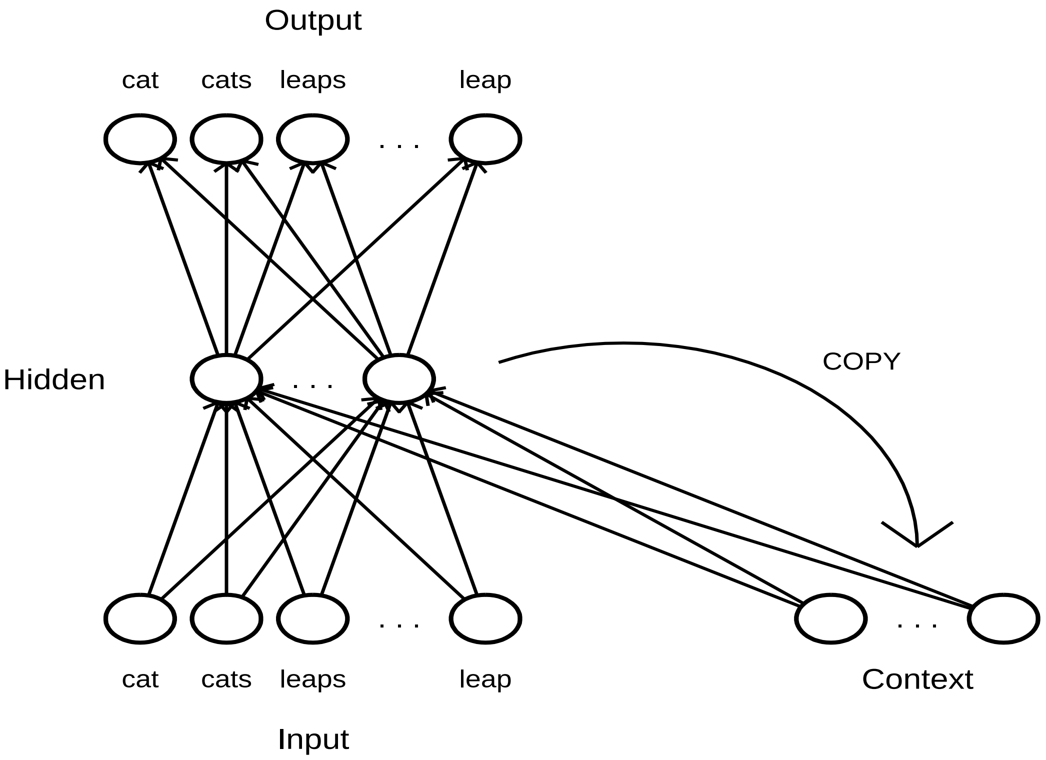


Figure 1. Simple Recurrent Network (Elman, 1990, 1991).

Since Elman's initial experiments, several projects have successfully used the Simple Recurrent Network and variants on it to learn complex grammatical dependencies in other human-language-like corpora (e.g., Christiansen & Chater, 1999; Rohde & Plaut, 1999; Rohde, 2002). Several other projects investigated the learning of abstract formal languages, related to human languages, in order to better understand the principles by which the network was succeeding at prediction (Bodén & Wiles, 2002; Rodriguez & Wiles, 1998; Rodriguez, Wiles, & Elman, 1999; Rodriguez, 2001; Wiles & Elman, 1995). Gers, Schmidhuber, and Cummins (2000); Gers and Schmidhuber (2001); Hochreiter and Schmidhuber (1997) invented a different, but related kind of recurrent network called "Long Short Term Memory" (LSTM) which could keep track of very long temporal dependencies (up to 1000s of intervening events).

The formal projects just mentioned focus on the learning of context-free languages, which provide the foundation for symbolic linguistic theories of language. Although the networks were successful on one simple case (the language, $a^n b^n$)[1] and the analyses provided clear insight into the method by which the networks were succeeding (see further discussion below), the networks were much less systematically successful with more complex cases

---

[1] $a^n b^n$ is the language consisting of all strings of a's and b's in which some number of a's is followed by the same number of b's: e.g., ab, aabb, aaabbb, etc.

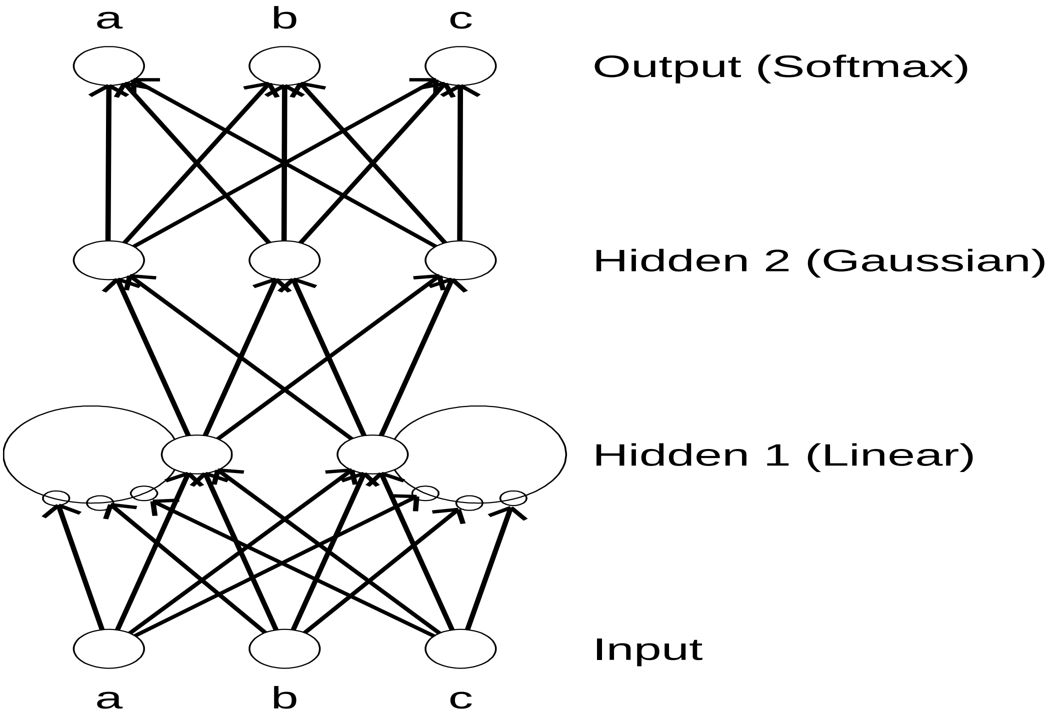


Figure 2. A Fractal Learning Neural Network (FLNN).

(e.g., palindrome languages), which more closely resemble human languages. Meanwhile, work on the encoding of complex symbolic structures in connectionist units independent of the learning question (Moore, 1998; Tabor, 2000) has provided some insight into how it is possible to encode complex syntax in connectionist architectures.

The construction of the Fractal Learning Neural Network (Tabor, 2003) was motivated by the combination of these research strands. It combines the representational results of Moore and Tabor with a hill-climbing algorithm that discovers weight settings embodying complex symbolic processes. I'll describe the architecture of this device, along with some simulation results, and then make some comments on light that this effort may shed on development and on the relationship between dynamical systems and connectionist approaches.

Architecture.

The FLNN architecture has four layers (Figure 2). The first layer is the input layer, where each unit stands for a word. The second layer has a linear activation function with linear recurrent connections that are gated by the input units. The third layer has Gaussian (radial basis function) units whose means are learned but whose variances are fixed (arbitrarily) at 0.25. The fourth layer is the output layer. In the output layer, as in the

Table 1: *ABC Grammar.* Parentheses denote optional constituents, which occur with probability 0.2 in every case. A probability, indicated by a decimal number to its left, is associated with each production.

1.0 S → A B C 1.0 A → a (S) 1.0 B → b (S) 1.0 C → c (S)

input layer, each unit stands for a word. The output units as a group have the normalized exponential (softmax) activation function because they specify probability distributions over words.

There is an important question about the number of units in the hidden layers. I chose the number of units to be the minimal number that I knew were necessary to solve the task well (in this case 2 for the recurrent hidden layer, and 3 for the Gaussian layer). This constraint may play a role in the success of generalization touted below. Further testing is needed to find out whether the network generalizes as well when it has an excess of hidden units.

Training.

I trained the network on simple formal languages that are more complex than those studied in prior work that has been successful at fully interpreting network symbol processors' hidden unit representations (Gers et al., 2000; Gers & Schmidhuber, 2001; Rodriguez & Wiles, 1998; Rodriguez et al., 1999; Rodriguez, 2001; Wiles & Elman, 1995). On the other hand, they are not as complex as the corpora used by researchers focused on the details of the resemblance of the network's behavior to natural language (Elman, 1990, 1991; Christiansen & Chater, 1999; Rohde & Plaut, 1999; Rohde, 2002). See Tabor (2003) for a discussion of this complexity issue.

The training data for the network were generated from the grammar shown in Table 1. In the simulation reported here, the training corpus consisted of one instance of each sentence from the grammar that has 9 or fewer words. Thus the maximum number of concentric phrases was 3 (i.e., the maximum number of center-embeddings was 2).

At the beginning of each sentence, the two recurrent hidden unit activations were set to 0. The words were presented one at a time, in order, on the input layer of the network. The grammar was used to determine output probabilities following each word.

The Gaussian Unit means and the Gaussian → Output weights were set to random initial values with mean 0 and variance 0.3. The gated self-connections in the hidden layer were set to initial value 1 (this is the unbiased choice (Tabor, 2000)). The recurrent connections between different units (as opposed to the recurrent connections from each hidden unit to itself) were clamped to 0 throughout training and testing. This choice also was motivated by the knowledge that these connections would not be needed for the task at hand.

The network was trained by hill-climbing with epoch-sized steps. The first step of hill-climbing was to present every sentence in the corpus to the network and compute the total error at the output layer. Kullback-Leibler Divergence was used as the error measure. Then, a sample of alternative weight settings lying on a sphere around the current weight setting was chosen. For each of these, the entire corpus was tested and error was again accumulated.

Table 2: Performance on training and test sets. RMSE = Root Mean Squared Error. % Cor. = Percent Correct. N = the number of networks that contributed to the computation of Standard Error (SE). Npoints = the number of words tested per network. All networks were trained by hill climbing.

Corpus	RMSE	SE	% Cor.	SE	N	Npoints
Training	0.013	0.001	100.000	0.000	9	129
Test	0.048	0.005	99.424	0.195	9	4755

If one of these made a maximal improvement in the error, than it was adopted as the new current weight setting. (Ties for best were broken arbitrarily). This process was repeated until total error dropped below a pre-specified threshold (0.02).

Test Results.

As noted, the network was trained on all sentences of length up to 9 (at most 3 concentric phrases). The network was tested on all sentences of length 10 to 15 (at most 5 concentric phrases). The results are shown in Figure 2.

Table 2 indicates that while the network's performance on the test corpus was not perfect, it was very nearly so. Tabor (2005) provides a careful analysis of both the principle underlying the network's success and the sources of its errors.

Analysis of the weights.

A fractal (or "fractal set") is a set that exhibits self-similarity over an infinite range of scales (scale-invariance), including arbitrarily small ones. The FLNN keeps track of complex linguistic dependencies by shrinking a vector (whose direction and length constitute information) whenever the information needs to be stored, and adding it on to the vector encoding new information at the original scale. When previously encoded information needs to be recalled, the component at the maximal scale is subtracted out and the remainder is expanded to make the information in it readily discernable. To handle arbitrary depths of recursion, the network contracts vectors to arbitrarily small sizes. Thus it performs its task by jumping around on a fractal.

Code.

Matlab code for running these simulations is available at
http://www.sp.uconn.edu/~ps300vc/Projects/FractalLearning/index.html
This code runs slowly but is fairly simple to interpret.

Connectionist networks and dynamical systems theory

Wiles and Elman (1995), Rodriguez and Wiles (1998), Rodriguez et al. (1999), and Rodriguez (2001) analyzed Simple Recurrent Networks trained on simple, context-free languages, related to the case discussed above. They showed that standard dynamical systems constructs (trajectories, attractors, repellers) provided a foundation for understanding how the network successfully generalized from the finite sample it was trained on. For example, Rodriguez et al. (1999) showed that a network trained on the language, a^nb^n, employed two quasi-linear dynamical objects: an attractor and a repeller. When the code for a was presented on the input layer, the network's hidden unit dynamics consisted of a single attractor basin with a stable fixed point. At the beginning of processing , the hidden unit activation vector was relatively far from the attractor. Each successive presentation of an a caused the state to jump a step closer to the fixed point. When the input was switched to b, the result was a new dynamical regime, similar to the a regime, but with a repeller that was located on the other side of hidden unit space. The first presentation of b moved the state to a location close to the repeller. The distance between this point and the repeller was a decreasing function of the number of a's that had been presented. Additional presentations of b caused the network to jump away from the repeller, and head back to the location associated with the beginning of a sentence. The hidden state arrived back at this location when the number of b's matched the number of a's that had been presented. The Hidden → Output weights of the network implemented linear separations of the hidden unit space so that when the network was near the periphery of the attractor's basin, it expected the start of a new sentence (only possible next word = a), when it was jumping toward the attractor, it predicted a and b, and when it was in the region of the repeller, it predicted only bs until it arrived back at the start-of-sentence region. Thus, the balancing of the number of a's with the number of b's was accomplished by matching the contraction effect of the attractor with the expanding effect of the repeller.

Rodriguez et al.'s network moves, in essence, on a simple fractal, the geometric series, to keep track of levels of embeddings. At the end of training, the FLNN described above moves on a slightly more complex fractal, a two-dimensional Cantor Set, to simultaneously track levels and types of embedding. As in Rodriguez et al.'s case, the system is driven by a collection of attracting and repelling fixed points with carefully coordinated interactions.

Thus, if these cases are typical, they suggest the following relationship between connectionist models, dynamical systems theory, and symbolic processes: certain connectionist networks with certain weight settings mimic symbolic processes in symbol sequencing behavior. Moreover, complex dynamical systems objects (e.g., a coordinated attractor and repeller, a fractal) can be systematically related to the symbolic entities. This suggests that dynamical systems theory is in a position to provide insight into the relationship between symbolic and connectionist models. The next section explores the process by which the network approaches such a symbolically organized state, and provides some evidence that the relationship proposed here is not a fluke, enforced by excessive training, but reflects a fundamental relationship that may extend to other cases. The next section also suggests that not all states of the network's development correspond to symbolically describable behaviors.

Phase Portraits and Review of Claims

Figure 3 is a succession of phase portraits[2] of the model over the course of training on the ABC language. The portraits show the points visited by the model when it processes sentences containing up to 5 levels of embedding. Only points visited during the processing of initial strings containing a's and b's are shown. If strings containing c's were shown, the pictures would look similar, but fuzzier, because the model imperfectly approximates the symmetry between contraction and expansion.

---

[2] Here, by "phase portrait", I mean a diagram of the state space at one stage of training that gives enough information that, with knowledge of the system at hand, one can deduce the dynamics at that stage.

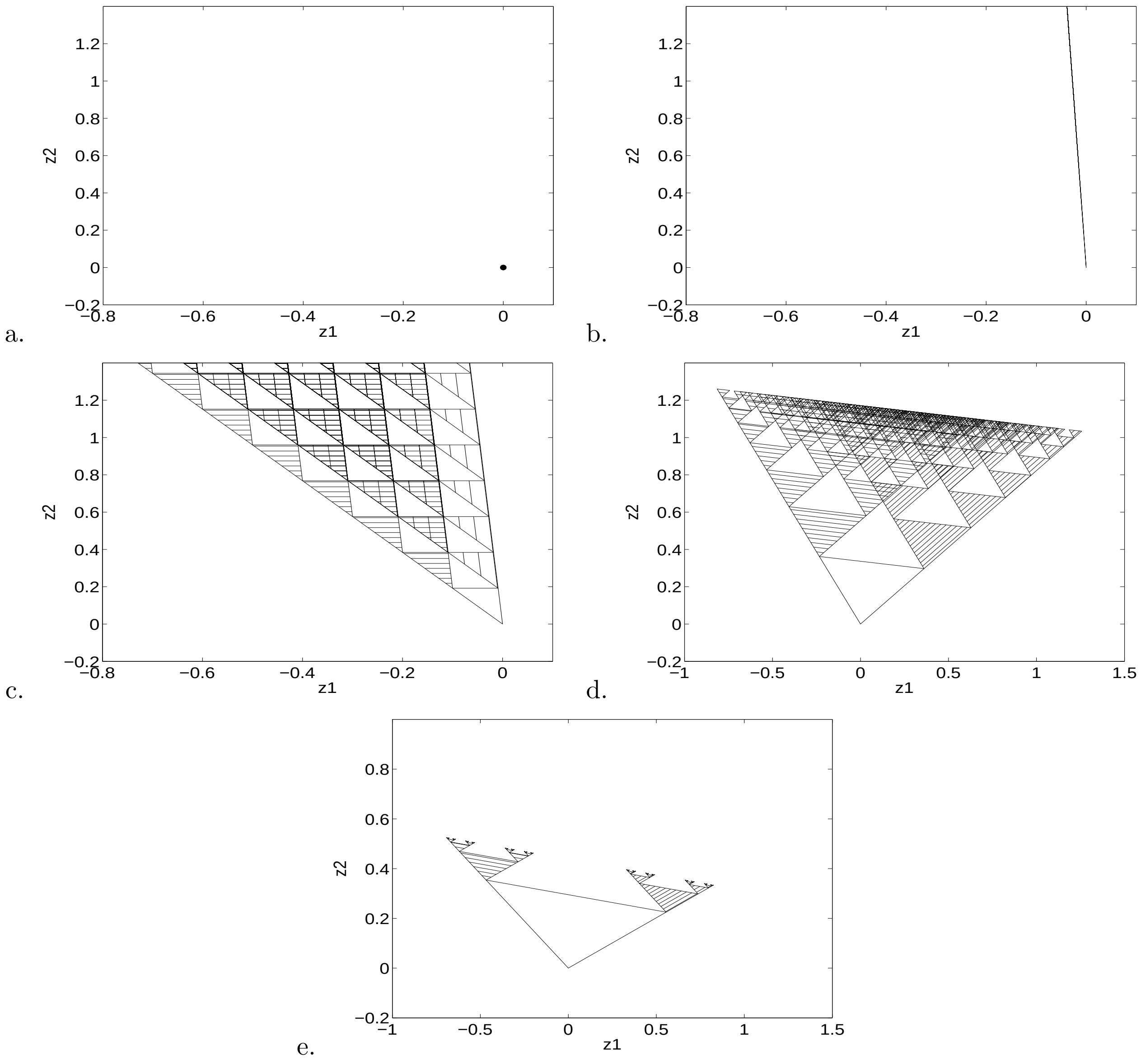


Figure 3. The evolution of the set of visited states in the linear hidden unit layer during processing of grammatical strings over the course of training an FLNN on the language of the grammar in Figure 1. Only states involving sequences of a's and b's are shown—i.e., only after sequences of push and control-state changes made by a pushdown automaton designed to process this language. (a) At the beginning of training, the visited states consist of the single point, [0, 0]. (b) At 550 epochs, the visited states lie on an infinite ray with endpoint at [0, 0]. (c) At 1000 epochs, the visited states lie on an infinite lattice (the visited states are at the vertices of the triangles). There is near-complete overlap between states in which a *b* (or *a*) is predicted next (horizontal shading) and states in which a *c* (or *a*) is predicted next (vertical shading). (d) At 10000 epochs, the lattice has become finite and the branches corresponding to *b*-prediction and *c*-prediction are spreading apart. (e) At 51174 epochs (the end of training), the branches have fully separated and the set of visited states lie on a totally disconnected fractal in the sense of Barnsley (1993[1988]). At this stage, the network's behavior is nearly identical to that of the corresponding pushdown automaton. It is not exactly identical because the expansion associated with the c inputs only approximately inverts the contraction associated with the a inputs.

Figure 3 illustrates several successive, qualitatively distinct stages of the model's behavior over the course of training. With this illustration in view, I review the claims about "appealing properties" made above and assess their validity:

(a) *The model learns.* The model clearly involves a tuning process which brings it from domain-oblivious behavior to highly domain-sensitive behavior. Hill-climbing itself is not biologically plausible but its success is a prerequisite for the success of currently known biologically plausible algorithms (e.g., Hebbian learning).

(b) *The model passes through qualitatively distinct stages.* Distinct stages can be observed in Figure 3. The dimensional expansions that have occurred between 3a and 3b and between 3b and 3c, along with the "unfolding" that completes itself between 3d and 3e appear to be associated with bifurcations, the mathematical analogues of phase transitions.[3] These bifurcations are associated with changes in the symbolic computational behavior of the system. This apparent correspondence between bifurcations and grammar-changes is one foundation of the claim that there is a link between the dynamical and symbolic computational perspectives—see also Crutchfield and Young (1990); Moore (1998).

The model exhibits "progressive quantitative change occurring during the stages". For example, during the "unfolding" stage illustrated in 3c and 3d, the depth at which the branches first overlap is increasing. Thus, the depth of embedding that the model can successfully parse is gradually increasing. This is a "progressive quantitative change" because it makes the model's behavior more like the fractal parser it eventually becomes, even though it has not yet crossed the threshold (between 3d and 3e) of becoming isomorphic to a pushdown automaton.

What is meant by the claim that the model transits from a stage at which well-formedness rules are inadequate for describing its behavior to a stage at which they provide a good approximation? In 3c, the region spanned by the model during the processing of all sentences of the ABC language has infinite area. This is because contraction has not developed yet and the inputs are causing incremental, equal-sized displacements. But the interpretation of the model as an approximation of a pushdown automaton depends on its use of the fixed-width gaussian units in the second hidden layer to classify each grammatical condition it encounters. When the model does not reliably land within the "spotlights" of these gaussians, the basis of the isomorphism with a symbolic device is lost. Viewing the case from the perspective of self-organization, one might say that the model has not yet reached the stage where the constituents of the emergent structure are successfully coordinating.

(c) *The stages derive from the learning.* The qualitatively distinct stages come about through the iterative application of a single, simple process: move to the best proximal weight setting. In this sense, the stages are derived from the learning. In this regard, the current model offers an advance over Catastrophe Theory models (Thom, 1985) which identify traits that phase transitions can be expected to have but do not provide a great deal of insight into what causes the transitions.

On the other hand, it is not clear that the bifurcations in the current model exhibit the full range of properties that Catastrophe Theory predicts (e.g., modality, sudden jumps,

---

[3] I have not yet formalized this claim, so it remains uncertain. I am assuming a standard interpretation of bifurcation: a bifurcation has occurred if there does not exist a homeomorphism between the before and after systems. Two systems are homeomorphic if one can be transformed into the other by stretching and squeezing without tearing.

hysteresis, divergence, critical slowing down, etc.) and for which there is compelling evidence in the developmental literature (van der Maas & Molenaar, 1992; van Geert, 1998). If this discrepancy is present, it may stem, in part, from the fact that the FLNN does not model real-time processing (its smallest grain is what might be called "sentence-time processing"). In any case, further assessment of the relationship between Catastrophe Theory and the current, fractal learning approach may be fruitful.

(d) *Complex, domain-specific behavior from a general mechanism.* Expert evidence from currently practicing linguists on the types of grammatical patterns present in natural languages finds all of them to be within the symbolic paradigm. I'm skeptical that this sample reflects the true distribution. My skepticism stems from the fact that, until recently, people have not known how to look for anything else. Even within the symbolic paradigm, there is currently no satisfactory model of how a specific grammar is chosen from among the immense number of possibilities the paradigm affords. For these reasons, again recognizing that further empirical justification is needed, I treat it as desirable to adopt a model with very general computational capabilities and to let learning do significant work. This is in contrast to established linguistic treatments of native capacity (e.g., (Gibson & Wexler, 1994)) which assume that the complex machinery is pre-fabricated by the genes, and only finitely-many discrete parameter-settings must be selected in order to achieve adult behavior.

That connectionist learning is a very general induction system is well attested. Here, we appear to see it building a fairly complex kind of mechanism, a *pushdown automaton*, to suit the needs of a particular task. A pushdown automaton is a device that stores a sequence of symbols in memory and then recovers them in reverse order, retrieving each at just the moment when it needs to be recalled. The question that needs to be asked is, To what extent was this result contrived by the constraints put on the model? If the observed pushdown stack is the only complex behavior the model is capable of, or if it is one of a small, finite number of possibilities, then the fact that the learning algorithm discovered it is not very revealing—under these conditions, the treatment would be very similar to current nativist accounts. On the other hand, if many behaviors are possible, then the model offers an appealing alternative to current versions of nativism in language learning theory. I know this much at present: close relatives of this model are capable of a great variety of behaviors, including all context free languages (Tabor, 2000), a crossed serial dependency language (Tabor, 1991), several finite-state languages (Tabor & Terhesiu, 1994), and several chaotic processes (Tabor & Terhesiu, 1994). I do not yet know how many of these behaviors can be learned.

(e) *Generalization.* When trained on shallow embeddings (e.g., down to depth 2) the model generalizes to greater depths (3+). In fact, its end-of-training behavior closely mirrors the behavior of the infinite-state pushdown automaton that inspired the training data. The fact that the neural model generalizes to this well-known and very useful (e.g., in programming languages and plant growth—Lindenmayer & Prusinkiewicz, 1989) type of mechanism on the basis of a minute number of examples (4 sentences in the depth 2 corpus for the ABC language) is surprising. Further tests show that when the model is trained on various other small subsets of the infinite state ABC language, it converges on the same, infinite-state behavior. This suggests that the infinite state behavior is a kind of attractor for the learning process. These observations constitute another piece of evidence in support of the view that there is a systematic relationship between the symbolic and dynamical perspectives.[4]

But how relevant are these infinite state languages to real, human behavior? This is a point of long-standing controversy. Chomsky (1957) and many other people have argued that we ought to be studying ideal language ability—called "competence"—which is said to include the ability to process embeddings of arbitrary depth. That people cannot manage more than 3 levels of center-embedding easily (e.g., [The essay [that the guinea pig [someone [I knew $_4$] cared for $_3$] shredded $_2$] was about Duchamp $_1$].) is taken to be a performance-limitation: it is supposed that although the mechanism involves a pushdown automaton, the memory buffer for the symbol stack is limited, so we run out of space after only a few embeddings (Kimball, 1973). Others have argued that the distinction between competence and performance is dubious and that connectionist networks like Elman's Simple Recurrent Network (SRN) predict the limited ability of people to process center embedded sentences without needing to adopt the full power of a pushdown automaton (Servan-Schreiber, Cleeremans, & McClelland, 1991) and without needing to make the apparently untestable distinction between competence and performance (Christiansen & Chater, 1999). The makers of the latter arguments offer connectionist models as examples of systems in which there is no distinction between competence and performance. The present results, however, suggest a different interpretation of at least some connectionist models: an FLNN only exhibits perfect recursive computation if there is perfect balance between contraction and expansion, and if weights and activation are computed with infinite precision. When any noise is added to the values, the system is computationally equivalent to a finite-state machine (Casey, 1996). This observation suggests a close parallel between Turing machines and FLNNs: the infinite precision is playing the role of the infinite tape (Pollack, 1987). Moreover, the possible attractor status of recursive weight settings suggests that even though real systems only approximate these settings, they are relevant theoretical constructs which can help us understand the principles by which the device is succeeding at the task. Put this way, the fractal structure is very much like an element of "competence" in the sense of Chomsky (1957). This line of reasoning casts some doubt on the claim that connectionist models provide a basis for rejection of the competence-performance distinction.

However, the argument does not stop here. It should be noted that there are importance differences between the symbolic and FLNN realizations of the competence/performance idea. First, in order to predict that center-embeddings will be hard to process beyond three or four levels, the symbolic approach must bound the length of a stack to three or four elements, a choice which seems unmotivated. By contrast, it is very plausible that noise would be present in the neural implementation of a fractal computer, and it is also attested that the growth of difficulty in center-embedding has a roughly exponential structure, as predicted by the fractal scaling (Miller & Isard, 1964). Second, the treatment of recursion as an emergent property rather than an architectural property gives the system a kind of flexibility that seems suited to generating the variety of recursive structures that natural languages exhibit (e.g., center-embedding vs. crossed-serial dependency Savitch, 1987) and the mutability, over periods of decades and centuries, of grammatical structures (Hopper & Traugott, 1993). Third, the treatment of ideal devices as attractors rather than

---

[4] Thanks to Denis Mareschal for pointing out the significance of these results.

mechanisms raises the possibility that multiple ideals could simultaneously shape behavior, a prediction born out by some psycholinguistic work (Tabor, Galantucci, & Richardson, 2004) and possibly relevant to cases of creole continua and language contact. Fourth, while the symbolic paradigm separates the ideal system from the perturbations of it caused by "performance factors" in a stark way, by defining the system as signal and the perturbations as noise, connectionist models claim that there is a continuum between the two. The latter position may be relevant for understanding how new structures arise in language change (Givon, 1971; Tabor, 1994, 1995) and for understanding why grammatical patterns in some languages parallel statistical patterns in others (Bresnan & Nikitina, 2003). In these regards, the FLNN properties are consistent with the claim put forth by (Servan-Schreiber et al., 1991) that recurrent connectionist networks are a new kind of computational device that may naturally be called a "graded state machines". A fractal is aptly described as a "graded recursive structure".

Conclusions

The original motivation for developing FLNNs was to find a way for connectionist networks to effectively learn complex syntactic dependencies and to discern how connectionist treatments differ from symbolic treatments of these behaviors. It is important to clarify how much of this task has been accomplished and how much not, and to draw attention to other elements of language, not treated by this model, which may play an important role in the developmental picture.

First, it is known that connectionist models with these general characteristics (symbol predictors with recurrent hidden layers employing second order weights) are capable of correctly processing all context free languages (Moore, 1998; Tabor, 2000). I've only tested learning on two such languages, so it remains an open question how many of these parameterizations are reachable via learning.

Second, it is important to note that human language knowledge contains a good deal of systematic structure beyond that of phrase structure. One major element of this "other" structure goes by the name "pragmatic" (or "semantic") knowledge. Some have argued that pragmatic knowledge plays a central role in bootstrapping syntactic understanding (e.g., Pinker, n.d.). Others have argued for the reverse causal relationship (e.g., C Fisher, 1991). I am disinclined to believe that the causality is serial in either direction, but a mutual causality view also attributes a significant role to pragmatic structure. Encouragingly, (Elman, 1990) found that pragmatic information reflected in the statistical properties of corpora was detected by a Simple Recurrent Network and was encoded via refinement of the syntactic encoding. (Servan-Schreiber et al., 1991) found that subtle statistical biases of the type that plausibly stem from pragmatic knowledge can help an SRN learn syntactic structure. (Weckerly, 1995) provided sentence processing evidence that syntactic and pragmatic information are not independent but are intertwined in a way predicted by the SRN. I have not studied this kind of information structure in the FLNN yet, but it is clearly important, and a good next step will be to see if the fractal analysis can help elucidate the nature of the intertwining.

Final remark: In some dynamical approaches to development as well as in some connectionist work, there is a tendency to view ideal concepts as misleading. For example, (Thelen, Schoener, Scheier, & Smith, 2001) remark:

> Does this [Dynamical Field Theory] model have anything to say about Piaget's issue: when do infants acquire the object concept? We believe this question is ill-posed and cannot be answered because there is no such thing as an "object concept" in the sense of some causal structure that generates a thought or a behavior (Smith, Thelen, Titzer, & McLin, 1999). There is only "knowledge" of objects as embedded in the immediate circumstances and the history of perceiving and acting in similar circumstances. (p. 34)

In keeping with this view, the present work portrays a world in which knowledge is nothing more than experience-tuned action (Elman, 2004), there is considerable flexibility of behavior, and virtually any behavior can occur at any time in an organism's life if strong enough biases encourage it. And yet, according to the view put forth here, there is still value in referring to ideal, or absolute, forms: these are the limiting dynamical systems which exert a force-like influence on the learning process. An example in the present case is the infinite-state grammar which the network converges to in the limit of infinite training. While it is important to recognize that a realistic network will never behave exactly like this infinite-state grammar, the grammar nevertheless provides insight into the nature of the system that the network "uses" to organize its knowledge of its training data. The current example thus encourages one to look for analogous attractors in other domains in which development appears to cleave to, or consort with, coherent patterns of behavior.

References

Barnsley, M. ([1988]1993). *Fractals everywhere, 2nd ed.* Boston: Academic Press.

Bodén, M., & Wiles, J. (2002). On learning context-free and context-sensitive languages. *IEEE Transactions on Neural Networks*, *13*(2), 491–493.

Bresnan, J., & Nikitina, T. (2003). *On the gradience of the dative alternation.* (Manuscript, Department of Linguistics, Stanford University. Available at http://www-lfg.stanford.edu/bresnan/)

Casey, M. (1996). The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction. *Neural Computation*, *8*(6).

C Fisher, L. G., H Gleitman. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, *23*, 331-392.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton and Co.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*, 157–205.

Crutchfield, J. P., & Young, K. (1990). Computation at the onset of chaos. In W. H. Zurek (Ed.), *Complexity, entropy, and the physics of information* (pp. 223–70). Redwood City, California: Addison-Wesley.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–225.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Science*, *??*(??).

Gers, F. A., & Schmidhuber, J. (2001). LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE Transactions on Neural Networks*, *12*(6), 1333–1340.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural Computation*, *12*, 2451–2471.

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, *25*(3), 407–454.

Givon, T. (1971). Historical syntax and synchronic morphology: an archaeologist's field trip. In *Proceedings of the 7th regional meeting of the chicago linguistics society* (Vol. 7, pp. 394–415).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization.* Cambridge, England: Cambridge University Press.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*(2), 15-47.

Lindenmayer, A., & Prusinkiewicz, P. (1989). Developmental models of multicellular organisms: a computer graphics perspective. In C. G. Langton (Ed.), *Artificial life* (pp. 221–250). Redwood City, CA: Addison-Wesley. (A proceedings volume in the Santa Fe Institute Studies in the Sciences of Complexity)

Miller, G. A., & Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control*, *7*(3), 292–303.

Moore, C. (1998). Dynamical recognizers: Real-time language recognition by analog computers. *Theoretical Computer Science*, *201*, 99–136.

Pearlmutter, N. J., & MacDonald, M. A. (1995). Probabilistic constraints and working memory capacity in syntactic ambiguity resolution. *Journal of Memory and Language*, *43*, 521-542.

Pinker, S. (n.d.). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pollack, J. B. (1987). *On connectionist models of natural language processing.* (Ph.D. Thesis, Department of Computer Science, University of Illinois)

Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation*, *13*(9).

Rodriguez, P., & Wiles, J. (1998). Recurrent neural networks can learn to implement symbol-sensitive counting. In M. Jordan, M. Kearns, & S. Solla (Eds.), *Advances in neural information processing systems 10* (pp. 87–93). Cambridge, MA: MIT Press.

Rodriguez, P., Wiles, J., & Elman, J. (1999). A recurrent neural network that learns to count. *Connection Science*, *11*(1), 5–40.

Rohde, D. (2002). *A connectionist model of sentence comprehension and production.* (Unpublished PhD Dissertation, Carnegie Mellon University)

Rohde, D., & Plaut, D. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Journal of Memory and Language*, *72*, 67–109.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal represenations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing, v. 1* (pp. 318–362). MIT Press.

Savitch, W. J. (Ed.). (1987). *The formal complexity of natural language.* Norwell, MA: Kluwer.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J. L. (1991). Graded state machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning*, *7*, 161–193.

Siegelmann, H. T. (1999). *Neural networks and analog computation: Beyond the turing limit.* Boston: Birkhäuser.

Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of action: The task dynamics of the a-not-b error. *Psychological Review*, *106*, 235–260.

Tabor, W. (1991). Sentence processing and linguistic structure. In S. Kremer & J. Kolen (Eds.), *Field guide to dynamical recurrent networks.* IEEE Press.

Tabor, W. (1994). *Syntactic innovation: A connectionist model.* (Ph.D. dissertation, Stanford University)

Tabor, W. (1995). Lexical change as nonlinear interpolation. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th annual cognitive science conference.* Lawrence Erlbaum Associates.

Tabor, W. (2000). Fractal encoding of context-free grammars in connectionist networks. *Expert Systems: The International Journal of Knowledge Engineering and Neural Networks*, *17*(1), 41-56.

Tabor, W. (2003). Learning exponential state growth languages by hill climbing. *IEEE Transactions on Neural Networks*, *14*(2), 444-446.

Tabor, W. (2005). *Fractal learning neural networks.* (Submitted manuscript, University of Connecticut, Department of Psychology: See http://www.sp.uconn.edu/ ps300vc/papers.html)

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355-370.

Tabor, W., & Terhesiu, D. (1994). *On the relationship between symbolic and neural computation.* (AAAI Technical Report FS-04-03. ISBN 1-57735-214-9)

Thelen, E., Schoener, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: a field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*, 1–86.

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

Thom, R. (1985). *Structural stability and morphogenesis.* Reading, MA: Benjamin.

van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review*, *99*(3), 395–417.

van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, vygotsky, and beyond. *Psychological Review*, *105*(4), 634–677.

Weckerly, J. (1995). *Object relatives viewed through behavioral, electrophysiological, and modeling techniques.* (Ph.D. dissertation, University of California, San Diego)

Wiles, J., & Elman, J. (1995). Landscapes in recurrent networks. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th annual cognitive science conference.* Lawrence Erlbaum Associates.

The dynamical field approach to cognition

John Spencer, Jeff Johnson, & John Lipinski

Introduction

Goal of this chapter is to provide an overview of the dynamic field approach to cognition and development

- Situated within the more general DST
- Builds on strong history of studying motor dev from a dynamics perspective
- Tries to stretch in cognitive domain largely populated by connectionist approaches

Given that situated within DST, shares core concepts.

Theme 1: Behavior reflects a dynamic balance among stability, instability, flexibility. Development is about changing this dynamic balance

- Identifies the nature of developmental change, i.e., this is one of the core things we need to be thinking about as developmentalists
- Changes the questions we ask as researchers

Theme 2: To study development, we must understand the integration of dynamics across multiple time scales. In particular, real-time dynamics constrain what is learned and what develops…

- The implication: we can't truly understand developmental change without understanding the dynamics that live at shorter time scales
- Might have to hold off on jumping to the longer time scales (though diversity is certainly a good thing in science)

Theme 3: given this focus on real-time behavior-in-context, how does this lead to a general theory of development that still provides rich theory-experiment relations?

- Principles generalize across domains
- But also need to trace ties among specific behaviors (same child developing in context 1, context 2, and so on…how link these?) We'll sketch our approach to this, but a clear challenge for both approaches (Connect & DST)
- Big result in the work we discuss: achieving both specificity of predictions and generality across tasks and time scales

Illustrate these concepts with our work on spatial cognition—so bridging the gap from motor control/development into cognition; Critically, though, we retain the same commitment to tight theory-experiment relations.

And because we’ve moved into the terrain occupied by connectionist networks, provides a fertile ground for comparison. Thus, at the end of the chapter, we’ll critique dynamic field approach and connectionism

Part 1. What is the dynamic field approach?

Overview of history of DFT

- Links to the many domains in which these concepts have been applied

Core ideas: be brief (see Schöner chapter)

Part 2. Real-time dynamics and dynamic balance

Overview of 7-layer model

- Going to focus on two time scales to illustrate concepts: real-time, dev time
- For discussion of our approach to learning time scale, see spacebook, Gregor’s chapter, etc.
- Importantly, we’ll highlight links to empirical data throughout this story to show tight theory-experiment links

Highlight stability concept

- Self-sustaining peaks
- Stability at midline

Highlight instability

- Early in development, lose peaks (ANB)
- Delay-dependent drift

Highlight flexibility; toward higher-level cognition

- From recall to change detection (with single item probe)

Part 3. Real-time (+ learning time) dynamics constrain what develops

Discuss Spatial Precision Hypothesis (SPH)

Some 'obvious' predictions

- Kids get better over development

Anything less obvious?

- Dev change in geometric bias (explain how real-time dynamics constrain this example)
- Getting qualitative change from quantitative change (tie this to Van Geert, Van der Maas, etc)

And offers powerful generalization

- Predicted developmental changes in position discrimination between 3 and 5 years (explain that same concepts as change detection; explain how real-time dynamics constrain this example)

Part 4. Evaluation

Weakness thus far: haven't closed the loop on development

- Point toward a potential way to do this: developmental robotics

Weakness: we generally tackle "low-level" issues, what about higher cognition

- Could refer to spatial language stuff here…but these examples are promissory notes. At the end of the day, we'll just have to keep plugging away to show that ultimately, our work will scale up.

Strength: generalizing across tasks and time scales while retaining strong behavioral predictions about real-time behavior-in-context

- Position discrim example strong here
- And within our framework, we have to start local (local task, local time scale) and generalize from there. Given that we've had some initial successes here, shows promise for the approach more generally.

Critique of connectionist approaches to dev

- They succeed at closing the loop; and more focus on learning/development; and more focus on higher level cognition

- But this comes at the cost of real-time specificity and groundedness
- Examples

Are these theories same/different? They are different.

But the great thing is: the strengths of one approach are the weaknesses of the other! Provides a fertile ground for learning from one another and collectively moving forward to achieve a richer understanding of development that captures the real-time dynamics of behavior-in-context AND how the system modifies itself via these real-time experiences over the longer time scales of learning and development.

Running Head: THE ROBOT AS A NEW FRONTIER

The Robot as a New Frontier for Connectionism and Dynamic Systems Theory

Matthew Schlesinger

Southern Illinois University Carbondale

Abstract

In this chapter, I suggest that the history of connectionism and dynamic systems theory (DST) is less important than their present and their future. In particular, I highlight the idea that what the two approaches have in common—including commitments to embodiment, multiple timescales of behavior, and the simulation of empirical data—provide a rich and fertile landscape for blending the two in a variety of novel ways. To illustrate some of the possible “hybrids” that might form by mixing elements of connectionism and DST, I review recent work in the field of robotics and adaptive behavior. This work not only exploits many of the same methods and concepts that play a central role in connectionism and DST, but also offers three valuable lessons for students of development: (a) robots are ideal tools for investigating learning and development, (b) robots also provide an appropriate platform for studying evolutionary changes, and (c) because robots are embodied and situated, they offer the means to systematically study embodied knowledge.

The Robot as a New Frontier for Connectionism and Dynamic Systems Theory

> By nature, men are nearly alike; by practice, they get to be wide apart.
>
> Confucius

Are connectionism and dynamic systems theory (hereafter, DST) separate approaches to development, or instead, different manifestations of a single, grand theory of development? One way to answer this question is to place the two approaches, side-by-side, and then to catalog all the similarities and differences, functionally, structurally, conceptually, and so on. One difficulty with such an analysis, however, is that neither approach is a static entity. Both have an "evolutionary past", in cultural terms, that is, the intellectual history that helped shape each approach. In much the same way, both connectionism and DST also have their own unique "developmental" trajectories, influenced by the major figures that have emerged and led their respective communities in a particular direction. More importantly, the simple act of dialog between the advocates (i.e., followers, practitioners, etc.) of connectionism and DST will inevitably influence the future of each approach, if only by the exchange of ideas and practices between the two.

The "compare-and-contrast" strategy, it seems, offers limited prospects. As an alternative, I suggest a more holistic approach, in which connectionism and DST are viewed in organismic terms, loosely speaking. What follows from this perspective? First, note that connectionism and DST share a similar historical heritage: both approaches became established in the latter quarter of the 20th century. In particular, both were motivated, in part, as a reaction to the prevailing winds of *mentalism* that were spread by the cognitive revolution. As a result, they share the view that knowledge is distributed and graded, rather than symbolic and centrally stored. Second, each has endeavored to formulate a broad, metatheoretical framework that identifies and investigates fundamental principles of development (e.g., continuity vs. discontinuity, endogenous vs. exogenous influences on development, etc.). Third, and perhaps most importantly, both connectionism and DST have constructed and utilized an impressive array of mathematical tools, methods, and models for describing, explaining, and predicting developmental processes.

To paraphrase Confucius, then: “By nature, connectionism and DST are nearly alike.” In any case, their developmental trajectories over the last 20 years have often traveled in parallel directions (though not always at the same time), confronting many of the same questions and challenges. Over time, though, many salient differences have also taken root, almost certainly as a consequence of the specific mathematical techniques that each approach has exploited (e.g., feed-forward networks vs. differential equations, principle component analysis vs. catastrophe theory, etc.). So it is probably also true that “by practice, they’ve gotten to be wide apart.”

My goal, then, is to provide an optimistic forecast for the future of connectionism and DST. In particular, I focus on the idea that—regardless of how similar or dissimilar connectionism and DST appear to be at this moment in their development—there are numerous signs that “hybridization” of the two approaches is not only possible, but also has already begun to occur. In the next section, I first provide a context for my forecast by reviewing three major, cross-cutting themes that are shared by connectionism and DST. Second, I highlight the evidence for an optimistic outlook by describing recent work in the field of adaptive behavior and robotics, which illustrates numerous examples of models that blend elements of connectionism and DST. Finally, in the conclusion I return to the cross-cutting themes and elaborate on each, in light of the progress that robotics researchers have made toward a hybrid approach.

Introducing Three Cross-Cutting Themes

To suggest that connectionism and DST are on converging paths is an admittedly hopeful prediction, especially in light of the fact that advocates for each hardly speak the same language! For example, some of the jargon used by connectionists includes *feed-forward activation, supervised learning,* and *gradient descent*, while the DST community talks about *stability versus instability, attractors*, and *bifurcations*. Still, these differences really beg the question: are they only skin deep, and if so, how deep are the "threads" they share and that bind the two together? In this section, I focus on three specific themes that cut across connectionism and DST, each helping to guide two separate developmental trajectories toward a common destination.

Core Concepts

Perhaps the strongest link between connectionism and DST is through the concept of *embodiment*. While the term is broad enough to be used in many different ways, both approaches generally agree that knowledge should not—maybe even cannot—be divorced from the physical context in which it is acquired and manifested. DST has often pushed this notion to its logical limits, that is, the *organism-environment coupling*, in which the concept of embodied knowledge includes both the organism's physical body and the "outside" world in which that body acts (e.g., Thelen, Schöner, Scheier, & Smith, 2001; Thelen & Smith, 1994).

While DST makes an explicit commitment to the concept of embodiment, connectionism appears at times to be less certain. In particular, throughout the 1990s a large number of neural network models were designed and studied that seemed to offer only a minimal, indirect role for the physical body and the physical environment during development (e.g., Munakata, McClelland, Johnson, & Siegler, 1997; Plunkett & Marchman, 1991; Quinn & Johnson, 1997; for a critical review, see Schlesinger & Parisi, 2001a). For example, consider the task of learning to convert present-tense verbs into their past-tense counterparts (i.e., verb inflection). By focusing in on a small piece of the information-processing stream, a model of this task might assume (1) that sensible auditory signals comes from “somewhere out there”, (2) that some disembodied perceptual system, “somewhere else” in the nervous system, parses the auditory stream into appropriate units (e.g., syllables and phonemes), and finally, that (3) after the morphology problem is solved, “some other part” of the nervous system will know what to do with that information (e.g., utter the correct form of the desired verb).

It would be incorrect, though, to suggest that connectionists are unaware or unconcerned with the concept of embodiment (e.g., Bechtel, 1997). Indeed, there are a number of recent models that explicitly incorporate both structural and functional elements of embodiment (e.g., Mareschal, Plunkett, & Harris, 1999; Mareschal & Shultz, 1993; Schlesinger, Parisi, & Langer, 2000; Schlesinger & Parisi, 2001a). For example, these models are more structurally complex than generic feed-forward models, reflecting the idea that neural embodiment requires a diverse range of neural structures and functions (e.g., “what” vs. “where” visual streams; see Mareschal, Plunkett, & Harris, 1999; Schlesinger, in press). In addition, these models pay careful attention to how sensory signals originate, sometimes simulating the transduction process itself (e.g., with a simulated retina; see Schlesinger, 2003). Finally, there are also a class of connectionist models—called *ecological neural networks*—that simulate *both* the organism and the environment, allowing the two to interact in exactly the tightly-coupled way that DST envisions (e.g., Parisi, Cecconi, & Nolfi, 1990).

As I emphasized earlier, a second, related concept shared by connectionism and DST is the idea of *distributed knowledge*. Again, while the language used by each approach is often quite different—for example, connectionism uses the term *distributed representations* while DST speaks of *soft assemblies*—the core concept is more or less the same. Each approach emphasizes the idea that that *whatever knowledge is* (and surely, there are important differences between the epistemology of connectionism and DST), it is not a passive store of information, located in a central location, waiting to be retrieved for a relevant situation. Insofar as connectionism and DST share a common view, it is that knowledge is fluid, dynamic, and somehow “spread out” over space and time.

Of course, advocates of DST may disagree with the idea that knowledge-as-distributed-representations is anything like the “in-the-moment”, ephemeral quality of the soft-assembly (e.g., the convergence of motor neuron activity in a particular context, at a particular moment, that gives rise to a particular reaching movement). Once again, this skepticism is likely the product of an over-abundance of standard, feed-forward network models in the literature (for a critical review of “backprop nets”, including an overview of alternative modeling approaches, see Schlesinger & Parisi, 2004). To alleviate this unfortunate bias, consider the wide array of connectionist models that (for some unknown reason) are less-often used and written about, but that more importantly, are intentionally designed so that their internal state is always a function of both (a) whatever sensory experience is occurring at that moment and (b) their internal state a moment before. For example, simple recurrent networks, hebbian networks, and auto-associator networks all exhibit the capacity for time-and state-dependent activity (e.g., Elman, 1990; Munakata & Pfaffly, 2004; Sirois, 2004). Watching one of these networks in action leaves the observer with the feeling that they are highly dynamic, “percolating”, “bubbling”, perhaps even “deliberating their options” in much the same way that a dynamic field model seems to do (egregious anthropomorphisms notwithstanding).

Time Scales

A second theme that links connectionism and DST together is the dimension of time. In particular, both approaches agree with the idea that to understand a given behavior, it must be analyzed over multiple timescales. Even the somewhat mundane example of an infant's reaching movement illustrates a behavior that spans a surprisingly wide range of timescales. To begin, there is (a) the microscopic timescale, including multi-millisecond events such as changes in neural and muscle activation patterns, (b) the macroscopic, psychologically-salient timescale during which overt behaviors occur (i.e., those visible to the naked eye, e.g., movement of the infant's hand toward a cup). Taken together, these first two timescales are conventionally referred to as *realtime*. In addition, there is also (c) the timescale including and extending just beyond realtime, sometimes called the *microgenetic* timespan (e.g., minutes, hours, and days), (d) the *ontogenetic* history of the infant, spanning days, weeks, and months, which allows the cumulative outcomes of past reaches to influence the one that is occurring right now, and finally (e) the *phylogenetic* history of the species, which imposes many of the biomechanical constraints that our young infant is learning to identify and exploit (e.g., the elbow and shoulder joints permit 2-dimensional versus 3-dimensional rotations, respectively).

A shared interest in multiple timescales of behavior is one thing. It is a different question, though, to ask how *in practice* connectionism and DST have addressed the issue. Indeed, DST makes a strong commitment to studying multiple timescales, while connectionism seems a bit more demure. Admittedly, it is probably a fair criticism to note that connectionist models of development have rarely investigated timescales at the "fast" end of realtime, which seems a bit ironic, given that the connectionist *zeitgeist* itself is inspired by the concept of neural processing. In addition, there are also few connectionist models of behavior at the speed of naked-eye realtime, though models of reaching are a notable exception (e.g., Kettner, Marcario, & Port, 1993; Schlesinger, Parisi, & Langer, 2000; Vos & Scheepstra, 1993).

Again, there may be an element of *déjà vu* at play here. The problem is not that connectionists are ambivalent about studying different timescales. Rather, it may follow from the pervasive practice of using generic feed-forward networks, which exist "out of time" insofar as they (more or less) instantaneously produce an output for a given input. In other words, their architecture lacks internal dynamics, and as a result, feed-forward networks are in a sense too blunt an instrument to study behavior as it unfolds.

To be fair however, there is a bright side to this story. While many connectionist models of development may treat the individual behavioral unit as a molar, indivisible whole, by definition (i.e., they *are* models of development, after all) virtually all of these same models are successful at relating the timescale of that behavioral unit to longer timescales. To illustrate, let us return to the problem of learning to inflect English verbs: for example, Plunkett and Marchman (1991) show how varying verb frequencies during learning can lead to the U-shaped acquisition pattern observed in English-speaking children (e.g., both the model and children produce "goed" after first learning to produce "went"). More generally, connectionist models of syntax acquisition, object permanence, and reasoning on the balance-scale, just to name a few, have addressed a variety of long-term developmental phenomena, including changes in knowledge representation, stages of development, and transition mechanisms (e.g., Elman, 1993; Munakata, 1998; Shultz, Mareschal, & Schmidt, 1994).

Constraints

Finally, note that all models, whether designed from the connectionist or DST perspective, are by necessity a simplified version of the behavioral phenomenon they are intended to represent. In other words, in order to capture a specific feature of development, a model must leave other things out. A model of perceptual categorization may ignore the fact that visual inputs include color and stereoscopic information. A model of semantics and word-learning may ignore the fact that different cultures interact with their children in different ways.

While the choice of what to leave in is by no means an exact science, there are several meaningful metrics for deciding whether a model "has enough" in it to be a reasonable simulation of a real-world phenomenon. One, of course, is parsimony (e.g., Occam's razor): if two models are equivalent in performance, the one with less "stuff inside" (i.e., fewer untested assumptions, free parameters, etc.) is to be preferred. A second metric concerns how closely the output of the model resembles its real-world counterpart. For example, one weather model may predict "rain tomorrow", while another may require several interpretive steps before its output can be converted into a weather forecast. A third, related metric involves what happens on the input side. In particular, the more desirable model of two is the one that requires as little massaging or preprocessing of the raw input data as possible.

To be sure, the common constraint that puts these metrics into a single, coherent package is *empirical data*. In other words, the story begins when a behavioral researcher gives the model-builder a behavioral dataset, and asks "Can you model this?" Three fundamental (but not exclusive) goals are to (a) capture performance differences in the data—with an emphasis on developmental patterns—while keeping the (b) input and (c) output of the model as similar (or analogous) as possible to the stimuli presented to, and the responses produced by, respectively, the participants in the behavioral study.

At first glance, DST models appear to have an advantage over connectionist ones with regard to modeling specific behaviors. In particular, a model that does a better job of capturing behavior in realtime is likely to have a close correspondence between its output and real-world behaviors (e.g., Schutte & Spencer, 2002; Thelen, Schöner, Scheier, & Smith, 2001). Despite their strengths on the output side, however, these models have had less success in exploiting input representations that are naturalistic, or at least close analogs to the kinds of stimuli that are used in comparable behavior tasks.

In contrast, connectionist models have some ground to cover before they can adequately claim to capture empirical data patterns in more than a coarse, roughly analogous sense. Indeed, I have already criticized generic feed-forward models as (a) suffering from disembodiment while (b) lacking the capacity for internal dynamics. To the list we can now add a third deficit: many connectionist models of the last decade only weakly resembled, if at all, the behavioral phenomena they were designed to study. A good example of this tendency is offered by looking at models of the balance-scale task, in which a child is shown a balance scale—with weights on each side of a fulcrum—and is asked to predict which side, if either, will tip. Connectionist models of this task "simplified" things by distilling the input into four numeric features (the two weights, and their distances from the fulcrum, respectively); the output, meanwhile, is two dichotomous variables, scaled so that three outcomes can be "read off" the model (i.e., tip left, tip right, or balance; e.g., McClelland, 1989; Shultz, Mareschal, & Schmidt, 1994).

Despite being somewhat impoverished analogs of real-world experience, these models remain a valuable research tool because they are able to account for a variety of developmental phenomena (e.g., four stages of reasoning on the balance-scale task). In other words, they are often good models of the ontogenetic timescale. Nevertheless, there is an emerging consensus that if connectionist models of development are going to progress beyond their infancy—metaphorically speaking—they must undergo a number of qualitative changes, including (a) capturing behavioral patterns at the level of realtime, and (b) providing a closer correspondence (on both the input and output sides) to behavioral data (e.g., Christiansen & Chater, 2001; Schlesinger & Parisi, 2001a). In the next section, I survey a wide variety of research programs in the field of robotics and adaptive behavior, with the goal of demonstrating not only that the metamorphosis of connectionism is already well underway, but also that DST is coming along for the ride.

Linking Connectionism and DST through Robotics

Sheldon Kopp (1972), in a quirky blend of Buddhism and psychotherapy, writes in *If you meet the Buddha on the road, kill him!*: "The Zen way to see the truth is through your everyday eyes. It is only the heartless questioning of life-as-it-is that ties a man in knots" (p. 187). Perhaps he is right, or perhaps not. In the meantime, what seems more certain is that as we collectively wring our hands, pondering whether connectionism and DST are one thing or two, researchers in other disciplines have gone their merry way, using their "everyday eyes" to explore reality. And to make matters worse, they have also been using our tools and our ideas, while our attention was focused elsewhere.

To be fair, it is more than a little disingenuous to suggest that the underlying theoretical tenets or mathematical tools that play a central role in connectionism and DST *belong*, in any concrete sense, to either approach. Of course, they are borrowed, or more accurately, shared. And so it is somewhat inevitable that a brief survey outside of psychology reveals a life for connectionism and DST that we might never have suspected or imagined. Indeed, engineers, computer scientists, roboticists, and others who study adaptive behavior have retooled and reshaped the two approaches, blending them into a variety of unique models, paradigms, and computational "hybrids".

In this section, I sample from recent work in robotics and adaptive behavior, highlighting convergent and complimentary applications of connectionism and DST. This survey is organized to focus on three fundamental lessons for students of development. First, the robotic platform—whether real or simulated—is an ideal tool for studying both microgenetic and ontogenetic changes. Second, the field of *evolutionary robotics* supplements the developmental perspective with the means to study phylogenetic (i.e., inherited) changes. And finally, robotics provides a comprehensive platform for studying embodied knowledge.

From Realtime to Developmental Time

A convenient fact about most robots is that they are normally designed to "exist" in the here-and-now. In other words, their behaviors are often temporally scaled so that (a) they are observable to the naked eye, and (b) they correspond to the behaviors of real organisms (e.g., a colony of ants, a mechanical dog, a humanoid robot, etc.). As a consequence, robots are interesting devices in their own right that elicit a wide range of social-interactive behaviors from curious humans (e.g., Adams, Breazeal, Brooks, & Scassellati, 2000). In more formal language, the "life" of a robot spans multiple timescales, ranging from realtime up to ontogenetic or developmental time (i.e. from milliseconds and minutes, up to days and years).

Realtime. The task of designing a robot that can operate in realtime is a fundamental challenge for a robot-builder. At a minimum, a robot is a *sensorimotor system*: it has *sensors* that operate in one or more modalities (e.g., tactile, infrared, sonar, digital video, etc.), which are used to gather information about its current state, and it has *effectors* (motor systems) that enable it to move, orient its sensors in a new direction, and more generally, interact with its environment. In engineering terms, the physical device that is the robot is called the *plant* (e.g., like a power plant). Between the sensors and effectors, the robot-builder designs and implements an input-output system called a *controller*. The controller converts sensory inputs into desired motor outputs, through a mapping or function called a *policy*.

Researchers in the field of *machine learning* pursue methods for computing, discovering, or learning *optimal policies*, that is, policies that enable a robot (or other input-output device) to reach a desired goal as efficiently (or reliably, quickly, etc.) as possible. It is important to note that finding an optimal policy depends on whether the number of states and actions for a given task is within computationally-tractable limits or not (e.g., tic-tac-toe vs. backgammon or chess). As a consequence, for some tasks an optimal policy can be computed in advance by mathematical analysis. For other tasks—particularly those with large, multidimensional state-spaces (i.e., possible sensor values)—an approximation method is used to estimate the optimal policy. In particular, the latter case is often the starting point for researchers in the field of *adaptive behavior*, who use *incremental learning methods* to gradually improve the behavior of their robot (or other autonomous agent) as it interacts with a real or simulated environment. In other words, analytical methods correspond to *fixed policies*, while incremental methods correspond to *adaptive policies* that are modified through experience.

As we focus on robotics researchers who employ incremental learning methods, three interesting details emerge. First, as a mechanical device with multiple limbs and joints, the robot can be interpreted as a nonlinear dynamical system. This is true because jointed or segmented robots, like biomechanical systems, are composed of interconnected, physical links that do not move independently (e.g., Turvey, Fitch, & Tuller, 1982). In particular, when the robot moves one portion of its body, that movement generates *reactive forces* which indirectly cause other parts of its body to move (e.g., shoulder rotation causes movement of the forearm). A traditional control strategy for avoiding this problem is to *linearize the plant* by overdampening (i.e., stiffening) each of the robot’s joints, so that reactive forces are minimized or suppressed (this helps explain the slow, stiff movement of “older” humanoid robots).

A second interesting detail that emerges from the study of robotics and adaptive behavior is that many researchers use artificial neural networks as a computational structure for both acquiring and storing the policy. In particular, because neural networks implement a form of *nonlinear function approximation*, they are able not only to transform sensory input into desired motor output (e.g., movement of the limbs), but also to predict the sensory inputs that are expected to occur as a result of a planned action (i.e., forward model; see Jordan & Rumelhart, 1992). Finally, a third detail is that because the space of sensor values is potentially massive (e.g., millions of possible combinations), a control policy that attempts to plan a movement—in realtime—in high-dimensional, kinematic coordinates will likely be too slow to handle the rapid flow of sensor data (e.g., 100 readings from a sensor per second), or to be able to adjust its behavior as circumstances change.

Taken together, these observations not only illustrate a creative mixing of ideas and methods from connectionism and DST, but more importantly, they also highlight an emerging theme in the field of robotics and motor control. In particular, while traditional robotics emphasizes eliminating the nonlinear properties of dynamic systems while planning in high-dimensional coordinates, a growing number of researchers have advocated the use of DST as a alternative to classical robotics methods, because DST seeks to exploit, rather than suppress, the natural dynamics of the mechanical system. For example, they argue that dynamic control methods are (1) more cost-efficient in energetic terms, (2) more cost-efficient in cognitive terms (e.g., “planning “ at a low-dimensional, more abstract level of behavior), (3) more biologically plausible, (4) generalizable or transferable across sensorimotor systems, and (5) robust over noisy, uncertain, or unpredictable environments (e.g., Bizzi, Mussa-Ivaldi, & Giszter, 1991; Flash, 1987; Grupen, 2003; Schaal, Kotosaka, & Sternad, 2000; Schweighofer, Arbib, & Kawato, 1998; Williamson, 1998).

Developmental time. Compared to the study of adaptive behavior in realtime, robotics researchers have devoted less attention to studying behavioral changes that occur over days, weeks, months, and years. Nevertheless, the disciplines of robotics and developmental psychology have recently begun to interact and exchange ideas, giving rise to a hybrid field of study called *developmental robotics* (e.g., Lungarella, Metta, Pfeifer, & Sandini, 2003; Weng, et al., 2001). The two goals of developmental robotics are reciprocal: (1) to exploit basic principles of learning and development in order to build better robots, and (2) to use robots and other artificial agents as a platform for studying developmental processes.

It should be stressed that while some developmental robotics research is, at best, loosely inspired by developmental psychology, there are numerous studies that take development seriously. For example, one active area of research focuses on the fundamental issue of development as a *sequence of qualitative changes* (i.e., stages). This question has been investigated in a variety of ways, including robotic models and simulations of spatial navigation, self-recognition, and hand-eye coordination (e.g., Hiraki, Sashima, & Phillips, 1998; Michel, Gold, & Scassellati, 2004; Schlesinger, Parisi, & Langer, 2000). A related area of research focuses on *growth of the body*, and corresponding developments in sensorimotor activity (e.g., Dominguez & Jacobs, 2003; Lungarella & Berthouze, 2002).

However, as Lungarella et al. (2003) note, the majority of papers in their review focus on sensorimotor development, that is, “competence in interacting with the local environment—in particular, basic visuo-motor competencies such as saccading, gaze fixation, joint attention, hand-eye co-ordination and visually-guided reaching” (pp. 175-176). The emphasis on sensorimotor development suggests that learning through interaction is indeed a difficult process to capture in a model, and raises the question: in contrast to learning a particular ability, when might it be better to use existing engineering knowledge to build in or hardwire the skill?

This question can be addressed in several ways. First, as I noted earlier, the use of analytical methods—in which a policy is engineered or computed by hand rather than learned—is limited to tasks or problems where the size of the state space is tractable. However, for problems with large state spaces, prior knowledge can be built into the search process, so that even when approximate or incremental methods are employed, less time is spent searching regions of the state space where sub-optimal solutions reside. On a related note, hardwired systems tend to do better when the environment is (a) *stationary* over time, and (b) predictable or *deterministic*. In contrast, systems that learn, adapt, or develop generally have an advantage over hardwired ones when (a) the state space is large, (b) the environment is non-stationary, and (c) sensory data are imperfect, uncertain, noisy, or *stochastic*.

Second, regardless of whether the performance of a developing robot is superior to that of a hardwired robot, many developmental researchers are more interested in the developmental process than in the outcome itself. Indeed, there is growing interest in the notion of “developmental immaturity”, that is, the idea that limiting or constraining early developmental capacities can facilitate long-term development (e.g., Bjorklund & Green, 1992). This approach has been successfully applied to a variety of perceptual and motor skills, including learning to swing, to reach, and to exploit binocular disparity as a depth cue (e.g., Dominguez & Jacobs, 2003; Lungarella & Berthouze, 2002; Schlesinger, Parisi, & Langer, 2000).

Evolutionary Time

Complementing the study of developmental robotics, there is a parallel community of researchers who investigate *evolutionary robotics* (e.g., Nolfi & Floreano, 2001; of course, there are some who do a mixture of both, i.e., “evo-devo”). The field of evolutionary robotics includes two major areas of research (e.g., Schlesinger, 2004). First, some focus on the use of *evolutionary computation* (i.e., neo-Darwinian principles of variation and selection) as a search strategy for exploring large state-spaces. Research from this perspective tends to focus on issues central to engineering and mathematics, including efficient search for optimal policies, tuning of algorithms, and scaling of policies from simulation to real-world environments. The first view is contrasted by a second area of evolutionary robotics research, in which the robot provides a platform for studying biological processes. In particular, many *artificial life* (Alife) researchers use the same neo-Darwinian algorithms, but with the goal of simulating all biological phenomena, both real and possible (Langton, 1995; Parisi, 1996).

The general research method of evolutionary robotics is to create a large population of agents, and then for all of the agents to compete on a given task. Because the individuals who are most fit are chosen to reproduce (i.e., determine the members of the next generation of robots), the performance of the population as a whole tends to improve over time. (This computational strategy illustrates the concept of *hill-climbing*, in contrast to *gradient-descent* algorithms like back-propagation-of-error, which seek to minimize performance errors.)

As a practical necessity, evolutionary robotics normally involves a mixture of simulation and real-world robots (e.g., Husbands, Harvey, Cliff, & Miller, 1997; for a detailed discussion, see Nolfi, Floreano, Miglino, & Mondada, 1994). In particular, note that there are no established methods for enabling robots to mate and reproduce! As a result, one research strategy is to (a) design a simulated robot and environment, and then (b) run the evolutionary algorithm in simulation. A related strategy is to first use the real robot to acquire a physical model of the sensors and effectors, and then to use the model as part of the simulated evolutionary process. Regardless of where learning occurs, it is important to ultimately verify the performance of the model by “downloading” the evolved control system or policy into a real robot, and testing it in a real environment.

It should be stressed that by including the evolutionary timescale, evolutionary robotics expands the range of issues that developmental researchers can investigate. For example, how does the evolutionary history of a species influence the development of a particular ability (e.g., motor skill, foraging and navigation, language acquisition; see Hinton & Nolan, 1987)? What precursors or components of the ability tend to be encoded genetically and inherited versus learned anew by each generation? How are different levels of biological adaptation related, during both evolution and development (e.g., genetic, neural, behavioral, etc.; see Cangelosi, Parisi, & Nolfi, 1994)? More generally, evolutionary robotics is an ideal platform for investigating *epigenetic processes*, that is, the role of genes and environment in the interaction between learning, maturation, development, and evolution.

An additional perspective offered by evolutionary robotics—one that is often overlooked by developmental researchers—is the population-based approach. In other words, evolutionary robotics also provides a comparative perspective, in which a variety of developmental patterns or trajectories can emerge within a population of artificial agents. As a concrete example, consider the development of *allocentric spatial orientation* (i.e., the use of external landmarks and spatial cues; e.g., Wang, Hermer, & Spelke, 1999). Using a mobile-robot within the evolutionary framework, Miglino and Lund (2001) studied this capacity and found that a number of component behaviors, including wall-following and orienting toward landmarks, emerged independently in a population of simulated robots, and then became coordinated, giving rise to successful navigation skills.

Of particular relevance to connectionism and DST is work in evolutionary robotics that synthesizes elements of both approaches. For example, Beer and colleagues (e.g., Beer, 1996; Slocum, Downey, & Beer, 2000) have studied populations of simulated agents, whose nervous systems are represented by continuous-time, recurrent neural networks (CTRNNs) that evolve over multiple generations. Like conventional recurrent networks, CTRNNs have a dynamic internal state that is time-dependent. In other words, their response to the same stimulus at two different points in time will vary as a function of their internal state (i.e., recent experience). From a functional perspective, then, CTRNNs are able to maintain a short-term internal representation that can serve as a memory trace of recent experience. Beer exploits this feature to show how a population of evolving CTRNNs can solve a number of “minimally cognitive” problems, such as (a) catching falling objects, (b) self-object discrimination, and (c) selective attention (e.g., Slocum, Downey, & Beer, 2000).

Robots as Embodied “Knowers”

To conclude the survey of robotics and adaptive behavior, I return to the issue of embodied knowledge. In particular, it is interesting to note that while there is wide variation in the types of platforms that robotics researchers use, and even wider variation in the questions they investigate, there is a broad consensus that the physical structure of the robot and the environment that it inhabits collectively exert a fundamental influence over how the robot learns, develops, and evolves (e.g., Brooks, 1991; Grupen, 2003; Lungarella et al., 2003; Weng et al. 2001).

This issue highlights a critical difference between connectionist models that are disembodied (i.e., have only an implied connection to the environment), and those that are embodied and situated (i.e., occupy and interact with an explicitly defined environment). On the one hand, the strategy of “classical connectionism” is to present input (sensory) patterns in random sequences, to minimize the influence of experiencing one input pattern before another. Alternatively, networks that have a body and occupy an environment (i.e., autonomous agents, ecological networks, etc.) *actively sample* or *self-select* their sensory inputs (e.g., Nolfi & Parisi, 1993; Parisi, 1997). In other words, by virtue of their own behavior, embodied and situated networks are able to (a) limit or constrain the kinds of sensory data that reach their nervous system at any given moment, and more importantly, (b) modulate the pattern of data-flow over time (e.g., Schlesinger & Parisi, 2001a).

The study of embodied networks raises two important questions. First, does self-selection of sensory input facilitate or simplify any particular developmental tasks? Second, what similarities, if any, are there between the developmental trajectories of embodied networks and real organisms?

A recent model of hand-eye coordination in infants can be used to address both of these questions (Schlesinger, Parisi, & Langer, 2000; Schlesinger & Parisi, 2001b). The model employs a population-based approach, in which a collection of artificial neural networks are used to represent a population of motor units that control the movement of an eye and arm. The position of the arm endpoint is determined by three joint angles: (a) the orientation of the body axis, (b) the shoulder joint, and (c) the elbow joint. Each of the arm joints, in turn, is controlled by two muscles that pull in opposite directions (i.e., flexor-extensor pairs). Note that because the arm reaches along a 2-dimensional surface, only two of the three joint angles (i.e., the body axis and the elbow joint) are needed to reach an arbitrary point. In effect, then, the model is placed in a two degree-of-freedom (DOF) workspace, leaving it with an extra or redundant DOF.

The model rapidly acquires a set of control strategies that exploit its particular physical design (Schlesinger, Parisi, & Langer, 2000). For example, it quickly learns to “lock out” the shoulder joint by rotating the shoulder to one side of its body (it’s interesting to note that the “preferred” direction for rotating the shoulder varies from one simulation run to another), while it keeps the elbow joint away from the fully flexed or extended position. In addition, the model also rapidly shifts from a pattern of co-activating the muscle pair that control the shoulder (which results in little or no net movement of the upperarm), to a pattern of *reciprocal activation*, in which one muscle pulls while the other remains in a relaxed state.

These findings bear directly on both of the questions raised earlier. First, by freezing or locking out the movement of the shoulder joint, the model effectively eliminates one dimension of the 3-dimensional joint search-space. In other words, rather than attempting to exhaustively sweep through the space of all possible torso-shoulder-elbow combinations, the model simplifies the search problem by holding the position of the shoulder constant, while exploring various torso-elbow configurations. Second, and more importantly, the strategy of “reducing the search space” does not appear to be an isolated phenomenon, but instead may be ubiquitous both during the development of reaching in infants (i.e., freezing degrees of freedom, muscle coactivation, etc.; e.g., Berthier, Clifton, McCall, & Robin, 1999; Spencer & Thelen, 2000), as well as during motor learning throughout the lifespan (e.g., Vereijken, van Emmerik, Whiting, & Newell, 1992).

To conclude the discussion on embodied knowledge in robots, it should be emphasized that research in this area is not limited to the development of motor control or visual-motor skill.

Indeed, there is a broad community of researchers in adaptive behavior who agree that sensorimotor activity is a fundamental starting point for cognition in general. The concept of *sensorimotor cognition*—that all forms of knowing are grounded in actively sensing and exploring the world—is not only historically rooted in Piaget's *genetic epistemology* (1952), but is also supported by work in linguistics, philosophy, and other areas of cognitive science (e.g., Johnson, 1987; Keijzer, 1998; O'Regan, 1992; Varela, Thompson, & Rosch, 1992).

As a concrete example of sensorimotor cognition, consider the related problems of (1) learning to perceptually discriminate a set of different objects, while (2) forming a set of categories, within which (either functionally or perceptually) similar objects can be grouped. Scheier and Pfeifer (1995) addressed this question by studying a mobile robot that develops a functional classification of "graspable" and "non-graspable" objects. Interestingly, Scheier and Pfeifer designed the robot's environment so that passively viewing the target objects results in ambiguous perceptual information (i.e., *perceptual aliasing*). As a result, the robot learns to exploit regularities in its own sensorimotor activity—that is, contingencies between its movement and changes in the activity of its infrared sensors—as the basis for differentiating and grouping objects. A similar approach has also been used to investigate other cognitive phenomena, including spatial navigation, intercepting or catching falling objects, and the perception of time (e.g., Nolfi & Parisi, 1999; Slocum, Downey, & Beer, 2000; Tuci, Trianni, & Dorigo, 2004).

Summary and Conclusions

This chapter began with an important question: Are connectionism and DST separate approaches to development, or instead, different manifestations of a single, grand theory of development? My strategy for addressing this question was to propose an analogy, in which the two approaches are interpreted within a biological framework. A number of interesting consequences follow from this perspective. First, the systems of interrelated concepts, scientific methods, and empirical data that comprise connectionism and DST are not viewed as static structures, but rather, as complex entities that change over time. A second consequence of the “organismic” analogy—and one that is clearly highlighted throughout this volume—is that there is no single definition or description that captures either approach. Surely there are prevailing *ways of thinking* like a DST or connectionist researcher, and common practices that complement these ways of thinking, but there are also a number of important and meaningful differences between the advocates that subscribe to each approach.

To extend the analogy one step further, it might prove helpful to think of connectionism and DST as two extant species, which may or may not share a common ancestor. In this sense, then, neither approach is one single entity, but rather, each can be better understood as a large family of distinct-but-related instances that have emerged from a common set of ideas. What is gained by this view? First, the fact that there are multiple “flavors” of connectionism and DST is not a weakness or a deficit, but instead, a healthy indication that each approach has both grown and diversified as a result of “adaptive pressures”. Second, and perhaps more importantly, the idea that connectionism and DST are each like multigenerational families argues against the likelihood of an ultimate, grand unification. Like a metaphorical Adam and Eve, or Noah’s ark, which pair of approaches would be chosen to define a unified theory? In other words, which specific instances (or general features) of DST and connectionism would be selected to unify the two?

My goal has been to argue (implicitly) *against* the strategy of distilling connectionism and DST to their fundamental essence(s), and then attempting to compare and contrast the two.

Instead, consider the following thought-experiment: imagine for a moment that you were a farmer, looking for a way to cross two strains of tomatoes into a new, more vigorous breed. With literally hundreds of different breeds to choose from, you would focus on those with the most appropriate or desirable features, such as color, shape, resistance to pests, shelf-life, and so on. I would argue, using the same logic, that rather than asking if connectionism and DST are one thing or two, we instead should focus on creative and novel ways of blending the two approaches into theoretical and methodological "hybrids" that bring together the best features of each approach, as appropriate for a particular empirical question, developmental task, or theoretical issue.

If indeed hybridization is a metaphor that will enable the developmental trajectories of connectionism and DST to align and perhaps even converge, then as I have already illustrated, there is ample evidence from the field of adaptive behavior that several different connectionism-DST hybrids are both viable and self-sustaining. In the final section, I conclude by returning once more to the cross-cutting themes that I discussed earlier and pulling together several of the issues that I highlighted in previous sections.

Discussion of Cross-Cutting Conceptual Themes

Core concepts. At the beginning of the chapter, I focused on two core concepts that play a fundamental role in connectionism and DST. One of these is the concept of *embodiment*. While connectionism and DST may agree that the physical body and the physical environment constrain (influence, shape, etc.) developmental processes, each approach has also explored different ways for investigating these constraints. For example, DST research has often focused on changes in behavior during realtime performance, while connectionist research has focused on long-term changes on the developmental timescale. The problem is also exacerbated by the fact that many connectionist models have provided only a vague, implicit role for the physical body and environment.

Fortunately, studying development in a (real or simulated) robot collapses all of these issues onto a single, coherent platform. First, a robot is, by definition, an embodied object that occupies a real or virtual environment. Second, it acts and exists through both realtime and developmental time (and perhaps even evolutionary time). Third, at the robot-builder's discretion, a number of physical properties can be incorporated into the robot's design that provide a basis for generating nonlinear forces, including elasticity, gravity, inertia, and friction (i.e., the robot *is* a dynamical system). Fourth, and also at the robot-builder's discretion, an artificial neural network can be used as the control system used for acquiring and storing the policy. Clearly, a robot provides connectionist and DST researchers with a unique opportunity to combine many of their ideas and techniques in an embodied, situated, and physically grounded system.

I also highlighted the concept of *distributed knowledge* as a theme that bridges connectionism and DST. Here, unfortunately, researchers in robotics, and perhaps also in AI more broadly, have not always been as sympathetic to the idea. This may be due to the influence of mechanical and computer engineers, who are often less interested in epistemology than in *building systems that work*. As a result, there is a community of researchers who favor the use of symbolic or centrally-stored knowledge in their models (e.g., representing the policy as a *lookup table*). Nevertheless, there are also numerous examples of researchers who employ distributed knowledge systems, including nonlinear neural networks, fuzzy-logic systems, classifier systems, and other forms of *nonlinear function approximation*. These computational approaches, and others like them, are obviously an ideal platform for connectionist and DST researchers to facilitate a dialog that focuses on defining and investigating distributed knowledge.

Time scales. A second major theme that cuts across both connectionism and DST is the issue of learning and development over multiple timescales. As I have already suggested, because robots are physical entities that exist through time, they provide a natural framework for studying change both within and between the realtime, microgenetic, ontogenetic, and phylogenetic timescales. Indeed, as robots acquire the capacity to design or repair themselves (i.e., robot self-replication!), the distinction between simulated or modeled changes in artificial systems, and biological or psychological changes in real organisms, may begin to break down. To put it another way, if ongoing research with robots indicates a future trend, the robots of the future may have a life of their own!

Constraints. The final cross-cutting theme that I described is the need for computational models of learning and development to map more directly onto the kinds of situations and behaviors that developmental psychologists actually study. In other words, for computational models—whether motivated by connectionism or DST—to become fully relevant and accessible to “good old fashioned” developmental researchers, they must be able to generate datasets that are more or less comparable to those generated by human subjects.

Unfortunately, the practice of using a model to replicate the performance of an infant or child in a particular study may be the exception, rather than the norm (cf. Simon, 1998). Note that by “replicate” I don’t mean “roughly” or “approximately” simulate the situation and the behavior. Instead, I suggest that we raise the standard as high as possible, and aim for placing robots or simulated agents in situations that are (more or less) the same as those experienced by human subjects, and enable the agents to respond to those situations in behaviorally “realistic” or meaningful ways (e.g., looking, grasping, vocalizing, etc.). Although this particular approach is comparatively new, there are a few promising examples. For instance, both real and simulated robots have recently been used to simulate infants' preferential-looking or visual attention during violation-of-expectation studies (e.g., Lovett & Scassellati, 2004; Schlesinger, in press; Weng, Zhang, & Chen, 2003). Hopefully, as this trend continues to grow, researchers from both the connectionist and DST approaches will not only find a common ground on which to stand, but also a common language with which to communicate, share ideas, and collaborate.

References

Adams, B., Breazeal, C., Brooks, R., Scassellati, B. (2000). Humanoid robots: A new kind of tool. *IEEE Intelligent Systems, 15*, 25-31.

Bechtel, W. (1997). Embodied connectionism. In D.M. Johnson & C.E. Erneling (Eds.), *The future of the cognitive revolution* (pp. 187–208). New York: Oxford University Press.

Beer, R.D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Mataric, J. Meyer, J. Pollack and S. Wilson (Eds.), *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (pp. 421-429). Cambridge, MA: MIT Press.

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

Bizzi, E, Mussa-Ivaldi, F.A., & Giszter, S. (1991). Computations underlying the execution of a movement: A biological perspective. *Science*, *253*, 287-291.

Bjorklund, D.F., & Green, B.L. (1992). The adaptive nature of cognitive immaturity. *American Psychologist*, *47,* 46-54.

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

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

Christiansen, M.H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, *5,* 82-88.

Dominguez, M., and Jacobs, R.A. (2003). Developmental constraints aid the acquisition of binocular disparity sensitivities. *Neural Computation, 15*, 161-182.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, *14,* 179-211.

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

Flash, T. (1987). The control of hand equilibrium trajectories in multi-joint arm movements. *Biological Cybernetics*, *57*, 257-274.

Grupen, R.A. (2003). A developmental organization for robot behavior. 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. 25-36). Sweden: Lund University Cognitive Studies.

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

Hiraki, K., Sashima, A., & Phillips, S. (1998). From egocentric to allocentric spatial behavior: A computational model of spatial development. *Adaptive Behavior*, *6*, 371-391.

Husbands, P., Harvey, I., Cliff, D., & Miller, G. (1997). Artificial evolution: A new path for artificial intelligence? *Brain and Cognition*, *34*, 130-159.

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

Jordan, M.I., & Rumelhart, D.E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, *16*, 307-354.

Keijzer, F.A. (1998). Doing without representations which specify what to do. *Philosophical Psychology*, *11*, 269-302.

Kettner, R., Marcario, J., & Port, N. (1993). A neural network model of cortical activity during reaching. *Journal of Cognitive Neuroscience*, *5*, 14-33.

Kopp, S.B. (1972). *If you meet the Buddha on the road, kill him! The pilgrimage of psychotherapy patients*. New York: Bantam.

Langton, C.G. (Ed.). (1995). *Artificial life: An overview*. Cambridge, MA: MIT Press.

Lovett, A., & Scassellati, B. (2004). Using a robot to reexamine looking time experiments. *Proceedings of the Third International Conference on Development and Learning* (pp. 284-291). Cambridge, MA: IEEE.

Lungarella, M., & Berthouze, L. (2002). On the interplay between morphological, neural, and environmental dynamics: A robotic case-study. *Adaptive Behavior, 10*, 223-241.

Lungarella, M., Metta, G., Pfeifer, R., & Sandini, G. (2003). Developmental robotics: A survey. *Connection Science, 15*, 151-190.

Mareschal, D., Plunkett, K., & Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviours in infancy. *Developmental Science, 2*, 306-31

Mareschal, D., & Shultz, T.R. (1993). A connectionist model of the development of seriation. *Proceedings of the fifteenth annual Conference of the Cognitive Science Society* (pp. 170–175). Hillsdale, NJ: Erlbaum.

McClelland, J.L. (1989). Parallel distributed processing: Implications for cognition and development. In R. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 9-45). New York: Oxford University Press.

Michel, P., Gold, K., & Scassellati, B. (2004). Motion-based robotic self-recognition. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2763-2768). Cambridge, MA: IEEE.

Miglino, O., & Lund, H.H. (2001). Do rats need euclidean cognitive maps of the environmental shape? *Cognitive Processing , 4*, 1-9.

Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A PDP model of the $A\overline{B}$ task. *Developmental Science*, *1*, 161-211.

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.

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science*, *7*, 141-148.

Nolfi, S., & Floreano, D. (2001). *Evolutionary robotics. The biology, intelligence, and technology of self-organizing machines*. Cambridge, MA: MIT Press.

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. (1999). Exploiting the power of sensory-motor coordination. In: D. Floreano, J-D. Nicoud, & F. Mondada (Eds.), *Advances in Artificial Life: Proceedings of the Fifth European Conference on Artificial Life* (pp. 173-182). Berlin: Springer-Verlag.

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.

Parisi, D. (1996). Computational models of developmental mechanisms. In R. Gelman and T. 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., Cecconi, F., & Nolfi, S. (1990). Econets: Neural networks that learn in an environment. *Network,* **1,** 149–168.

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

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition*, *38***,** 1–60.

Quinn, P.C., & Johnson, M.H. (1997). The emergence of perceptual category representations in young infants: A connectionist analysis. *Journal of Experimental Child Psychology, 66***,** 236–263.

Schaal, S., Kotosaka, S., & Sternad, D. (2000). Nonlinear dynamical systems as movement primitives. In *Proceedings of the First IEEE-RAS International Conference on Humanoid Robotics*. Cambridge, MA: IEEE.

Scheier, C. & Pfeifer, R. (1995). Classification as sensorimotor co-ordination: A case study on autonomous agents. In F. Moran, A Moreno, P. Chacon, & J.J. Merelo (Eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life* (pp. 657-667). New York: Springer Verlag.

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

Schlesinger, M. (2004). Evolving agents as a metaphor for the developing child. *Developmental Science*, *7*, 158-164.

Schlesinger, M. (in press). Decomposing infants’ object representations: A dual-route processing account. *Connection Science*.

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

Schlesinger, M., & Parisi, D. (2001b). 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. (Eds.). (2004). Beyond backprop: Emerging trends in connectionist models of development. [Special section]. *Developmental Science, 7*, 131-132.

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

Schutte, A.R., & Spencer, J.P. (2002). Generalizing the dynamic field theory of the A-not-B error beyond infancy: Three-year-olds' delay- and experience-dependent location memory biases. *Child Development, 73*, 377-404.

Schweighofer, N., Arbib, M.A., & Kawato, M. (1998). Role of the cerebellum in reaching movements. I. Distributed inverse dynamics control. *European Journal of Neuroscience, 10*, 86-94.

Shultz, T.R., Mareschal, D., & Schmidt, W.C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning*, *16*, 57-86.

Slocum, A.C., Downey, D.C. and Beer, R.D. (2000). Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In J. Meyer, A. Berthoz, D. Floreano, H. Roitblat and S. Wilson (Eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior* (pp. 430-439). Cambridge, MA: MIT Press.

Simon, T.J. (1998). Computational evidence for the foundations of numerical competence. *Developmental Science*, *1*, 71-78.

Sirois, S. (2004). Autoassociator networks: Insights into infant cognition. *Developmental Science*, *7*, 133-140.

Spencer, J.P. & Thelen, E. (2000). Spatially-specific changes in infants' muscle coactivity as they learn to reach. *Infancy, 1*, 275-302.

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*, 1-86.

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

Tuci, E., Trianni, V., & Dorigo, M. (2004). ‘Feeling’ the flow of time through sensorimotor co-ordination. *Connection Science*, *16,* 301-324.

Turvey, M.T., Fitch, H.L., & Tuller, B. (1982). The Bernstein perspective: I. The problems of degrees of freedom and context-conditioned variability. In J.A.S. Kelso (Ed.), *Human motor behavior: An introduction* (pp. 239-252). Hillsdale, NJ: Lawrence Erlbaum Associates.

Varela, F.J., Thompson, E., & Rosch, E. (1992). *The embodied mind: Cognitive science and human experience.* Cambridge, MA: MIT Press.

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.

Vos, J.E., & Scheepstra, K.A. (1993). Computer-simulated neural networks: An appropriate model for motor development? *Early Human Development*, *34*, 101-112.

Wang, R. F., Hermer, L., & Spelke, E. S. (1999). Mechanisms of reorientation and object localization by human children: A comparison with rats. *Behavioral Neuroscience*, *113*, 475-485.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., & Thelen, E. (2001). Artificial intelligence: Autonomous mental development by robots and animals. *Science*, *291*, 599-600.

Weng, J., Zhang, Y., and Chen, Y. (2003). Developing early senses about the world: “Object permanence” and visuoauditory real-time learning. *Proceedings of the International Joint Conference on Neural Networks* (pp. 2710 – 2715). Cambridge, MA: IEEE.

Williamson, M.M. (1998). Exploiting natural dynamics in robot control. In R. Trappl (Ed.), *Cybernetics and Systems ’98*. Austria: Austrian Society for Cybernetic Studies.

Connectionist approaches to perseveration:

Understanding universal and task-specific aspects of children's behavior

J. Bruce Morton & Yuko Munakata

Abstract

Infants and children often perseverate, repeating prior behaviors when they are no longer appropriate. These intriguing behaviors have been the focus of intense empirical and theoretical scrutiny, resulting in thorough documentation of numerous factors that influence perseveration. Formal models, both connectionist and dynamic systems, have played an important role in providing a unified framework for understanding these factors, and perseveration and development more generally. In this chapter, we focus on case studies of connectionist models of perseveration – for understanding infant reaching behavior in A-not-B tasks, and children's rule-guided behavior in card-sorting and speech interpretation tasks. These models demonstrate how general principles of graded and distinct types of representations provide a common framework for understanding perseveration across ages, task paradigms, and response modalities, while also capturing important details of children's performance. These models also generate novel and testable predictions. We compare and contrast these contributions from connectionist models with those from dynamic systems accounts of the A-not-B task, which place greater emphasis on the details of visually guided reaching. Such details may be essential for understanding nuances of infant performance on the A-not-B task, but the more general principles may have greater utility for understanding perseverative behaviors across the life span.

1. Introduction

Everyone perseverates at one time or another, repeating behaviors that were adaptive in an earlier context but that are no longer appropriate. After reorganizing the objects on our desk, for example, we might reach for the phone in its old rather than its new location. Infants and children are particularly robust perseverators. After repeatedly watching as an object is hidden in one location (A) and searching for it there, infants perseverate by reaching back to that location after watching as the object is hidden in a new location (B) (Diamond, 1985; Piaget, 1954). Infants will even make such perseverative A-not-B errors when the visible environment clearly specifies the appropriate behavior. For example, after repeatedly pulling a towel in one location (A) to retrieve a distant toy on the towel, infants perseverate by pulling the towel in that location when the toy is visibly no longer on it, instead of pulling a towel in a new location (B) that has a toy on it (Aguiar & Baillargeon, 2000). Similarly, three-year-old children often perseverate in simple card-sorting tasks by continuing to sort cards one way (e.g., by color) when asked to switch and sort them in a new way (e.g., by shape) (Zelazo, Frye, & Rapus, 1996). Finally, 6-year-old children often persist in judging a speaker’s feelings from what she says when asked to switch and judge her feelings from the sound of her voice (Morton, Trehub, & Zelazo, 2003).

These kinds of perseverative errors, and children's eventual success on such tasks, have been studied extensively, leading to rich characterizations of children's performance and the many factors that influence it (e.g., Brace, Morton, & Munakata, in press; Brooks, Hanauer, Padowska, & Rosman, 2003; Diamond, 1985; Jacques, Zelazo, Kirkham, & Semcesen, 1999; Kirkham, Cruess, & Diamond, 2003; Marcovitch &

Zelazo, 1999; Morton & Munakata, 2002b; Perner & Lang, 2002; Smith, Thelen, Titzer, & McLin, 1999; Towse, Redbond, & Cook, 2000; Wellman, Cross, & Bartsch, 1986; Zelazo, Mueller, Frye, & Marcovitch, 2003). As elaborated below, this work has established both commonalities in perseverative behaviors across task paradigms, as well as unique aspects of perseveration that are tied to specific details within a given task.

In terms of commonalities across task paradigms, children show similar patterns of perseveration across tasks that vary greatly in their response requirements (e.g., reaching vs. answering verbal queries vs. error detection; Jacques, Zelazo, Kirkham, & Semcesen, 1999; Munakata & Yerys, 2001; Piaget, 1954). In addition, children's performance improves with age across task paradigms (Diamond, 1985; Jacques & Zelazo, 2001; Zelazo et al., 1996). And, children show similar dissociations across task paradigms, demonstrating some knowledge of appropriate behaviors despite responding perseveratively. For example, in the A-not-B task, infants will occasionally look at the correct hiding location for a toy (Diamond, 1985; Hoftstadter & Reznick, 1996) or appear to recognize where the hidden toy should be (Ahmed & Ruffman, 1998), while nonetheless reaching perseveratively to a previous hiding location[1]. Similarly, in the card-sorting task, children will reliably answer questions correctly about what they are supposed to be doing (e.g., “Where do trucks go in the shape game?”), while sorting cards perseveratively according to a previously relevant rule (Zelazo et al., 1996).

In addition to common patterns across tasks, specific details of individual tasks strongly influence whether children perseverate. For example, the number and

---

[1] Although some researchers have challenged the reliability and interpretation of such dissociations, for example, because looking and reaching behaviors are often closely coupled (e.g., Smith & Scheier, 1998; Spencer & Schutte, 2004), looking and reaching measures nonetheless reveal differential sensitivity when compared systematically in the A-not-B task.

distinctiveness of hiding locations affects whether infants perseverate in the A-not-B task (Bremner, 1987b; Cummings & Bjork, 1983a, 1983b; Wellman et al., 1986). Whether a toy is placed on a towel versus physically attached to the towel affects whether infants perseverate in the towel-pulling task (Aguiar & Baillargeon, 2000). The types of cards used and how they are labeled affects whether children perseverate in the card-sorting task (Kirkham et al., 2003; Perner & Lang, 2002; Yerys & Munakata, in press; Zelazo et al. 2003). And, children show decalages in their behavior, succeeding on one task (e.g., card-sorting) at a given age while perseverating on another task (e.g., speech-interpretation) that is similar in structure but different in its details (Morton et al., 2003).

Both connectionist and dynamic systems models have played an important role in helping to understand this diversity of findings. In this chapter, we focus on case studies of connectionist models of perseveration, for understanding infant reaching behavior (Munakata, 1998; Stedron et al., 2005) and children's rule-guided behavior (Morton & Munakata, 2002). These models demonstrate how general connectionist principles: (1) provide a unified framework for understanding perseveration across ages, task paradigms, and response modalities; (2) capture important details of children's performance; and (3) lead to novel and testable predictions.

We compare and contrast these contributions from connectionist models with those from dynamic systems models of the A-not-B task, which place greater emphasis on the details of visually guided reaching. Such details may be essential for understanding nuances of infant performance on the A-not-B task, but the more general principles may have greater utility for understanding perseverative behavior across the life span.

The connectionist models we discuss focus on the nature of the representations supporting behavior, and specifically the importance of graded and distinct types of representations. First, these models capture the assumption that knowledge is graded in nature rather than all-or-nothing, and can be weaker or stronger depending on a variety of factors, including an individual's state of development, experience with a task, and the amount of support from the environment (McClelland, 1989; Munakata, 2001a; Sitton et al., 2000). Second, these models capture the assumption that knowledge can take multiple forms rather than being monolithic. One important distinction is between active versus latent representations (Cohen, Dunbar, & McClelland, 1990; Munakata, Morton, & Stedron, 2003). Active representations take the form of sustained neuronal firing, which can be communicated to other processing regions. Such active representations are thought to tap prefrontal cortical regions and develop relatively late. Latent representations take the form of changes in connections between neurons, which influence subsequent processing of stimuli by those neurons without otherwise being accessible to other processing regions. Such latent representations are thought to tap posterior cortical and subcortical areas and develop relatively early. This active-latent distinction is captured naturally in connectionist models. Active representations take the form of sustained activation of processing units, and latent representations take the form of changes in the strength of connection weights between processing units. Both kinds of representations can vary continuously in their strength, capturing the notion of gradedness. In what follows, we discuss models that demonstrate how such graded, active and latent representations provide a unified framework for understanding both general and specific aspects of children's perseveration.

2. Case studies: Connectionist models of infant perseveration in A-not-B tasks

We focus first on a model of the Piagetian A-not-B task with hidden objects (Munakata, 1998), and then briefly describe extensions of this model to other variants of the A-not-B task (Stedron et al., 2005).

2.1 Architecture, stimuli, and learning algorithm

The model (Figure 1) includes three input layers, a hidden layer, and two output layers. Input layers allow the network to encode information about task-related events including: (1) location of events, with a separate unit encoding each of three different locations (A, B, and C); (2) covers presented in the events, with one unit encoding the default cover (C1) and a second unit encoding an alternate (C2); and (3) toys presented in the events, with one unit encoding a default toy (T1) and a second unit encoding an alternate (T2). Hidden and output layers each contain three units that correspond with the three possible hiding locations. The two output layers allow the network to indicate separate gazing and reaching responses and are essentially the same except that units in the gazing layer respond to every event, whereas units in the reaching layer respond only to events that are within reaching distance. This difference captures the fact that during the actual A-not-B task, infants are able to gaze to particular locations throughout the entire procedure but can reach only at select times.

Recurrent and feedforward connections link the units of the model and allow the network to form active and latent representations of its experience. Feedforward connectivity includes connections between each of the three input layers and the hidden layer, as well as connections between the hidden layer and the two output layers. Consistent with infant behavior, this initial feedforward connectivity includes a bias to respond appropriately to location information such that events that occur at A lead the network to represent and look to A. This bias is instantiated as a set of connection weights such that connections between corresponding units (e.g., A input unit to A hidden unit) have larger weights than connections between non-corresponding units (e.g., A input unit to B or C hidden units. These feedforward connection weights change with experience according to a Hebbian learning rule, such that units that fire together (i.e., show correlated activity) wire together. For example, input at the A location leads to simultaneous activity in the A units of the internal representation and gaze layers. Consequently, with repeated experience representing and gazing at a toy presented at A, connections between A input, hidden, and output units are further strengthened and establish a latent representation of A. This latent representation in turn causes the network to gaze preferentially to A on subsequent trials.

In addition to feedforward connections, hidden and output units also make connections to themselves. These so-called recurrent connections allow the model to form active representations of its experience—essentially self-sustaining activity of select units of the model that can be maintained in the absence of continued input from the input layers. The strength of recurrent connections is varied to simulate changes in active memory hypothesized to occur with development.

The task is administered as a series of trials, with each trial consisting of four segments that are presented as patterns of activity over the input units and that correspond directly with the procedure administered to infants. These segments include:

Start: covers sit in place on the apparatus until the experiments draws the infant’s attention to a particular location;

Presentation: Experimenter presents a toy at a particular location;

Delay: The toy has been hidden, and the apparatus sits with the lids in place;

Choice: Experimenter pushes the apparatus toward the infant until the locations are within reach. Reaching is permitted only in this segment of the trial.

To illustrate, Figure 2 displays patterns of input corresponding to an A trial. Inputs in each segment correspond with the visible aspects of the display, with more salient aspects of the display producing more activity. At the Start, there is only a small amount of activation for each location and the cover, reflecting the fact that at this point, the experimenter has not drawn infants' attention to the apparatus. At Presentation, the toy and A location units are highly active, reflecting the presentation of the toy at the A location. During Delay, the cover and the location units are moderately active, as these aspects of the display remain visible to the infant, but the toy unit is not active, as the toy is hidden. Finally, at the Choice point when the apparatus is pushed within reach of the infant, activity in all location units increases to reflect the increased salience of the three locations.

2.2 Performance

The model provides a comprehensive framework for understanding general aspects of infant performance in the A-not-B task, including perseveration, age-related change, and dissociation, as well as more detailed aspects of children's performance, and also generates counter-intuitive predictions that have been supported in subsequent studies.

2.2.1 General Aspects of Infant Performance

2.2.1.1 Perseveration

Figure 3 depicts the network’s gazing and reaching performance on A trials and B trials as a function of age, or recurrence strength. On A trials, the model gazes and reaches correctly on A trials independent of age (i.e., recurrence strength). By contrast, when recurrence is low, the model perseverates strongly, gazing and reaching correctly on only 15% of trials. Why does the model search correctly on A trial but perseverate on B trials?

When recurrence is low, the model is unable to actively maintain information about prior events at A, but gazes and searches correctly at A by virtue of changes in latent weights that favor A. Figure 4, which shows an A trial of a young model with relatively weak recurrent connections, illustrates this idea. Because the model receives several practice trials at the A location, the strength of connections between A units changes due to Hebbian learning, and leads to a relatively strong latent bias to respond to A. Therefore, in the presentation period, when a toy is presented at A, the model represents that something is happening at A and gazes at A[2]. Even during the delay period though, when the model is given identical input to the A, B, and C locations, the model has no difficulty representing and gazing to the A location even though it cannot actively maintain an internal representation of A. This is because feedforward connections are so strongly biased in favor of A. Similarly in the choice period, even though the model receives equal input from the A, B, and C input units, it still actively represents A, gazes at A and reaches to A. This would correspond to the following account of very young infants: while they are not terribly good at actively remembering that a toy was hidden at A, they have built up such a strong bias to looking at A, representing an object at A and reaching to A, they're very inclined to go in that direction.

The model's ability to maintain activity is however critical for B trial performance, as the model needs to maintain a representation of B in order to overcome a latent bias to respond to A. This is illustrated in Figure 5, which shows the same model performing a B trial. At the beginning of the trial, when the toy is presented at B, the model represents and gazes toward B, just as infants gaze correctly when the toy is being hidden. However, in the delay period, A, B, and C locations are equally active, and the model has a strong bias to A and a weak ability to actively maintain new information. Consequently, the model's internal representation of B begins to fade away. By the time the choice period comes, the active representation of B has faded away almost entirely

---

[2] The model does not reach to A because the reach output layer is not allowed to update until the choice period

and the model is representing A, and perseveratively reaching and gazing to A. Thus, in the face of a strong bias to A, the model's weak ability to actively maintain information about recent events at B is inadequate for representing B, and reaching and gazing to B, and the model perseverates by looking and reaching to A.

2.2.1.2 Effects of Age

Performance in the A not B task is related to age, with older and younger infant performing comparably on A trials but older infants performing better than younger infants on B trials given a particular delay (Marcovitch & Zelazo, 1999; Wellman, 1986). The model's performance on B trials improves incrementally with age, simulated as an increase in the strength of recurrent weights. This is illustrated in Figure 6, which shows an older model performing a B trial. The critical difference in performance begins to emerge in the delay period of the trial when the network is receiving equal input to the A, B, and C input units. Whereas before, when recurrent weights were low, the model's internal representation of B began to fade away, now the model's internal representation of B is strongly maintained, despite a strong bias to represent A. The model's strong active representation of B carries though to the choice period, and therefore, even in the face of a strong bias to A, the model correctly represents, reaches, and gazes to B.

2.2.1.3 Dissociation

Although infants' memory for hidden toys seems poor when tested by means of reaching—most infants perseverate by reaching to A when a toy is hidden at B—their memory seems considerably better when tested by means of gazing. For example, infants that reach perseveratively to A in the A-not-B task nevertheless look longer when a toy hidden at B is revealed at A than when it is revealed at B (Ahmed & Ruffman, 1998).

Similarly, infants are more likely to gaze correctly than reach correctly to B on B trials (Hofstadter & Reznick, 1996). Thus, infants exhibit dissociations in their behavior, appearing more knowledgeable when tested one way than another. The model exhibits the same dissociation in its performance by showing strong perseveration in its reaching system but more accurate performance in its gaze system. Evidence of this dissociation can be observed in Figure 3 which shows the gazing and reaching performance of the model during A trials and B trials at various levels of recurrence. A moderately old model with recurrence of 0.4, for example, reaches correctly on approximately 60% of B trials but gazes correctly on over 80% of B trials. Why does the model show this dissociation?

The discrepancy arises because the underlying representation of B is adequately strong to drive correct gazing but not correct reaching. This effect is illustrated in Figure 7, which shows activity in a moderately old model (recurrence = 0.4) during the presentation, delay, and choice periods of a B trial. In the presentation period, a toy is presented at B, and the model represents and gazes to B. During the delay period, when A, B, and C input units are equally active, moderately strong recurrent connections allow the model to actively maintain a representation of B and correctly gaze to B, but are not sufficiently strong to completely override the model’s strong latent representation of A. Thus, the internal representation is gradually pulled back toward A. When the choice period begins and the reach system is able to update, A is now more strongly represented than B due to the influence of the model’s latent bias to A, and the model reaches incorrectly to A. This weak representation of B is, however, sufficient to drive a correct gaze to B. This is because the gazing system has updated throughout the delay period and allowed activation of the B gaze unit. Therefore, even though inputs to the gaze system in the choice period favor A, the gaze system responds correctly. This parallels what is observed with infants, who may show earlier success in gaze/expecation variants of the A-not-B task because they can constantly update their gazing and their expectations about B during the presentation, delay, and choice periods of B trials which would allow them to more effectively counter perseverative tendencies in their gaze responses. In contrast, infants can only reach at the choice point, by which time their representations have become more susceptible to perseverative biases. Therefore, in the model, even though inputs to the gaze system in the choice period favor A, the gaze system responds correctly. Dissociations of this kind illustrate a central idea of graded representation accounts, namely that a single shared representation (in this case a weak representation of B) can be sufficient to drive correct responses in one system (i.e., the gaze system), and yet insufficient to drive correct responses in a different system (i.e., the reaching system).

2.2.2 Detailed Aspects of Infants' Performance

In addition to capturing general aspects of performance like perseveration, age-related change, and dissociation, the model can also account for a number of detailed aspects of infant performance in the A-not-B task. These include the effects of the number of hiding locations, the distinctiveness of the covers, the use of covers alone versus covers and toys, and merely observing toys being hidden at A rather than reaching for them. The following discussion focuses on two of these effects, the effects of covers alone and observation alone.

2.2.2.1 Effect of covers alone and observation alone

One finding that seems to challenge the idea that perseveration in the A-not-B task relates to memory of hidden objects is that infants perseverate in a covers-only version of the task that has no hidden objects. In this cover-cover version, infants watch and then reach for a cover that is lifted and waved at A and then perseverate by reaching back to A when a cover is subsequently lifted and waved at B (Munakata, 1997; Smith, Thelen, Titzer, & McLin, 1999). These findings suggest that motor history rather than memory for hidden objects is critical for eliciting perseverative search (Smith et al., 1999).

However, although the covers remain visible throughout cover-only versions of the task, memory is still required to determine which of the two covers was waved. Moreover, hidden toys do in fact influence infant performance in variants of the covers-only paradigm. For example, infants perseverate less in a cover-toy version of the task in which they reach for a cover waved at A and then watch a toy hidden and covered at B than they do in the cover-cover version (Munakata, 1997).

The model captures all of these effects, in terms of levels of input activity (higher for toy and cover than for cover alone) and resulting changes in latent representations. As seen in Figure 8, which shows performance on two types of A trials (cover-only ("cover") and cover + toy ("toy")) and the four types of B trials (i.e., two types of B trials that each follow two types of A trials). The model performs accurately on both the cover and toy A trials at all levels of recurrence. The network has a bias to respond to A based on weight changes that occur in pre-training and A trials. Consequently differences in the level of input activity for toy and cover events and ability to maintain information have little effect on performance. The model perseverates strongly and comparably on B-trials of toy-toy and cover-cover versions, as strong activation of B in the hidden layer competes against a strong latent representation of A in the former, and weak activation of

B competes against a weak latent representation of A in the latter. The model performs better in the cover-toy version because in this condition, strong activation of B competes against a weak latent representation of A. Thus, for these three conditions, the model's B trial performance reflects the relative strength of active and latent representations.

Somewhat counter-intuitively, performance on toy-cover B trials is comparable instead of worse than performance in the toy-toy and cover-cover versions, in both the network (Munakata, 1998) and infants (Munakata, 1997). The reason is that A trials and B trials are more similar to one another in toy-toy and cover-cover versions than in the toy-cover version. As a result, B trials "remind" the network of A trials more in toy-toy and cover-cover versions than in the toy-cover version. For example, on B trials, the toy in the toy-toy version provides strong activation to B *and* A, since the toy is presented at B but has also been previously associated with A. In effect, the model is more strongly reminded of A in the toy-toy version than in the toy-cover version, leading to comparable levels of preservation in both versions.

Further evidence that motor history alone cannot adequately account for infant perseveration in the A-not-B task is that infants perseverate after merely observing events at A (Butterworth, 1974; Diamond, 1983; Evans, 1973; Spencer & Schutte, 2004). The model showed the same behavior, due to latent biases to attend to A that were formed after merely observing events at A (Munakata, 1998). To simulate the absence of reaching during A trials, inputs corresponding to regular A trials were administered to a standard network but the learning rate for connections between hidden and reaching units were set to 0. Consequently, these connection weights did not change across A trials, which effectively simulated an absence of experience reaching to A. The model accurately simulated infant performance by perseverating strongly on B trials. The model perseverated in this way because although it had never reached to A, it nevertheless learned to attend to A. Repeated experience watching a toy hidden at A strengthened input to hidden unit weights, causing the model to preferentially represent A on its hidden units. Although hidden to reaching units weights did not change, preferential attention to A was sufficient to drive preservation on B trials. The model therefore provides a comprehensive account of specific effects involving various combinations of hidden toys and reaching experience, and clarifies how motor history is just one of many factors influencing infant perseveration in the A-not-B task.

2.2.3 Predictions

The model generates a number of novel and testable predictions about infant performance in the A-not-B task. One counterintuitive prediction of particular relevance for current purposes is that infants should show a developmental U-shaped curve in their A-not-B performance. This is illustrated in Figure 9, which shows B-trial performance at various levels of delay and recurrence. For delays greater than 1, as recurrence strength increases from the lowest to more moderate values, B trial performance actually gets worse before getting better. This pattern of performance is related to the fact that during A trials, the models latent representation of A gets stronger partly as a function of how active A hidden units are, which is in turn a function of recurrence strength. The parallel in infants would be that really young infants do not think very much at all about A after the toy is hidden, and therefore build up less of a bias in that direction compared with older infants, who continue to maintain information about A across the delay. Thus the model generates the novel prediction that young infants would actually do better on B trials than slightly older infants since they do not maintain information about A across the delay period. This prediction has been supported using a lids-only variant of the A-not-B task administered to 5-, 6-, 7-, and 8-month-old infants (Clearfield et al., 1999). Consistent with the predictions of the model, B-trial performance actually declines between 5- and 8-months and then improves thereafter.

2.3 Perseveration in the absence of apparent memory demands

One criticism of the idea that perseveration is linked to under-developed active memory mechanisms is that perseveration can occur in instances in which all of the information required to solve a problem is immediately available. After reorganizing the office, for example, it is not uncommon to reach for the telephone in its old location even though its new location is in clear view. Young infants show a similar form of perseveration in a simple task depicted in Figure 10. An attractive toy is placed on an “A” towel, while a similar toy is placed behind a “B” towel. Although initially infants successfully retrieve the toy by pulling the “A” towel, when the position of the toys is reversed, infants perseverate by continuing to pull the A towel (Aguiar & Baillargeon, 2000). Why do infants perseverate when the solution to the problem is in full view?

Although these findings appear to challenge the idea that infant perseveration is linked to active memory, active maintenance mechanisms may support not only memory, but also attention to task-relevant information in the environment (Chafee & Goldman-Rakic, 2000; De Fockert, Rees, Frith, & Lavie, 2001; Desimone & Duncan, 1995; Fuster, Bauer, & Jervey, 1985).

This hypothesis was tested using an extension of the active-latent neural network model of infant performance in the A-not-B task (Stedron et al., 2005). The main difference between the basic A-not-B model and the extension for the towel-pulling task was in the input representations; the towel-pulling model inputs included information about whether toys were on or attached to cloths (Figure 10).

2.3.1 Performance

As seen in Figure 11, the model simulated general aspects of infant performance, including accurate performance on A trials, perseveration on B trials, and age-related performance improvements, as well as more detailed aspects of performance related to the earlier mastery of a toy-attached version of the task.

General aspects of infant performance

2.3.1.1 Accurate A-trial performance across ages

The model performs accurately on A trials across all ages (i.e., levels of recurrence) because of the greater input activation associated with the correct side. This greater activation to the A side propagates through feedforward connections and allows A units in the hidden, PFC, and output layers to win the competition for representation over the B units. Successful reach to A subsequently strengthens these feedforward connections, thereby establishing latent bias to attend and reach to A.

2.3.1.2 Perseveration

On B trials, at low levels of recurrence, the model perseverates strongly, inappropriately pulling the A towel. On B trials, greater input activation is again associated with the correct side, but this activation has to compete against a latent bias to attend and respond to A. When recurrence strength is low, the representation of the B location in PFC units is insufficiently strong to override this bias, and the model perseverates by reaching to A, even though all the information required to make the correct response is available to the network on the input layers.

2.3.1.3 Age-related performance improvements

Hypothesized age-related changes in active maintenance mechanisms were simulated by increasing the strength of recurrent connections PFC units made to themselves. Stronger recurrent connections amplified the representation of B in the PFC layer, leading to greater top-down support for the B unit in the hidden layer, and improved B trial performance. Thus, the models offer an existence proof that active maintenance can provide a common mechanism for improving memory for hidden objects as well as attention to relevant information in the environment.

2.3.1.4 Detailed aspects of infant performance

By 9-months, infants succeed in a toy-attached version of the towel pulling task in which the toy is attached to one of the two towels. However, it is only by 11-months that they succeed in a very similar toy-unattached version of the task in which the toy is placed on but not attached to one of the two towels (Aguiar & Baillergeon, 2000). To allow for the administration of these two versions of the task, the input layers included units that indicated whether the toy was merely on or was also attached to a particular towel. Consistent with infant performance, the model showed earlier mastery of the toy-attached than unattached version of the task (see Figure 11). Why did the model show this specific effect?

The model mastered the toy-attached version of the task earlier than the toy-unattached version because there is a greater difference in input activation between the correct and incorrect side in the toy-attached version (4 versus 2) than in the toy-unattached version (3 versus 2). Consequently, the correct location “stands out” more in the toy-attached than in the unattached version and is mastered earlier (i.e., at lower recurrence). The simulation results suggest that the toy-attached version is easier because additional features of the task make the correct location more salient to infants and help support attention to that location despite under-developed active maintenance mechanisms. When fewer features highlight the correct location, as is the case in the toy-unattached version, infants have to rely more on active maintenance mechanisms to overcome bias and attend to the relevant information in the environment.

2.3.1.5 Predictions

The model leads to the novel prediction that infants should show a developmental U-shaped curve in their response times on B-trials (Figure 11). In particular, the model predicts that young infants who perseverate strongly to A and older infants that reach correctly to B should show shorter response times than infants of an intermediate age. Differences in response times across different levels of recurrence reflect the degree of imbalance between the PFC unit's representation of B and the latent bias to A. When this imbalance is great, either because PFC representations are very weak (as in young networks) or very strong (as in old networks), the network requires relatively few processing cycles to settle and response time are short. However, when this imbalance is small, the network requires more cycles to settle and response times increase.

3. Case studies: Connectionist models of perseveration in children's rule-guided behavior

The general principles of graded, active and latent representations also provide a framework for understanding general and specific aspects of perseveration in children's rule-guided behavior. Children often persist in using old rules for sorting cards (e.g., sort by shape) or judging speakers' feelings (e.g., judge by message content) when asked to switch and use new rules (e.g., sort by color/judge by tone of voice), despite apparent knowledge of the new rules. Although these tasks are similar in structure, children switch flexibly in the card-sorting task by 4- or 5-years of age, but perseverate in the speech-interpretation task through 7-years (Morton et al., 2003; Zelazo citation for 4-5 years in DCCS).We briefly describe a model which captures general aspects of children's card-sort performance and specific differences between card-sort and speech interpretation performance, and which leads to novel and testable predictions.

3.1 Architecture and task-administration

The card-sort model, shown in Figure 12, features separate pathways that connect visual features of the test cards (i.e., color and shape) with potential responses. Additional input layers (verbal features, and rule) allow the network to encode sorting rules, descriptions of the cards, and verbal queries. As in previous models, latent representations take the form of changes in feedforward connections weights, whereas active representations take the form of sustained activity in PFC units made possible by recurrent connections. The task is administered as patterns of input delivered to the input units and includes pre-switch trials in which the network sorts cards by one rule (e.g., color; see Figure 12) followed by post-switch trials in which the sorting rule changes. The network is then asked questions about the post-switch rules.

3.2 Perseveration and age-related change

Figure 13 shows the performance of the network across pre-switch and post-switch performance as a function of age (i.e., recurrence). At all ages, models sort accurately in pre-switch trials and establish a strong latent representation of the pre-switch rule, but young models perseverate in post-switch trials. When recurrence strength is low, PFC units only maintain a weak representation of the post-switch rule and can only minimally bias competition among units in the internal representation layer. As a result, they fail to override a strong latent bias to sort by the pre-switch rule. Older models with stronger recurrent connections robustly maintain an active representation of the post-switch rule, and more effectively override latent biases. As a result, older models are more likely to switch appropriately in post-switch trials. The model therefore simulates critical aspects of children's card-sorting performance, including accurate pre-switch performance across age and incremental improvement in post-switch performance with increasing age.

3.3 Dissociations

Like 3-year-olds, young models also show apparent knowledge-action dissociations by perseverating strongly on post-switch trials while responding correctly to simple questions about the post-switch rules (Zelazo et al., 1996). This apparent dissociation can be understood in terms of differences in the amount of conflict contained in knowledge and action measures (Morton & Munakata, 2002a, 2002b; Munakata & Yerys, 2001). Standard knowledge questions pertain to only a single feature of the cards (e.g., "When sorting by shape, where do trucks go?") and therefore contain no conflict. Thus, young networks that form weak active representations of the sorting rules can nevertheless answer these questions correctly. By contrast, the sorting task (i.e., the action measure) presents conflicting shape and color information (i.e., cards that would be sorted one way if sorted by shape, but a different way if sorted by color). Thus, only older networks that can form strong active representations of the new shape rule can resolve the conflict between shape and color in post-switch trials. Thus, knowledge-action dissociations in the card-sort task may be more apparent than real, in that they reflect differences in the amount of conflict contained in measures of knowledge and action.

Consistent with this account, apparent dissociations disappear when differences in conflict across measures of knowledge and action are controlled (Morton & Munakata, 2002a; Munakata & Yerys, 2001). In one study, 3-year-olds were presented with the standard card-sort task and were asked different questions about the post-switch rules. Simple questions contained no conflict (e.g., “In the shape game, where do the trucks go?”) whereas conflict questions contained conflicting shape and color information (e.g., “In the shape game, where do red trucks go?”). Children’s knowledge of the post-switch rules appeared to outstrip their sorting performance when knowledge was assessed by the simple knowledge question, but this apparent dissociation disappeared when knowledge was assessed by means of the conflict question (Munakata & Yerys, 2001). These processes underlying apparent dissociations may be general across the lifespan (Morton & Munakata, in preparation). Adults show an apparent dissociation in their reaction times on card-sorting tasks, by responding relatively slowly when switching under conditions of conflict but relatively quickly when answering non-conflict questions about the new rule. However, consistent with the graded representations account, these apparent dissociations disappeared when knowledge and action measures contained the same amount of conflict.

3.4 Detailed aspects of performance

Although the card-sort and speech interpretation tasks both involve switching, children master card-sorting long before the SIT. One potentially important difference is that children have a strong prepotent bias to judge a speaker's feelings from what she says rather than how she speaks (Friend, 2000; Morton & Trehub, 2001) but no strong prepotent bias to sort by color or shape (Zelazo et al., 1996). Consequently, the degree of conflict operative in post-switch trials of the SIT is greater than in the card-sort task and may demand a more robust active representation of post-switch rules. Consistent with this hypothesis, a card-sort and a SIT model that were identical except for the definitions of the constituent units and the presence/absence of a pre-existing latent bias for one sorting dimension showed the same general performance patterns (i.e., perseveration, age-related change, dissociation), except that the SIT model required stronger recurrence to achieve post-switch performance comparable to that of the card-sort model (Morton & Munakata, 2002). Thus, the active-latent distinction provides a parsimonious framework for understanding differences in when children master formally similar tasks.

3.5 Predictions

The model makes the novel and counter-intuitive prediction that guided practice may be more effective in eliciting flexible behavior in preschoolers than direct instruction. The prediction stems from the fact that young models have difficulty actively maintaining a strong representation of a new sorting rule due to weak levels of recurrence in PFC units. Consequently, instructions are relatively ineffective in causing young networks to switch sorting criteria. By contrast, guided practice in the use of the new rule is effective in causing young networks to switch because such experience increases the strength of the model’s latent representation of the new rule. This prediction has been confirmed (Brace, Morton, & Munakata, in press). In a modified version of the card sort task, 3-year-olds sorted cards by one rule and then received either guided practice in the use of a new rule or instructions to switch to a new rule. Consistent with the predictions of the model, significantly more children switched to the new rule given guided practice alone than instructions alone.

4: Connectionism and Dynamic Systems Theory: Discussion of cross-cutting conceptual issues

The neural network model case studies described in this chapter provide a unified framework for understanding perseveration and flexibility -- both more general aspects observed across ages, task paradigms, and response modalities, and important details specific to particular tasks. In this section, we consider the implications of these case studies for broader issues about core concepts in cognition, claims about what develops, time scales and levels of explanation, and data constraints. We indicate where answers from these case studies may not be representative of neural network approaches more generally, and draw comparisons and contrasts with dynamic systems approaches throughout (see our related discussion in Munakata, 2006; Munakata & McClelland, 2003; Munakata, Sahni, & Yerys, 2001).

4.1 Core concepts

Our neural network case studies capture many assumptions about core concepts of cognition that are representative of neural network approaches more generally and highly overlapping with dynamic systems approaches. However, the focus of these case studies on the distinction between active and latent representations may be unique.

Common aspects of neural network and dynamic systems approaches include a dynamic, highly interactive view of behavior as reflecting a confluence of multiple, graded factors that interact in non-linear ways (Elman et al., 1996; Kelso, 1995; McClelland & Rumelhart, 1986; Smith & Thelen, 2003). One example from our A-not-B neural network case study (Munakata, 1998) concerned the effects of changing the strength of recurrent connections. Very young models (with very low levels of recurrence) could actively maintain information about a recent hiding location for only a brief period of time, but actually performed better on B trials than slightly older models. As discussed earlier, this counterintuitive effect occurred because recurrence strength affected not only how well the network could override latent biases on B trials, but also how strongly latent biases to A were established during A trials. Thus, changes in recurrence could not be understood in isolation at any given point in time, but needed to be considered in terms of their influence on the dynamics of the entire model across time.

Multi-causality and interactivity are also at the heart of dynamic systems accounts, such as a dynamic field model of infant A-not-B performance (Schoner and Spencer, 2003). Central to this approach is the concept of stability, which refers to a model's ability to maintain a pattern of activation corresponding to a previously viewed object, for example. The ability to enter into self-sustaining stable states has important implications for A-not-B performance, because it allows models to establish “representational states” (Spencer & Schöner, 2003) and carry forward activation related to prior events that are no longer perceptually available so that these events can influence subsequent processing. This capacity is not determined by one factor but arises through the confluence of multiple factors, including the nature of inputs and the strength of excitatory and inhibitory intra-field interactions.

Another parallel in neural network and dynamic systems approaches to cognition is that no clear distinction is made between processing and knowledge in these approaches. That is, knowledge is not treated as separable from processes that operate on knowledge; instead, knowledge and processing are intrinsically related. One example from the A-not-B and card-sorting case studies (Morton & Munakata, 2002a; Munakata, 1998) concerned behavioral dissociations, where infants look to the correct hiding location for a toy but reach perseveratively, or children answer questions correctly about a new card-sorting rule but sort perseveratively. Standard accounts explained these dissociations in terms of a disconnect between children's knowledge (viewed as intact) and their processing (e.g., inhibitory processing, viewed as underdeveloped) (Ahmed & Ruffman, 1997; Diamond, 1985; Hofstadter & Reznick, 1996). In contrast, the case study models demonstrated how such dissociations could arise through highly interactive systems, as a result of the strength of representations formed during processing. Similarly, dynamic systems accounts do not draw a strong distinction between knowledge and processing. For example, dynamic systems accounts emphasize that the A-not-B error is not about what infants have or don't have as enduring concepts or traits, but about what infants do and what they have done (Thelen et al., 2001). Thus, knowledge emerges through the course of infant processing, rather than taking the form of reified entities in the infant's head.

Our case studies point to one important potential difference between neural network and dynamic systems approaches to cognition: the focus on distinct types of representations. The distinction between active and latent representations is central to neural network accounts of perseveration and flexibility. This distinction is motivated by behavioral, physiological, and computational data (reviewed in Kharitonova, Chien, Colunga, & Munakata, 2006; Ashby & Todd, 2005; Bunge, Kahn, Wallis, Miller, & Wagner, 2003; Casey, Durston, & Fossella, 2001; Frank, 2005; Jog, Kubota & Graybiel, 1999; Miller & Desimone, 1994; Miller, Erickson, & Desimone, 1996; Patalano, Smith, Jonides & Koepee, 2001; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005; Wallis, Anderson, & Miller, 2001). The active-latent distinction has also led to novel, counterintuitive predictions that have since been confirmed. In contrast, dynamic systems approaches have not emphasized such representational distinctions; instead, various factors such as environmental stimuli and memories for prior events may be summed as equivalent inputs to a dynamic field (Spencer & Schöner, 2003; Thelen et al., 2001).

4.2 Development -- what changes?

Both neural network and dynamic systems approaches have emphasized the importance of multiple interacting factors and nonlinear, emergent effects in development (e.g., Elman et al., 1996; Smith & Thelen, 2003). For example, neural network models have been used to investigate how small changes in the starting state of a system (e.g., in the firing rates of processing units) could interact with other processes to lead to specializations of different brain regions (O'Reilly & Johnson, 2002), or to developmental disorders that appear to be highly modular (Joanisse & Seidenberg, 2003; Thomas & Karmiloff-Smith, 2003). Similarly, dynamic systems models have been used to investigate how small changes in the visual environment could lead infants to very different trajectories in perceptual state space, in ways that might explain their patterns of looking in habituation studies (Thelen & Smith, 1994).

However, neural network approaches have typically focused more on learning as an important aspect of developmental change, with emphasis on the mechanisms underlying such learning. Dynamic systems approaches have instead focused on the importance of control variables that change with age, without an explanation of the mechanisms that lead to such changes. For example, all of our case studies focused on Hebbian learning mechanisms for explaining how habitual behaviors could build across development. These case studies also linked the reduction of perseveration across development with age-related advances in active maintenance mechanisms. Although these advances were implemented by hand in these case studies, through increases to recurrent connections, other neural network models have focused on understanding how such active maintenance abilities could develop through the course of learning (e.g., Munakata et al., 1997; Rougier et al, 2005).

In contrast, in dynamic systems models of the A-not-B task, development is simulated in terms of changes to a parameters without explanation of the mechanisms driving such changes. For example, in one model development was simulated through changes to a parameter, h, the resting level of a dynamic field (Thelen et al., 2001). In another model, development was simulated through changes in a local excitation/lateral inhibition function (Spencer & Schöner, 2003). This function determined the extent to which an activated site in the field increased activation in neighboring sites and suppressed activity in distant sites. Fields that were weakly interactive simply reflected current inputs, as they had difficulty establishing a self-sustaining stability. By contrast, fields that were strongly interactive could establish self-sustaining stabilities or attractors and could therefore maintain a representation of prior input, even when the input was no longer delivered to the field. This concept of stability is very similar to active maintenance in our case studies, but neural network approaches have focused more on learning mechanisms that explain developmental changes in these processes

4.3 Time scales and levels of explanation

Both neural network and dynamic systems approaches have the potential to address change over very broad time scales, from milliseconds to months. For example, our case study A-not-B model led to predictions about U-shaped developmental curves on the order of months. These predictions have since been confirmed, through studies motivated by a dynamic systems approach (Clearfield et al., 1999). Similarly, both our case study A-not-B model and dynamic systems approaches can explain effects on a much smaller time scale, for example, trial-by-trial effects of toys versus covers on A-not-B performance. Although dynamic systems discussions of the A-not-B task have focused on the precise dynamics of infant reaching (e.g., Diedrich et al., 2000) more than neural network approaches, dynamic systems models have not always addressed such real-time reaching dynamics (Spencer & Schöner, 2003; Thelen et al., 2001). Moreover, neural network models have been used to explore real-time dynamics at the millisecond level in other domains (e.g., Huber & O'Reilly, 2003; Usher & McClelland, 2001). Thus, we do not see inherent differences between neural network and dynamic systems approaches in their potential to address change over multiple time scales in an integrated framework.

Similarly, both neural network and dynamic systems approaches have the potential to address multiple levels of explanation. For example, as we have emphasized throughout this chapter, our case study neural network models explain both very general aspects of perseveration observed across ages, paradigms, and response modalities, as well as more detailed aspects of perseveration that are specific to particular tasks. In addition, abstract neural network models have been used to investigate a range of behaviors (e.g., Arbib, 2002; Elman et al., 1996; McClelland & Rumelhart, 1986), while more detailed neural network models have been used to investigate neurobiological mechanisms (e.g., Arbib, 2002; O'Reilly & Munakata, 2000). Again, although dynamic systems discussions of the A-not-B task have focused on the importance of embodied cognition and the details of perceptual motor systems (e.g., Diedrich et al., 2000; Thelen et al., 2001) , dynamic systems models have not always incorporated such details at this level, for example, regarding the kinematics of reaching (Spencer & Schöner, 2003; Thelen et al., 2001). Moreover, neural network models have been used to investigate the detailed kinematics of learning to reach (e.g., Berthier, Rosenstein, & Barto, 2006). Thus, again, we do not see inherent differences between neural network and dynamic systems approaches in their potential to address multiple levels of explanation, each of which is important. Higher level explanations and general principles may be critical for understanding general aspects of perseveration, such as the fact that perseveration is observed across response modalities, and even after infants have merely observed hidings at A in the A-not-B task (Butterworth, 1974; Diamond, 1983; Evans, 1973; Spencer & Schutte, 2004). More detailed models of visually guided reaching may be needed to explain effects of body orientation and arm-weights on infant performance on the A-not-B task (Smith & Thelen, 2003).

4.4 Constraints from empirical data

Both neural network and dynamic systems models are constrained by various types of empirical data: behavioral, neurobiological, qualitative, quantitative, etc. Neural network models may have a relative strength in incorporating constraints from neurobiology. Dynamic Field Theory models establish stabilities or attractor states through local excitatory and lateral inhibitory intra-field connections, a well-known “brain-like” interactive mechanism. However, neural network models also capture these more abstract aspects of neurobiological processing, while providing a more transparent mapping between the computational elements of units and weights and the corresponding biological elements of neurons and synapses, and between network architectures and learning algorithms and the corresponding neural systems and processes (e.g. Frank, 2005; O'Reilly & Rudy, 2001).

5. Conclusions

As many others have argued (e.g., Elman, 2003; Spencer & Schoner, 2003; Thelen & Bates, 2003), neural network and dynamic systems approaches to development show both important similarities and differences. We believe that the similarities highlight essential characteristics of development, such as nonlinear, emergent effects that arise in highly interactive systems in which knowledge and processing are intertwined. These characteristics are ignored or even challenged by alternative approaches to development, such as some modular or nativist characterizations (e.g., Baron-Cohen, 1998; Fodor, 1983; Leslie, 1992). We believe that many of the differences between neural network and dynamic systems approaches, for example in time scales and levels of explanation, reflect the foci of particular models or modelers rather than inherent differences between these frameworks. Each approach alone may serve to provide a unified framework for understanding behavior and development, as demonstrated by the case studies of neural network models of perseveration discussed here. However, the most complete accounts and understanding may emerge through a synergy of the complementary strengths of neural network and dynamic systems approaches.

References

Aguiar, A., & Baillargeon, R. (2000). Perseveration and problem solving in infancy. In H. Reese (Ed.), *Advances in Child Development and Behavior* (vol. 27, pp. 135-180). New York: Academic Press.

Ahmed, A., & Ruffman, T. (1998). Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task? *Developmental Psychology, 34*, 441-453.

Arbib, M. (2002). (Ed.), The Handbook of Brain Theory and Neural Networks (2nd Ed.), Cambridge, MA: MIT Press.

Ashby, F. G. & Todd, M. W. (2005). Human category learning. *Annual Review of Psychology, 56*, 149-178.

Baron-Cohen, S. (1998). Modularity in developmental cognitive neuropsychology: Evidence from autism and Gilles de la Tourette syndrome. In J. A. Burack, R. M. Hodapp, & E. Zigler (Eds.), Handbook of Mental Retardation and Development (pp. 334-348). Cambridge University Press, Cambridge, UK.

Bremner, J. G. (1978). Spatial errors made by infants: Inadequate spatial cues or evidence of egocentrism? *British Journal of Psychology, 69*, 77-84.

Bunge. S. A., Kahn, I. Wallis, J. D., Miller, E. K. & Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *Journal of Neurophysiology, 90*, 3419-3428.

Butterworth, G. (1974). *The development of the object concept in human infants.* Unpublished doctoral dissertation, University of Oxford.

Casey, B. J., Durston S. & Fossella J. A. (2001). Evidence for a mechanistic model of cognitive control. *Clinical Neuroscience Research, 1,* 267-282.

Chafee, M. V., & Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *Journal of Neurophysiology, 83*, 1550.

Clearfield, M. W., Thelen, E., & Smith, L. B. (1999). The role of specific input and delay on infant perseverative reaching errors. Presented at the Biennial Meeting, Society for Research in Child Development, Albuquerque.

Cummings, E. M., & Bjork, E. L. (1983a). Perseveration and search on a five-choice visible displacement hiding task. *Journal of Genetic Psychology, 142,* 283-291.

Cummings, E. M., & Bjork, E. L. (1983b). Search and behavior on multi-choice hiding tasks: Evidence for an objective conception of space in infancy. *International Journal of Behavioral Development, 6,* 71-87.

De Fockert, J., Rees, G., Frith, C., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science, 291*, 1803-1806.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

Diamond, A. (1983). Behavior changes between 6-12 months of age: What can they tell us about how the mind of the infant is changing? *Dissertation Abstracts International, 44*(01B), 337.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants’ performance on AB. *Child Development, 56,* 868-883.

Elman, J., Bates, E., Karmiloff-Smith, A., Johnson, M., Parisi, D. & Plunkett, K. (1996). Rethinking Innateness: A Connectionist Perspective on Development. Cambridge, MA: MIT Press.

Evans, W. F. (1973). The stage-IV error in Piaget's theory of object concept development: An investigation of the role of activity. *Dissertation Abstracts International, 34*(11B), 5651.

Fodor, J. (1983) The Modularity of Mind. Cambridge, MA: MIT Press.

Frank, M.J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience, 17*, 51-72.

Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Research, 330*, 299-307.

Hofstadter, M. C., & Reznick, J. S. (1996). Response modality affects human infant delayed-response performance. *Child Development*, *67*, 646-658.

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology 20, 573-591.*

Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection and rule execution in preschoolers: An error detection approach. *Developmental Psychology, 35*, 770-780.

Joanisse, M. F., Seidenberg, M. S. (2003) Phonology and syntax in Specific Language Impairments: Evidence from a Connectionist model. Brain and Language. 86, 40-56.

Jog, M.S., Kubota, Y. & Graybiel, A.M. (1999). Building Neural Representations of

Habits. *Science, 286,* 1745.

Kirkham, N. Z., Cruess L. & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimensional-switching task. *Developmental Science, 6,* 449-476.

Kharitonova, M., Chien, S., Colunga, E., & Munakata, Y. (2006). The emergence of human intelligence: An early synergy of flexibility and generalization. Manuscript in preparation.

Leslie, A.M. (1992). Pretence, autism, and the Theory-of-Mind-Module. Current Directions in Psychological Science, 1, 18-21.

Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development, 70,* 1297-1313.

Miller E. K. & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263,* 520-523.

Miller E. K., Erickson C. A. & Desimone, R. (1996). Neural mechanisms of visual working memory in the prefrontal cortex of the macaque. *The Journal of Neuroscience, 16,* 5154-5167.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development, 72,* 834-843.

Morton, J. B., Trehub, S. E., & Zelazo, P. D. (2003). Representational inflexibility in children's interpretation of emotion in speech. *Child Development, 74,* 1857-1868.

Morton, J. B., & Munakata, Y. (in preparation). The appearance (and disappearance) of knowledge-action dissociations in adults.

Munakata, Y. (1997). Perseverative reaching in infancy: The roles of hidden toys and motor history in the A-not-B task. *Infant Behavior and Development, 20*, 405-416.

Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A PDP model of the AB task. *Developmental Science, 1*, 161-184.

Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science, 12*, 335-337.

O'Reilly, R.C. & Munakata, Y. (2000). Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. Cambridge: MIT Press.

Patalano, A.L., Smith, E.E., Jonides, J. & Koeppe, R.A. (2001). PET evidence for multiple strategies of categorization . Cognitive, Affective, and Behavioral Neuroscience, 2, 360-370.

Piaget, J. (1954). *The Construction of Reality in the Child.* New York: Basic Books.

Rougier, N.P., Noelle, D., Braver, T.S., Cohen, J.D. & O'Reilly, R.C. (2005). Prefrontal Cortex and the Flexibility of Cognitive Control: Rules Without Symbols. Proceedings of the National Academy of Sciences, 102, 7338-7343.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*, 343-348.

Smith, L.B., Thelen, E., Titzer, B. & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review, 106*, 235-260.

Spencer, J. P., & Schoner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science, 6*, 392-412.

Spencer, J. P., & Schutte, A. R. (2004). Unifying representations and responses: Perseverative biases arise from a single behavioral system. *Psychological Science, 15,* 187-193.

Thelen, E., Schoner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences, 24,* 1-86.

Thomas, M., & Karmiloff-Smith, A. (2003). Connectionist models of development, developmental disorders, and individual differences. In R. J. Sternberg, J. Lautrey, and T. I. Lubart (Eds.), *Models of intelligence: International perspectives.* Washington D.C.: American Psychological Association.

Wallis J. D., Anderson K. C. & Miller E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411*, 953-956.

Wellman, H. M., Cross, D., & Bartsch, K. (1986). Infant search and object permanence: A meta-analysis of the A-not-B error. *Monograph of the Society for Research in Child 51*(3 Serial No. 214).

Yerys, B. E. & Munakata, Y. (in press). When labels hurt but novelty helps: Children's perseveration and flexibility in a card-sorting task. *Child Development*.

Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In M. Stamenov (Ed.). *Language Structure, Discourse, and the Access to Consciousness* (pp. 113-153). Amsterdam & Philadelphia: John Benjamins.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11,* 37-63.

Figures and Captions

Figure 1: The AB network

Figure 2: Input patterns for five A-trial segments (start, toy presentation, cover presentation, delay, and choice) in the standard 2-location AB task: The amount of activity on a given input unit is represented by the size of the square in that unit. The three sets of units, from left to right, represent location (A, B, and C), cover type (C1 and C2), and toy type (T1 and T2).

Figure 3: Percent correct responses as a function of age: On A trials, the network is accurate across all levels of recurrence shown because any ability to maintain activity for A is supported by latent weight changes (built up across previous trials) that favor A over B. On B trials, the network responds non-perseveratively only as the recurrent weights get stronger. The gaze/expectation system shows non-perseverative responding earlier than the reach system.

Figure 4: Patterns of network activation during the Presentation, Delay, and Choice periods of an A-trial when recurrence is low. The network correctly represents, reaches, and gazes to A.

Figure 5: Patterns of network activation during the Presentation, Delay, and Choice periods of a B-trial when recurrence is low. During Presentation, the network represents B. However, because recurrence is low, this representation fades during the Delay period. By Choice, the model more strongly represents A, and reaches and gazes perseveratively.

Figure 6: Patterns of network activation during the Presentation, Delay, and Choice periods of a B-trial when recurrence is high. The model accurately represents, gazes and reaches to B, despite a strong latent bias to represent, gaze, and reach to A.

Figure 7: Patterns of network activation during the Presentation, Delay, and Choice periods of a B-trial when recurrence is moderate. A weak representation of B at Choice is adequate to drive correct gaze but inadequate to drive correct reach.

Figure 8: Percent correct responses as a function of age and toy presence: On A trials, the network is accurate with both cover and toy trials across all levels of recurrence shown, because any activity for A is supported by latent weight changes that favor A over B. On B trials, the network perseverates on all task variants at some point in development, but shows non-perseverative responding earliest on the A-cover/B-toy (cover-toy) variant.

Figure 9: Percent correct responses to B on B trials as a function of age and delay. Data show a U-shaped developmental curve, such that when delay > 1, B-trial accuracy initially declines with increasing recurrence before increasing at higher levels of

recurrence.

Figure 10: A and B trials in toy-unattached (A) and toy-attached (B) versions of the cloth-pulling task (Aguiar & Baillargeon, 2000). On A trials, infants pull the towel with the toy on it. On B trials, infants reach perseveratively to the prior location, although the correct towel is specified by the visible environment.

Figure 11: Percent correct and reaction time results from networks at different ‘‘ages’’ (with different strengths of recurrent connections supporting sustained activations), on A trials (shown for just the toy-attached condition, with similar results in the toy-unattached condition) and B trials. Networks showed the developmental progression observed in infants. Changes in recurrent connections alone allowed networks to progress from perseverating to succeeding on B trials, first in the toy-attached and then in the toy-unattached condition. Thus, performance improved through the same changes that increased working memory abilities in prior models, despite the lack of apparent memory demands in these tasks. Reaction times were slowest around the transitions from perseverating to succeeding.

Figure 12: A simplified version of the card-sort network (R= red, B=blue, T=truck, F=flower), with the five inputs for a trial shown in the five rows below. Each trial included a presentation of the rule (first row), a description of the rule (second and third rows), a description of a test card (fourth row), and a presentation of a test card (fifth row). The description of the rules also included activation of the relevant output units (not shown).

Figure 13: Percent correct responses as a function of age. The Network performed perfectly on pre-switch trials, but perseverated on postswitch trials when recurrence was low.

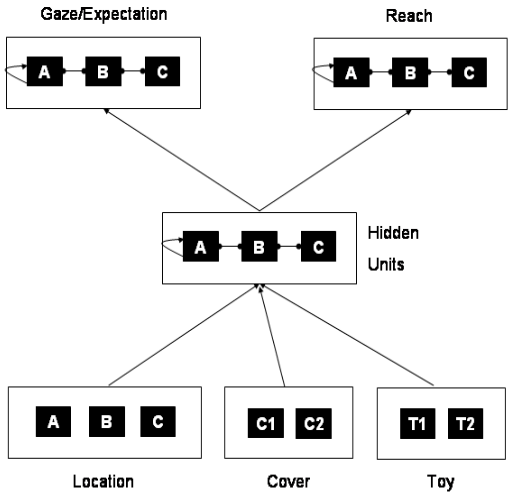

Gaze/Expectation
A
B
C
Reach
A
B
C
A
B
C
Hidden
Units
A
B
C
C1
C2
T1
T2
Location
Cover
Toy

Figure 1

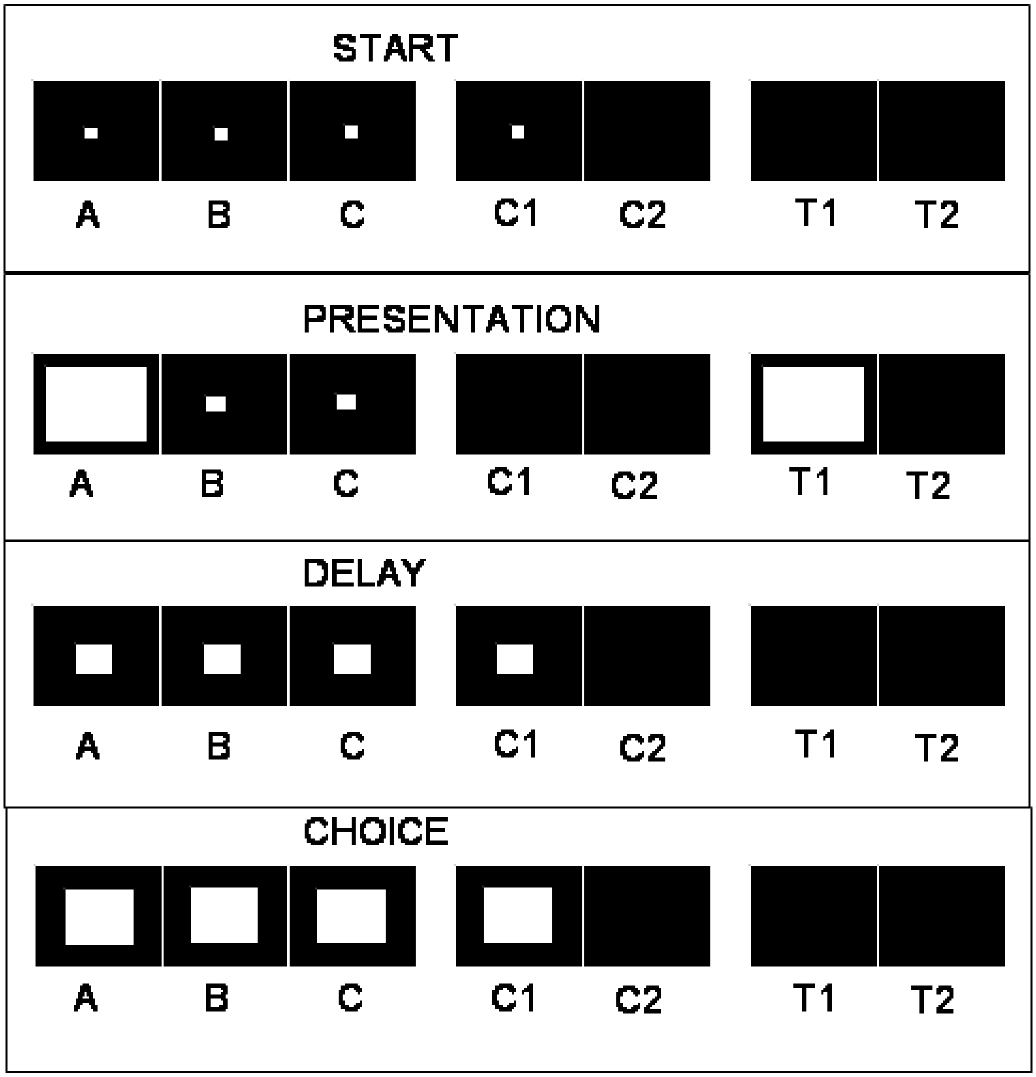

START
A
B
C
C1
C2
T1
T2
PRESENTATION
A
B
C
C1
C2
T1
T2
DELAY
A
B
C
C1
C2
T1
T2
CHOICE
A
B
C
C1
C2
T1
T2

Figure 2

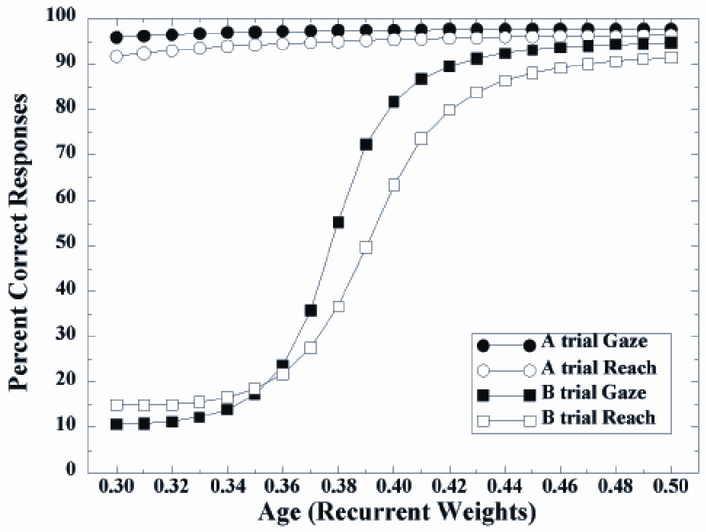

Percent Correct Responses
Age (Recurrent Weights)
0
10
20
30
40
50
60
70
80
90
100
0.30
0.32
0.34
0.36
0.38
0.40
0.42
0.44
0.46
0.48
0.50
A trial Gaze
A trial Reach
B trial Gaze
B trial Reach

Figure 3

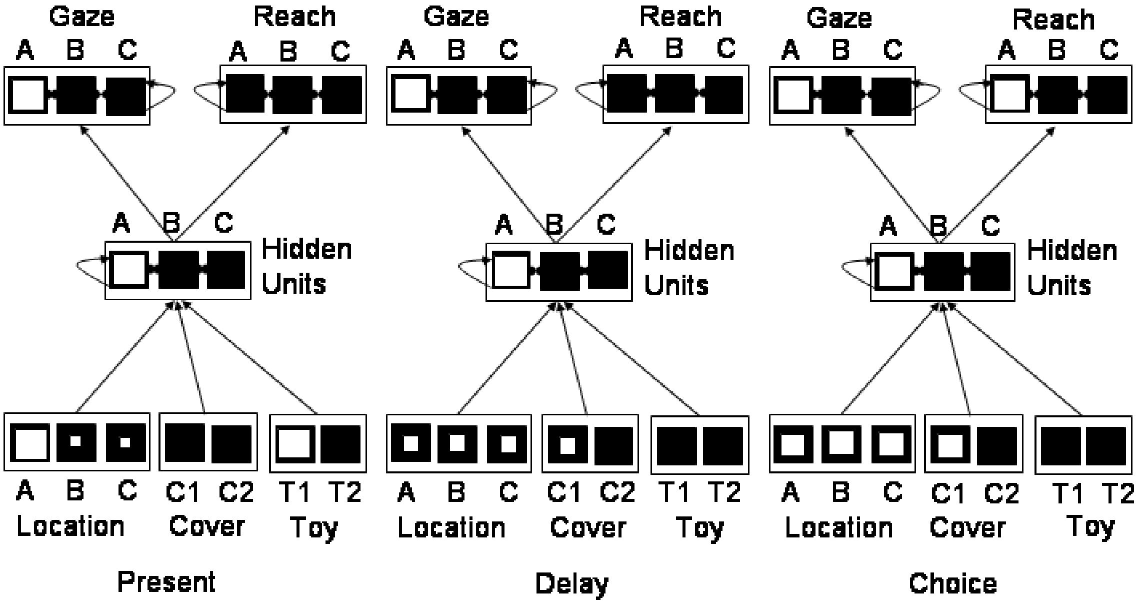

Gaze
A B C
Reach
A B C
A B C
Hidden Units
A B C
Location
C1 C2
Cover
T1 T2
Toy
Present
Delay
Choice

Figure 4

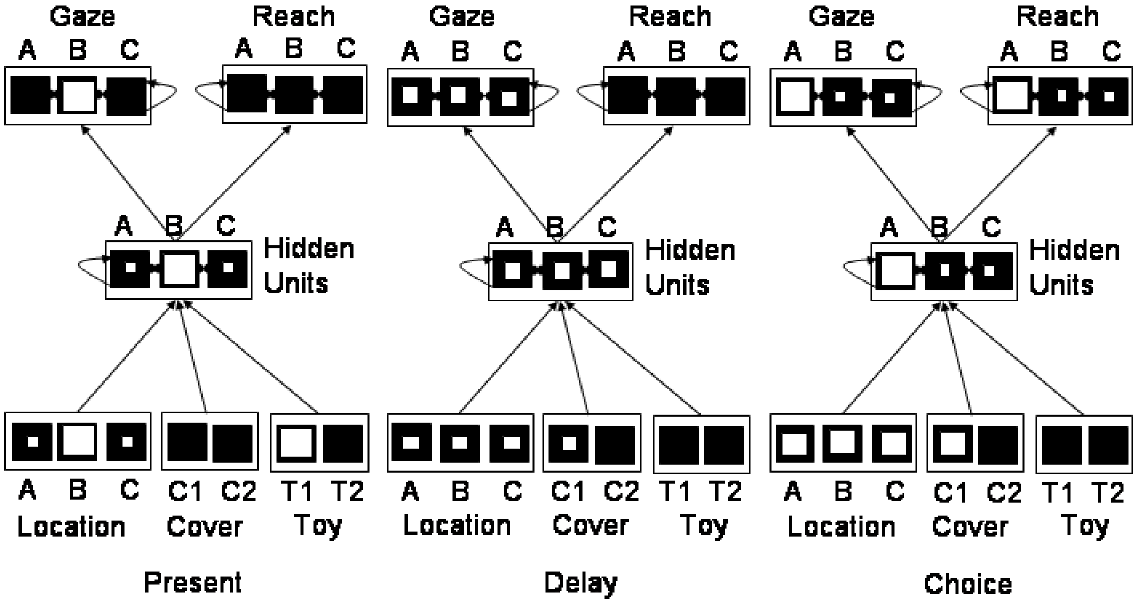

Gaze
Reach
A
B
C
Hidden Units
Location
Cover
C1
C2
Toy
T1
T2
Present
Delay
Choice

Figure 5

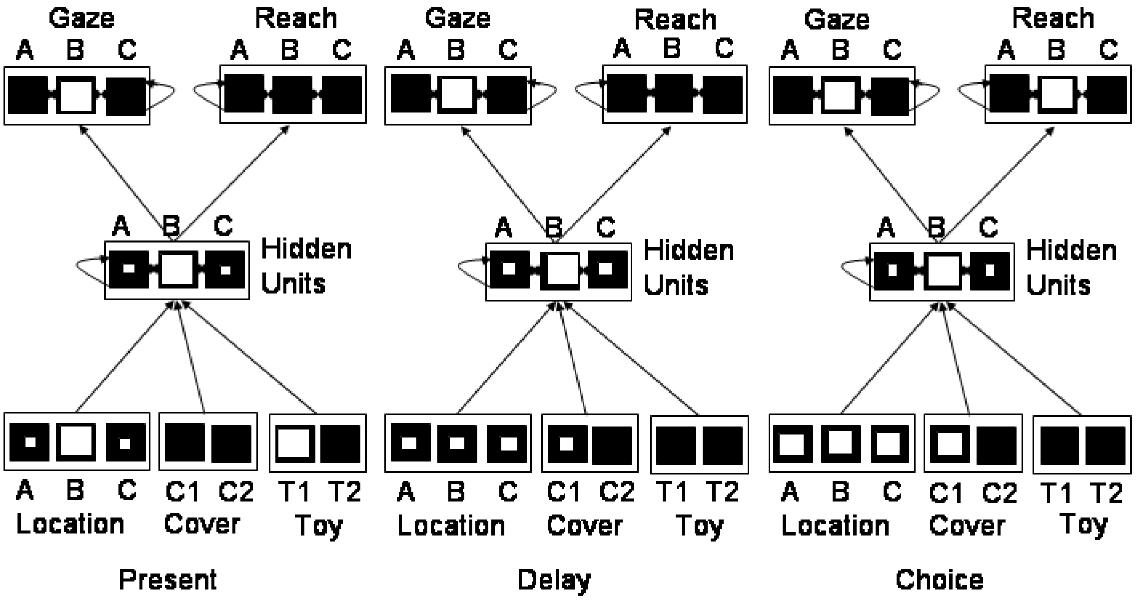

Gaze
A B C
Reach
A B C
A B C
Hidden Units
A B C
Location
C1 C2
Cover
T1 T2
Toy
Present
Delay
Choice

Figure 6

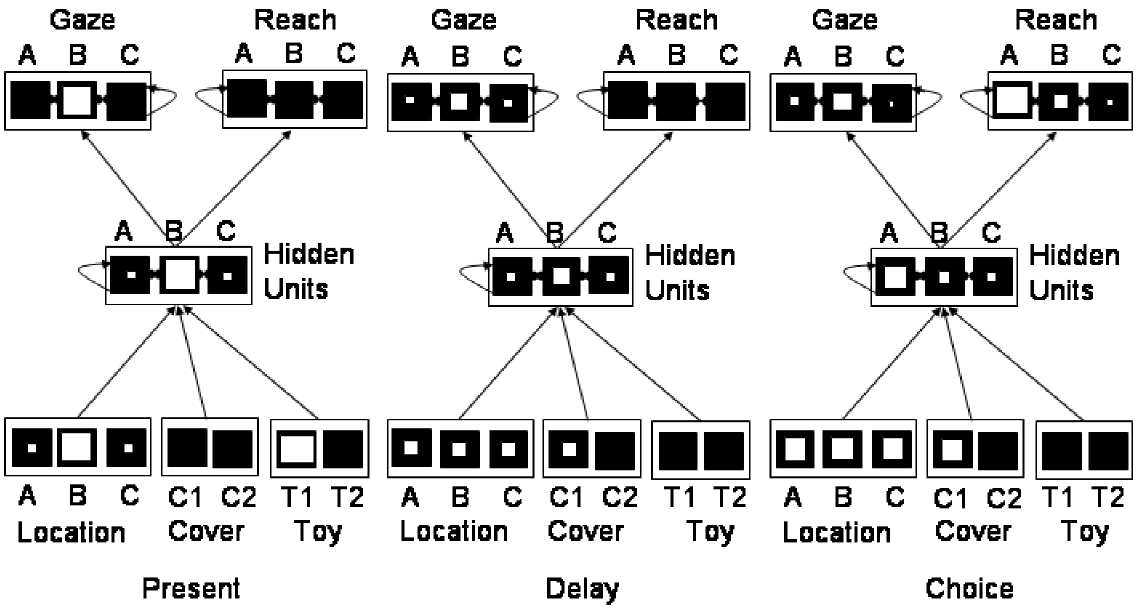
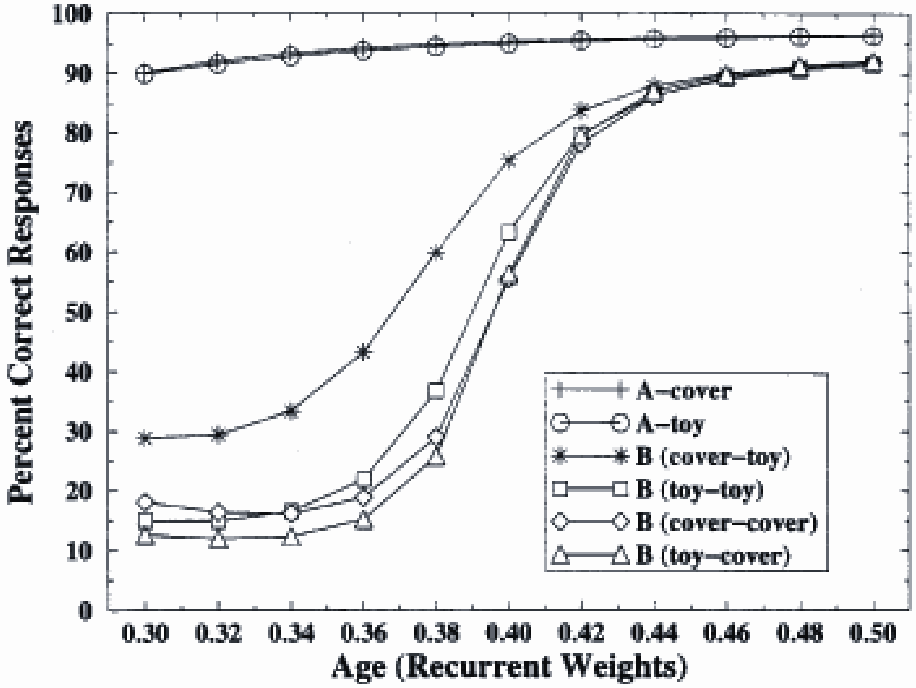

Gaze
Reach
A B C
Hidden Units
Location
Cover
Toy
C1 C2
T1 T2
Present
Delay
Choice

Figure 7



Percent Correct Responses
100
90
80
70
60
50
40
30
20
10
0
0.30
0.32
0.34
0.36
0.38
0.40
0.42
0.44
0.46
0.48
0.50
Age (Recurrent Weights)
A–cover
A–toy
B (cover–toy)
B (toy–toy)
B (cover–cover)
B (toy–cover)

Figure 8

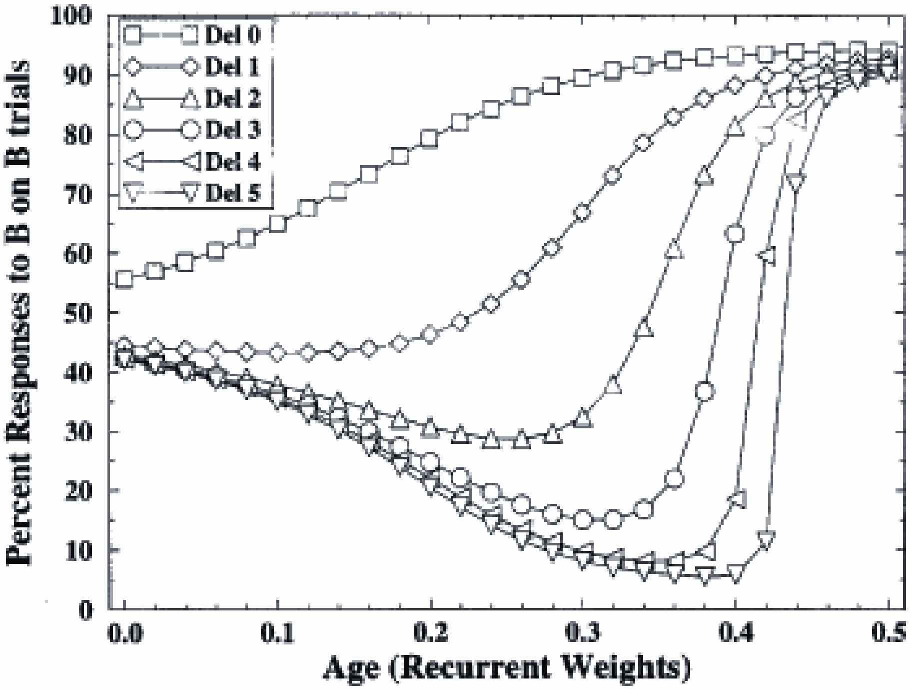

Percent Responses to B on B trials
Age (Recurrent Weights)
Del 0
Del 1
Del 2
Del 3
Del 4
Del 5
100
90
80
70
60
50
40
30
20
10
0
0.0
0.1
0.2
0.3
0.4
0.5

Figure 9

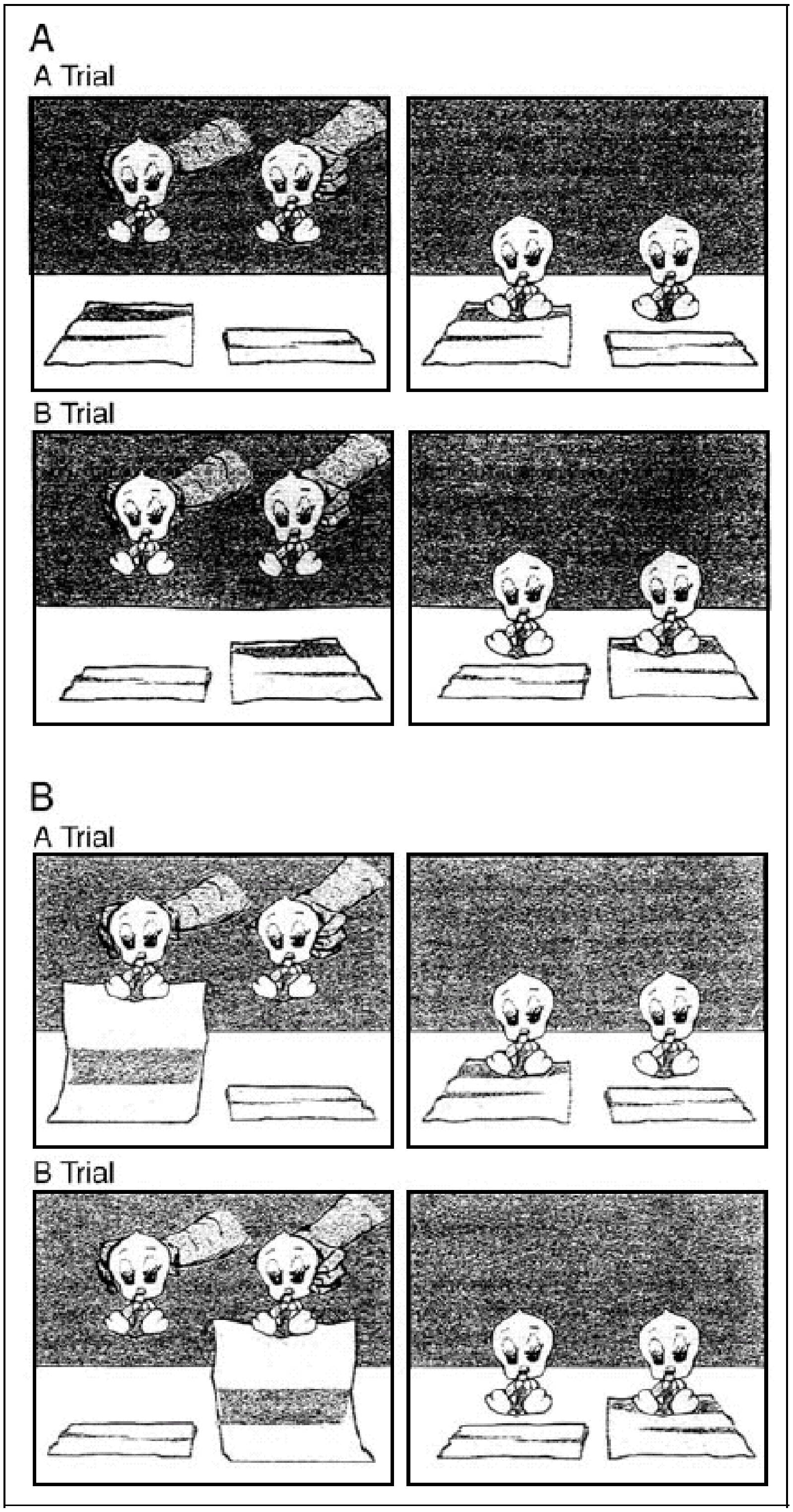


Figure 10

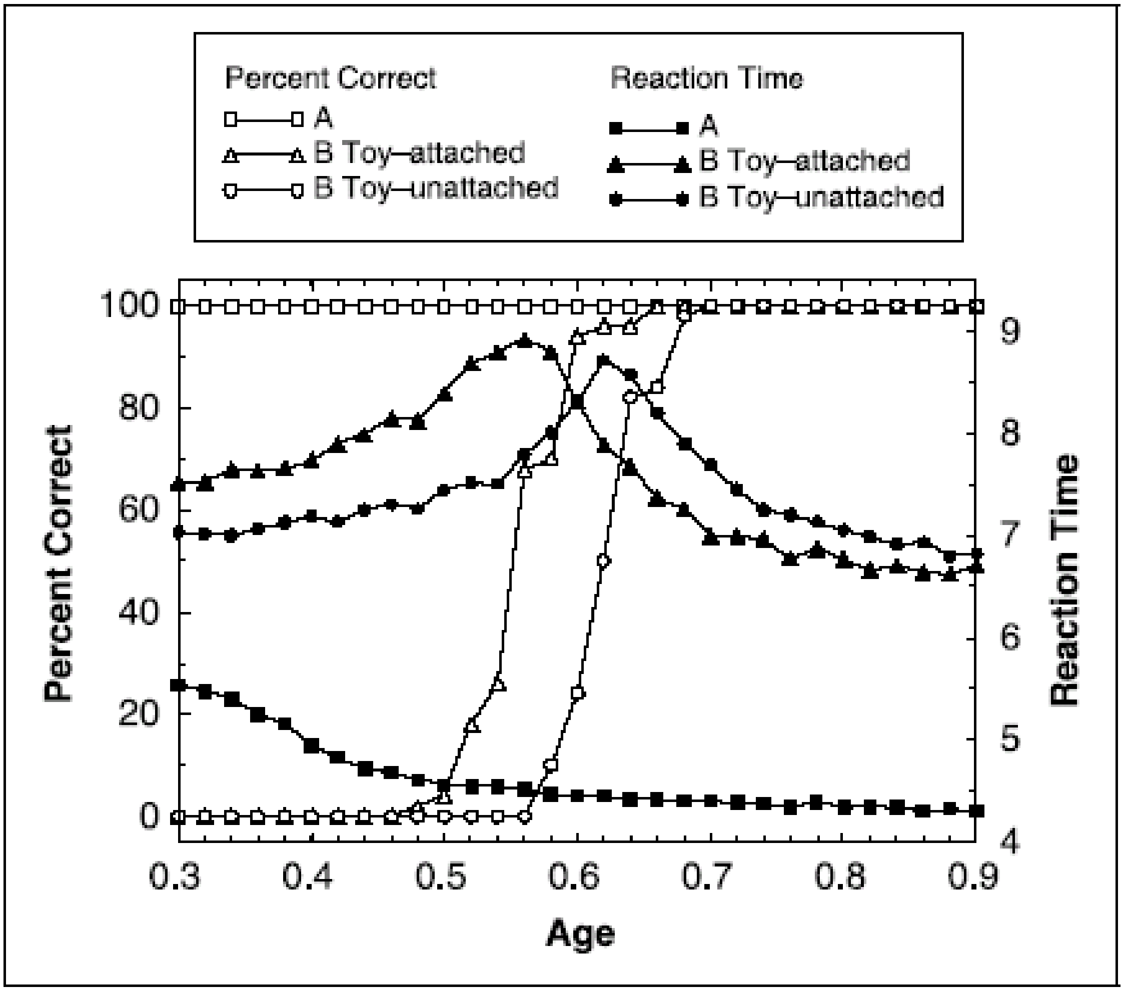

Percent Correct
A
B Toy—attached
B Toy—unattached
Reaction Time
A
B Toy—attached
B Toy—unattached
Percent Correct
100
80
60
40
20
0
Reaction Time
9
8
7
6
5
4
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Age

Figure 11

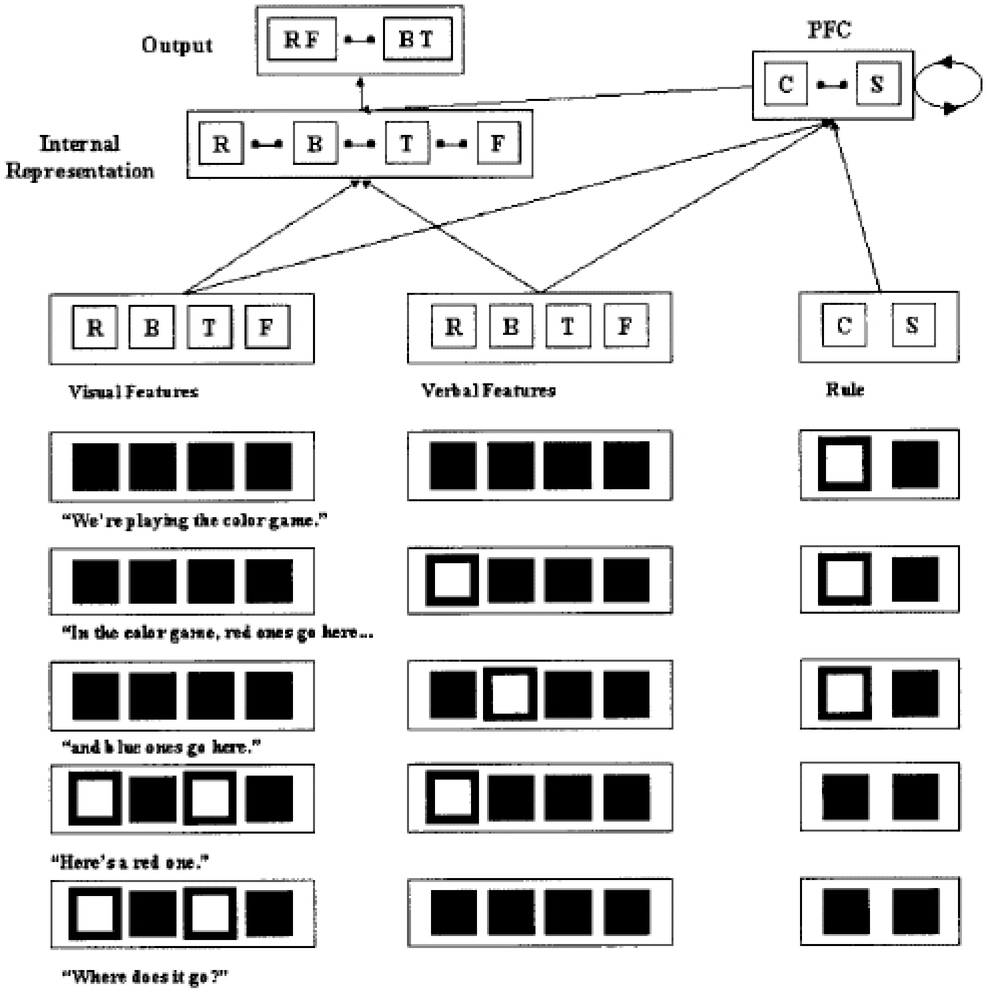

Output
RF
BT
PFC
C
S
Internal Representation
R
B
T
F
R
B
T
F
R
B
T
F
C
S
Visual Features
Verbal Features
Rule
"We're playing the color game."
"In the color game, red ones go here...
"and blue ones go here."
"Here's a red one."
"Where does it go?"

Figure 12

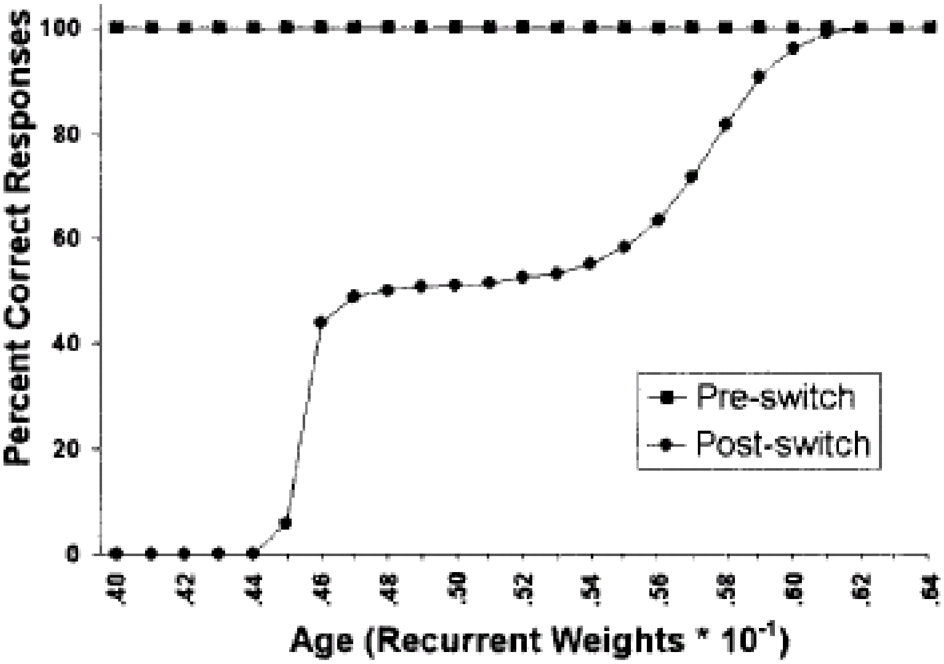

Percent Correct Responses
100
80
60
40
20
0
.40
.42
.44
.46
.48
.50
.52
.54
.56
.58
.60
.62
.64
Age (Recurrent Weights * 10-1)
Pre-switch
Post-switch

Figure 13

Embodiment and dynamic systems: Binding objects to locations in space

Linda Smith

The idea of emergence – the temporary but coherent coming into existence of new forms through ongoing intrinsic processes–is fundamental to the idea of dynamic systems. Complex systems composed of very many individual elements embedded within, and open to, a complex environment can exhibit coherent behavior: the parts are coordinated without an executive agent, plan, or program. Rather, coherence is generated solely in the relationships between the components and the constraints and opportunities offered by the environment. This self-organization means that no single element has causal priority. When such complex systems self-organize, they are characterized by the relative stability or instability of their states. These ideas have been used to explain a variety of phenomena in the physical and biological world, and increasingly they have been used – as this volume attests – in the study of developmental process.

These foundational tenets of dynamic systems have particularly radical implications for how we think about cognition. In particular, they imply each that thought is an in-the-moment unique event, open to a continually changing world, and the product of the intrinsic dynamics of a non-stationary system.

The embodiment of cognition would seem to demand a dynamic systems approach. The embodiment hypothesis is the idea that intelligence emerges in the interaction of an organism with an environment and as a result of sensori-motor activity. Notions of embodiment strongly challenge one key component of the traditional cognitivist approach, a component that lives on in much connectionist modeling: that cognition (and representation) is separate from and made of different stuff than perception and action. This case study of the A not-B error suggests instead that knowledge may be emergent from, embedded in, distributed across, and inseparable from real time processes of perceiving and acting.

To this end, this chapter will consider a dynamic systems account of the A-not-B error. The chapter will present theory and data indicating that the knowledge that underlies performance in the task emerges in mechanisms of perceiving and acting, and more specifically in visually guided reaching and in memories that are specifically linked to the body's own current position. Finally, the chapter will show how these very same –and very embodied processes – underlie infants and older children's ability to connect events over time, and specifically to bind names to things. That is the processes that make the A-not-B error are not about a deficit in children's thinking but are instead about how cognition –for all of us, for infants, children adults – is bound to the world in real time through the body. This fact demands dynamic systems.

Running head: EMBODIMENT

Brain, body, and mind: Lessons from Infant Motor Development

Daniela Corbetta

The University of Tennessee, Knoxville

Address for correspondence:

Daniela Corbetta, Ph.D.
Department of Psychology
The University of Tennessee
Austin Peay Building
Knoxville, TN 37996-0900

(865) 974-3346
dcorbett@utk.edu

Brain, body, and mind: Lessons from Infant Motor Development

> *"One cannot simply 'peel away' the body to understand the nervous system's role in adaptive behavior"*
>
> *(Chiel & Beer, 1997, p. 554)*

The body is the host of our thoughts. It is with our body that we think, express emotions, reach states of mind, and formulate intentions. Although these mind processes are usually thought of as the direct product of our brain activity, they are inseparable from the body in which they reside, and with which they are actualized. These embodied expressions of our brain activity occur continuously, moment by moment, whether we are letting our thoughts wonder while strolling along a pathway or whether we are thinking hard while sitting still in front of our computer. Our body is also the crucial pathway through which our language, gestures, and actions are actualized and perceived by others. It is in concert with our body, and by mastering its intrinsic physical and physiological properties that we learn to express, direct, perceive, receive, feel, and continuously exchange information with our environment. In sum, our body with its physical properties is our vital liaison between our mind and the outer world and this liaison is constant throughout lifespan as we grow and interact with the environment.

Can we ignore the body when studying the development of the brain and the mind? Is it possible to conceive a mind, pure and autonomous, that would not take into account, would not be influenced by, or would function independently from the body in which it resides? Could our thinking and reasoning, even in their most abstract form, grow and evolve without such fundamental connection to the outer world – our body? Despite the fact that body and mind are continuously and tightly interconnected in time, cognitive sciences have largely neglected the former to the benefit of the later. Body and mind have been defined mostly hierarchically. The mind guides, commands, decides, discovers, and organizes, while the body executes, expresses, or simply reflects the commands or states dictated by the mind. This neglect of the body has become even more extreme today as a great deal of theoretical and technological advances have made possible to study the brain – the site of our mind – more closely. Computer models simulating complex neural networks have emerged as a powerful way to understand the mind and its formation, however, on many instances, such models study a mind denudated from its body.

The role that the body plays in the formation of the mind is a major point of contention between the dynamic systems and the connectionist's approaches. While dynamic systems consider the body – its physiological and biomechanical properties – a crucial component in the formation of behavior and thought, connectionists have largely ignored the body by focusing solely on the brain and neural processes underlying mind formation. The goal of this chapter is to discuss how the body contributes to brain and mind development. I will begin by discussing briefly how the notion of embodiment is central to the dynamical systems approach (more extended tutorials on dynamic systems, brain, behavior, and development can be found in other chapters in this book or through other sources such as Kelso (1995) or Thelen & Smith (1994)). Then, the core of this chapter will rely on three examples from studies on infant motor development. These examples will show how the biomechanical and dynamic action properties of the body constitute an integral part of the process of behavioral and mind formation. They will be based on classical developmental questions on the development of reaching in infancy. Infant reaching has been traditionally studied from a cognitive perspective. As a result, the body and its action properties have not been considered as having a primary or influential role in the formation of developmental changes. In this chapter, I will contrast the traditional understanding of the process of change in infant reaching with recent work that paid particular attention to the infants' motor activities and led to the formulation of novel interpretations of what changes and how change occurs in reaching development. Two examples will show that some characteristics intrinsic to the movement can influence the infants' intentions or decisions when learning to reach. These examples, pointing to influencing properties of the body while performing the tasks, contrast sharply with the traditional cognitive approaches viewing the body as primarily guided by the mind. The third example will show that the emergence of novel sensory-motor experiences at specific points in development can influence prior skills, and significantly alter preferred patterns of response, which will shed some light on the processes underlying behavioral discontinuities in development and also contrast with the idea that some aspects of development are preprogrammed or inborn and thus impervious to change. Overall, these examples will emphasize the strong coupling that exists between the brain and the body in development and will stress the fact that studying a brain without a body may provide a partial, incomplete understanding of the developmental problems examined. The conclusions will sketch how the view of an embodied mind applies to situations where the body is not actively involved in the task. They will stress that development is a process marked by perceptual and sensory-motor experiences which involve a brain with a body that can act and sense. I will argue that those perceptual and sensory-motor experiences are crucial for brain, mind, and behavioral development and that they form the foundation of most forms of thinking in development.

Dynamics systems and the embodied brain: A brief overview

In 1997 Chiel and Beer published a wonderful paper in *Trends in Neurosciences* in which they offered a contrasting view to the commonly held view that the body is a mere acting device, subordinated to the commands of the mind. In that paper, they made a very simple, seemingly obvious, however extremely fundamental point, namely, that the brain is completely embedded in the body, and that the body, in turn, is completely embedded in the environment. Because of this embeddedness, the body is an obligatory pathway between the mind and the environment and therefore understanding the brain cannot be done without including the body in which it resides. Chiel and Beer (1997) also argued that the body, just like the brain and the environment, is a complex and highly structured dynamical system that has unique intrinsic mechanical and musculoskeletal properties. These properties are inherent to patterns of sensory and muscle activation and constitute a fundamental influence to the process of neural activity and change. Finally, Chiel and Beer (1997) argued that in development and time, brain, body and environment are continuously coupled to one another, and that adaptive behavior emerges from the interaction of all three coupled components.

Chiel and Beer’s view of an embodied brain is fully compatible with the dynamic systems approach. From a dynamic systems perspective, brain, body, and mind cannot be dissociated from one another. In fact, as pointed by Thelen (2000), there are no points in time where the mind is not embedded in the body, and there are no points in time where the brain does not receive information from the body. What changes with development and time is the nature of the coupling between brain, body, and environment. Thus, the mind and the behaviors observed are the product of this continuous change in coupling between brain, body, and environment.

From a dynamic systems approach, the body and its musculoskeletal properties is an integral component of the process of change. The body with its particular perceptual and motor capabilities fully participates to the process of mind, neural, and behavioral formation.

Furthermore, because change is the product of continuous, intricate, and multileveled interactions between brain, body, and the environment, one cannot establish a hierarchical order between these elements, nor is there one element that is more privileged than the others in driving change. Thus, the brain does not hold a particular or privileged status; rather, the brain contributes in concert with the body and the environment to the dynamic process of behavioral and mind formation. Just as a symphony emanates from the cooperative activity of the many musicians and their instruments in the orchestra, behavior and thought emerge from the active interactions of the many parts and components of the organism in its environment.

For researchers interested in development, studying change and capturing the processes of change is of primary importance. The fact is that change occurs continuously, no matter the time scale or magnitude of change. Thus, behaviors, thoughts, or states of mind do not have an enduring, definite entity; they evolve continuously, form, dissipate, and eventually reform as a function of time and interactions between elements of a system. Those changes are also closely linked to the kind of experiences that are associated to having a body with a particular set of perceptual and motor capabilities. Those experiences, their behavioral and cognitive outcome, and the developmental changes they entail are not independent from the way our body – its perceptual and motor systems – is organized.

Brain, body, and environment all play a central and intertwined role in the process of change. Whether one looses a limb, gains weight, discovers how to use computers, carries a heavy load, falls in love, or acquires the meaning of a new word, all these examples have the potential to alter dramatically the way individuals, their body, brain, and mind, perceive, act, and interact with their environment. Thus, the way the brain grows and develops cannot be separated from its developmental history and from the way it interacts with the body and the environment.

In the following sections, I use examples from research on motor development to illustrate specifically how the body and its characteristics play an integral part in the process of behavior and mind formation, and impact neural organization. The behaviors I present and discuss here after relate to different aspects of the development of reaching during the first year of life of infants. They relate to the emergence of voluntary reaching around 3-4 months of age, reaching in the context of the A-non-B task around 8-to-10 months old, and the fluctuation of patterns of reaching throughout the first year of life. These three aspects of the development of reaching have been heavily studied by many researchers before, but the explanations provided to account for the observed developmental changes in reaching are consistent and more in line with traditional, non-embodied cognitive views. Those views assume, for example, that infants reaching patterns are guided by vision and the mind, that they directly reflect the infant's state of reasoning or understanding of a particular situation at a particular point in time, or that they are the product of inborn or autonomous maturational changes in the brain. In my presentations, I will revisit those interpretations and show that when we acknowledge the role that the body and its sensory and motor characteristics play in the observed behavioral phenomena, another level of understanding emerges. In all three cases, I begin by presenting the behavioral facts and their classic interpretations, and then I present the corresponding alternate interpretation from a dynamic system, embodied view.

Learning to reach

The question of how young infants learn to reach for objects located in their nearby environment dates back to the first observations of Piaget (1952). How do infants, who have never reached for an object voluntarily before, come to figure out how to move their arm to where the desired toy is located? What are the characteristics of their movement on their first reaching attempts? And how do infants progress over time? Piaget (1952) contended that early reaching emerged from seeing the hand and the toy simultaneously in the same visual field. From that moment, reaching developed from a process of gradual guidance of the hand toward the toy, in which the infant glanced successively and alternately between hand and toy to bring the hand step-by-step closer the toy. Piaget's observations have pretty much set the stage for the long lasting interpretation that reaching develops primarily under the guidance of vision. Subsequent studies that examined the development of reaching in finer details by analyzing the movement kinematics of young reachers, found that during the few months following the onset of reaching, infants produce indeed highly meandering, indirect, zigzagging, discontinuous hand paths to the toy (Fetters and Todd, 1987; Mathew & Cook, 1990; von Hofsten, 1979, 1991). These indirect trajectories were assumed to reflect this stepwise process of visual guidance of the hand, where the infant reassessed the position of the hand relatively to the toy's position after each move and made successive corrections to bring the hand progressively closer to the toy. This early period of discontinuous hand trajectories was later labeled "visually guided" (Bushnell, 1985; Streri, 1991) and was very much in line with a cognitive approach to development where the mind commands the body.

This interpretation that early reaching is visually guided came into question in the late 80s, early 90s by a number of studies performed by Clifton and collaborators (Clifton, Muir, Ashmead, & Clarkson, 1993; Clifton, Rochat, Robin, & Berthier, 1994; Perris & Clifton, 1988). These researchers have demonstrated through a series of studies that young infants do not really need to see their hand to bring it to the toy; young infants can attain a target in the dark without seeing their hand or solely on the basis of auditory information, that is, without seeing the target either. This work, nonetheless, raised a new question. If young infants do not need to visually guide their hand to reach for a toy, why is their movement trajectory to the toy so convoluted? Why do infants show so many corrections in hand trajectory before attaining the target?

Thelen and colleagues discovered that the deviations in arm trajectory that are typical of infants' first attempts to reach arise from the interactions of the infants' own level of activity and intrinsic biomechanical properties of the body and arm at work (Thelen, Corbetta, Kamm, Spencer, Schneider, & Zernicke, 1993; Thelen, Corbetta, & Spencer, 1996). Thelen *et al.* observed that infants who were more active and thus moved their arm toward the target faster, also tended to generate more discontinuous movements, while infants who were calmer and moved their arm much more slowly, appeared to display much straighter trajectories to the toy. To understand this link between activity level and movement discontinuity, Thelen *et al.* analyzed the movement speed and trajectory patterns of the reaches in relation to the inertial forces that were generated by the infants' movements as they directed their hand toward the toy. They found that fast movements were associated with high motion dependent torques which tended to drag the arm away from the intended goal location. Thus, the active infants who produced these high speed motion patterns had to use their muscles not only for initiating movement, but also for breaking these deviations in trajectory and repeatedly redirecting their arm toward the toy. This ultimately caused the formation of convoluted hand paths to the toy. In contrast, the slower movements of the quieter infants appeared a lot more direct simply because they did not generate disruptive torques during the movement.

Thus, from this work it appeared that the discontinuous trajectories of the arm to the toy were not the result of stepwise visual guidance of the hand, but rather the product of the amount of energy that the infants were bringing into the task when reaching for the objects. High energy level generated high speed movements and high torques which interacted with the infants' plan or intent to direct their hand to the toy. Alternatively, slow energy level required mainly generating muscle strength to lift the arm and maintain it against gravity. Thus, learning to reach, attaining the toy, and improving movement directness over time, required that infants discover unique control solutions that were relative to their individual levels of energy, that is, active infants needed to slow down their movement and quiet infants needed to speed up their movement in order to attain the toy more readily. This process of scaling the movement speed to the goal involved discovering the biomechanical properties of the arm in motion. This could only be done through repeated sensory-motor experience as infant produced and practiced a wide range of movement solutions – some fast, some slow – and matched their respective outcome to the goal. In this scenario, the brain and specific neural networks involved in the formation of reaching trajectories are not solely responsible of monitoring the trajectory of the hand; they are also involved in encoding and mastering the specific motion-dependent dynamics of the arm as it moves towards the goal. Those biomechanical, motion-dependent properties of the body and the arm are integral pieces and parts of the larger plan and process of learning to reach for an object and the brain cannot ignore them when directing an arm to a specific goal location. Thus, the mind intending to plan an action with a goal also needs to deal and integrate in the plan the characteristics of the “tool” involved in the action, i.e., the body. In this context, body and mind are inseparable. Consistent with this scenario, Thelen *et al.* who followed the development of those infants over time were able to show that within a couple weeks and later during the year, infants learned to adjust their movement to achieve more direct reaches: the active infants learned to slow down their movements, while the slow infants were able to increase movement speed. In sum, infants learned to reach and map their action to their intention through the discovery of their body’s properties and more specifically through the discovery of the appropriate pattern of movement dynamics that allowed them to attain the target more directly. Such developmental change in pattern of action involved the coupled activity of the mind, brain, body, and environment.

Reaching in the context of the A-not-B task

The ability of infants to represent mentally and retrieve objects hidden in particular locations in space is at the core of this second issue. Piaget (1954), who was interested in the origins of intelligence, devised a task – the A-not-B task – to assess exactly that. In this task, infants in the 8 to 12 months old range are presented with a surface containing two wells located a few inches next to each other and both are covered with an identical cloth or lid. The infant first watches an experimenter hide a toy in one of those wells (location A). After a short delay, the infant is invited to retrieve the toy. This hiding/retrieving event to the A location is repeated a few times, then, the experimenter hides the toy in the second well (location B) as the child is still watching. Eight to 11 months old infants typically make an error; they still go to search for the toy in the A location despite having witnessed the experimenter hiding the toy in location B. Piaget's (1954) interpretation is that infants fail to search in location B because they have a weak or incomplete representation of where the object is located, but also because, at that age, infants cannot dissociate their thought from their actions.

Since Piaget, this error in the A-not-B task has been replicated by a very large number of studies involving many variations of the task. In those studies, interpretations of the error were almost always of a cognitive nature and they completely neglected the role that the action may have had in this task. For example, interpretations stressed the limited memory capacity of infants, their lack of object permanence, or their inability to coordinate information in time and space (see Diamond, 1990; Smith, Thelen, Titzer, & McLin, 1999; Thelen, Schöner, Scheier, & Smith, 2001: Wellman, Cross, & Bartsch, 1986 for reviews).

Curiously enough, the A-not-B error in infants is not limited to the context of a disappearing object. Some studies found that infants continue to make the error even when the object placed in location B remains fully visible (Butterworth, 1977), or when the task does not involve any hiding of the toy at all, but simply the repetition of reaches for a visible toy in a identical previous location (Munakata, 1997; Smith et al., 1999). Thus, from these studies it appears that the interpretations based solely on location memory or object representation are not sufficient to explain why infants continue to make the error and persevere reaching to the A location even when the object is perfectly visible in location B. Why then do infants make the A-not-B error?

An alternative explanation involving the body and infants' inability to flexibly control their arm during reaching was provided by Thelen and colleagues (Diedrich, Thelen, Smith, & Corbetta, 2000). This group of researchers recorded the kinematics of the arm movement while infants were reaching to the A location for 6 consecutive trials and then when cued to the B location for 2 additional trials. They observed that at each new reaching trial to the A location, the speed profiles of the arm movements to the target were becoming more and more alike, as if infants were becoming progressively locked into a specific reaching pattern as they were aiming repeatedly to the same location. Infants who displayed such growing pattern similarity in movement speed as a function of the repeated A trials were also more likely to continue to reach to A on the B trials, thus, making the A-not-B error. It turned out that those infants made the error to reach to A on the B trials because they lacked movement flexibility to break the strong motor pattern (or "motor habit") that had formed as a result of the previous reaches to A (see also

Clearfield & Thelen, 2001). Thelen *et al.* demonstrated that such “motor habits” and the associated likelihood of making the error did not form when infants performed a lesser number of reaches to A before the object was switched to B (Smith et al., 1999). Furthermore, when the infants’ arm was loaded with a weight, increasing movement inertia and making it more difficult for the infants to change movement trajectory, infants made the error even more frequently (Smith, Clearfield, Diedrich, & Thelen, 2006). Lately, further formal modelings of this A-not-B task and the perceptual-motor responses associated with it have confirmed that the number of prior reaches to A can contribute to form a memory of the motor pattern (see also Spencer, in this volume). This memory can be strengthened or weakened depending on the number of previous reaches to A, the motor experience of the child, the perceptual layout of the task, and other contextual factors, hence, increasing or decreasing the likelihood of making the A-not-B error depending on how these factors interact with one another (Thelen et al., 2001). Thus, here again, the response and errors of the infants in the A-not-B task are not solely the product of infants’ decisions or lack of mental representations. The mechanical and musculoskeletal properties of the body, the lack of movement flexibility required to “break” the forming reaching habit, and the number of previous trials to A, all contributed to the observed erroneous response on the B trials. Once again, mind, body, and environment are tightly coupled in time in the formation of the response.

Developmental fluctuations in hand preference in infant reaching

This last example examines the development of hand preference in infant reaching. Hand preference is linked to specific brain asymmetries (Corballis, 1991; Hellige, 1993), thus, this example is useful to explore forms of coupling between brain asymmetries, body, and behavior and their evolution in early development. One strong characteristic of hand preference is that 75 to 90% of the human population displays a preferred right hand use when achieving tasks requiring a certain level of dexterity, such as writing or drawing. Furthermore, this preferred hand use is quite manifest and steady in adults. One question that has puzzled scientists for many decades relates to how such a manual preference forms. This has prompted a significant number of developmental studies and interpretations (see Corbetta & Thelen, 2002; Provins, 1997, for reviews). Many of the interpretations however, boiled down to two main views: the one claiming that hand preference is an inborn behavioral trait, and the other, contending that hand preference develops during childhood.

Because right-handedness is such a strong and widespread characteristic among many humans, the view that hand preference is inborn has strongly predominated (see for example Annett, 1985; McManus, 1985). A number of developmental studies have significantly contributed to support this view of an inborn preferred hand use. One of the strongest evidence has come from studies that examined newborns' spontaneous head turn (Coryell, 1985; Michel, 1981; Viviani, Turkewitz, & Karp, 1978). These studies found that most newborns prefer to rotate their head to the right, and that this right preferred head turn correlates with a preferred right hand use later in early childhood. However, this connection between preferred spontaneous head turn at birth and later hand preference is not so obvious. Following the assumption that newborns' spontaneous head turn reflects an inborn behavioral bias which is setting the direction of hand preference very early on, one would expect to see some continuity in hand use as infants begin to reach and manipulate objects consistent with the early bias. Yet, studies that followed the development of hand preference during infancy did not find such consistent bias in hand use; rather, they found that infants' hand preference fluctuated many times during the first year, where periods of right hand use alternated with periods of left, two-handed, or no preferred hand use (Corbetta & Thelen, 1999; Gesell & Ames, 1947; Ramsay, 1985). Some authors proposed that these fluctuations were part of the natural process of brain lateralization and neuromotor maturation (Gesell, 1956; Ramsay, 1984). Nevertheless, if one assumes that hand preference is determined from birth, or genetically driven as it has been suggested (Annett, 1985; McManus, 1985), the question remains as to why hand use fluctuates so much during early development. This also raises the question of how stable hand preference emerges from such a background of fluctuating movements.

I became interested in these developmental questions while collaborating with Esther Thelen and her collaborators on a longitudinal study on infant reaching in which 4 infants were followed at close time intervals over their first year of life (see Thelen, et al., 1993, 1996). In that study, we observed many aspects of the motor development of those infants, including which hand they used for reaching and when they developed new posture-related skills. We noticed that developmental transitions in hand use for reaching occurred at particular times when the infants were beginning to develop new locomotor skills. All infants displayed alternating periods of one- versus two-handed reaching (Corbetta & Thelen, 1996) intertwined with periods of right versus left or no preferred hand use in reaching (Corbetta & Thelen, 1999). Nonetheless, besides these fluctuations, we could discern a developmental pattern in shifts in hand use that seemed to follow infants' postural and locomotor progression. The developmental pattern suggested by these preliminary data was as follow. During the early period following the emergence of reaching, infants displayed a pattern of preferred right hand use. This right bias dissipated toward the middle of the year when infants began to self-locomote on hands-and-knees. From that time, infants alternated hand frequently, trial after trial and week after week. Finally, toward the end of the first year, one infant began to perform his first upright independent steps. The emergence of independent walking in that infant coincided with a strong return of two-handed patterning in his reaching. This was particularly surprising because at the end of the first year all infants had developed fine reaching skills, the objects were always small and easily graspable with one hand, and they were always presented at midline. Plus, all infants were highly familiar with the task, they knew exactly what to expect, and had had months of practice reaching for those small objects with one hand. Thus, why did that one infant return to two-handed reaching at the end of the first year? That was quite puzzling, particularly knowing that nothing had changed in the task, other than the fact that that infant began to walk upright.

These early observations pointed to a developmental coincidence between the emergence of locomotion and change in patterns of hand use. The question arose as to whether new levels of postural control, acquired as infants learn to self-locomote, entailed successive reorganizations of the upper limb system and changed the way infants preferred to reach for and manipulate objects. That question was examined with a couple of follow-up studies with a larger subject pool and varied tasks. The first study investigated whether underlying preferred lateral biases in hand use would dissipate when infants begin to crawl on hands-and-knees (Corbetta & Thelen, 2002). In that study, two groups of infants were followed over several weeks as they learned to retrieve an attractive toy concealed in a box with a lid. One group of infants was tested weekly from the time they were 6 ½ months old and the other group was followed weekly as well from the age of 8 months. In the younger group, infants were not crawling. In the older group, infants just began or were beginning to crawl on hands-and-knees right around the onset of the study. All infants were tested with the same task; they had to learn how to coordinate their hands to lift and maintain open the lid of the box with one hand while retrieving the toy from the box with other hand (see Corbetta & Thelen, 2002, for more details).

The findings from that study were quite clear. The young infants, who were not crawling, displayed a stunningly stable division of hand role as they learned to solve the task over the successive weeks. Many infants systematically and consistently used the same hand to open and hold the lid and used the other hand to retrieve the toy. This consistent division of hand roles to retrieve the toy from the box was very steady trial after trial and week after week to the point that most infants displayed either no or little variability in response over the whole period of testing. The older infants, who were crawling, in contrast, displayed extremely changeable patterns of hand use to retrieve the toy from the box. They did so within sessions and across sessions to the point that it was impossible to discern a preferred pattern of hand use in that group of infants (Corbetta & Thelen, 2002). These findings were consistent with our prior longitudinal observations on reaching, namely, that preferred biases in hand use present before the onset on hands-and-knees crawling disappeared when infants learned to locomote on their four limbs.

Next, I turned to the development of walking and investigated whether a return to two-handed reaching was associated with learning to walk (Corbetta & Bojczyk, 2002). In this study, only one group of infants was followed longitudinally. Infants were seen weekly over the transition to upright locomotion, from the time they were 8 months old, and thus unable to walk, to about 2 months post upright locomotion. Every week, infants were presented with small and large objects at midline to test their one- or two-handed reaching tendencies, and they solved an object retrieval task identical to the one described above for the crawling infants. Additionally, the infant's postural and locomotor progresses were tested every week in order to identify the week in which they took their first independent steps without support. After infants began to walk, improvement in walking was monitored by indexing the height at which they kept their arms during walking. Indeed, when infants are learning to walk, they keep their arms in high guard to control balance, but as they practice walking and gain better upright balance they lower their arms along the body sides.

The results from this study were again very clear. During the *pre-walking* period, infants reached for the large objects mainly with two hands and used mainly one hand for the small ones. They also learned to open the lid of the box with one hand and coordinate the activity of their two arms to retrieve the toy out of the box without difficulty well before they began to walk. *When the infants began to perform their first independent steps, all these patterns changed.* Infants began to reach more with two hands for the small objects, they further increased their rate of two-handed reaching for the large objects, and began to open the lid of the box with two hands as well. All these increases in two-handedness were significant (Corbetta & Bojczyk, 2002). Again, this study supported our prior observations that infants return to two-handed reaching when they are learning to walk.

These studies confirmed that the development of early hand preference is not necessarily a behavioral trait set at birth, nor does it grow progressively and steadily over time. More importantly, our studies revealed that fluctuations in hand use were not happening randomly or autonomously, rather, they occurred as a function of the successive locomotor skills that infants developed over the course of their first year. The remaining problem was to elucidate why patterns of hand use happen to fluctuate with infants' progression in locomotion. This was not clear because in either studies, infants were always reaching or retrieving objects from the box while they were fully supported in an infant seat; they were not crawling or walking while they were reaching and there was no balance threat while in the seat that could have prompted a change in manual response. Furthermore, the same seat was used for the same tasks across infants and prior they learned to walk, thus the postural support provided during reaching remained unchanged at all times. So, why was the emergence of crawling associated with a disappearance of preferred hand biases and the emergence of walking associated with a return to two-handed reaching?

An answer to these questions was suggested by the last piece of data from our study on walking and reaching (Corbetta & Bojczyk, 2002). We looked at how long the period of two-handed reaching lasted after infants began to walk independently. Interestingly, we found that the rate of two-handed reaching declined significantly a few weeks later after infants began to lower their arms along the sides of their body during walking. In other words, coupling between arms in reaching while sitting dissolved at about the same time infants began to decouple and lower their arms while walking. Thus, the return and decline in two-handed reaching seemed to mirror the emergence of and subsequent progress in independent walking. It is this particular finding that suggested to us the mechanism that may have tied the observed developmental shifts in reaching with the development of infants' posture and locomotion. To explain this fully, I first need to make a detour and review some work in the neuroscience literature.

There are now several classic studies on brain plasticity showing that specific and novel sensory-motor experiences entail a functional reorganization of the sensory-motor cortex (Jenkins, Merzenich, Ochs, Allard, & Guic-Robles, 1990; Karni, Meyer, Rey-Hipolito, Jezzard, Adams, Turner, & Ungerleider, 1998; Kleim, Barbay, & Nudo, 1998; Merzenich & Jenkins, 1993; Merzenich, Allard & Jenkins, 1990; Petersen, Mier, Fiez, & Raichle, 1998, to cite a few). These studies are not limited to rats, non-human primates, and human adults. A similar cortical reorganization was reported in human infants, in particular, following the emergence of hands-and-knees crawling (Bell & Fox, 1996). These cortical reorganizations are closely related to the task that is learned and practiced, and hence, to the specific limbs, body parts, and sensory organs that are used to achieve these tasks. Particularly relevant to our developmental story, is the finding revealing that the acquisition of bimanual skills increases coupling of the cortical sensory-motor areas (Andres, Mima, Schulman, Dichgans, hallet, & Gerloff, 1999). Likewise, the use of one arm more than the other can entail a larger upper limb representation in the hemisphere contralateral to the hand used (Nudo, Milliken, Jenkins, & Merzenich, 1996). Finally and importantly, it has been shown that such experience-dependent neural reorganizations can, in turn, immediately shape motor performance (Dorris, Paré, Munoz, 2000). Thus, together, these studies show that there is a constant mapping between brain and behavior as new sensory-motor skills are learned and practiced. Moreover, such mapping occurs as these tasks are performed intentionally, that is, with a goal in mind.

Returning to our data on the development of infant's hand preference, my co-authors and I thought that a similar kind of mapping was taking place between reaching, crawling or walking, and the brain during critical periods of motor skill learning. When infants begin to crawl or walk, they develop a new behavior that was not present in their motor repertoire before. This new behavior involves using the body in a new way, by coordinating patterns of muscles differently, in order to achieve a new goal – the one of locomoting. When infants discover how to perform these new skills, they practice them a lot (Adolph & Berger, 2006). It is the practice of these novel motor skills that, we believe, has temporarily transferred to reaching. Figures 1 and 2 illustrate how this transfer may have occurred. We describe the case of learning to crawl first (figure 1). When infants are learning to crawl on hands-and-knees they need to figure out how to alternate hands and arms, and legs in order to locomote. As they do so, both arms acquire an identical status or role since they are now both used alternately but equally to support and move the body around. Note that despite similarity of roles, the activities of the hands remain uncoupled since they are used in alternation. Our data on hand preference revealed that such uncoupled, alternated, but similar hand role became the predominant mode of response in reaching, right at the time when infants began to practice hands-and-knees crawling (Corbetta & Thelen, 1999, 2002). Prior to the emergence of crawling, infants displayed a preferred hand use, this preferred hand use dissipated after they began crawling to the point that it became impossible to discern whether one hand was favored for reaching or manipulating objects. It seemed that infants could use either hand as well to reach and retrieve the object out of the box. The same scenario can be applied to the transition to upright locomotion. When infants are learning to perform their first independent steps, they raise their arms in high guard and couple them tightly to control their precarious balance during walking (figure 2). Again, when this new skill emerges, infants practice it intensely for many days and weeks with their arms up and tightly coupled. We observed that infants who were beginning to walk, also began to reach for objects at midline with two hands, suggesting that the tight arm coupling adopted and practiced during walking transferred to reaching. Furthermore, infants maintained this coupling as long as infants were coupling their arms during walking. When infants improved upright balance control, lowered their arms and decoupled them, then two-handed reaching declined as well (Corbetta & Bojczyk, 2002).

Recently, we obtained further support for our interpretation on transfer of learning by studying the reaching patterns of two young infants who adopted unique forms of self-produced locomotion (Corbetta, Williams, & Snapp-Childs, in press). One child who began to locomote by scooting on his buttock while in a sitting posture, also began to couple his arms during reaching over the same period. The coupling in reaching occurred, presumably, as a result of the arm coupling performed to generate the scooting motion. Another infant, who preferred to crawl on his belly and dragged his body on the floor by using the same steady lateral pattern between hands and legs continued to maintain a strong right hand use for reaching. Unlike infants who alternated arms for crawling on hands-and-knees and displayed a disappearance in hand preference, this infant maintained a strong right bias as a result of never alternating arm movements during belly crawling.

Together, these examples on the development of hand preference in relation to the emergence of different forms of locomotion reveal that brain, body and mind maintain a tight match over time depending on the type of locomotor pattern adopted by the child. Furthermore, this work reveals that developmental fluctuations in hand preference are intimately linked to the sensory-motor experiences that infants develop in the course of their first year. Thus, hand preference in early development is not necessarily a rigid, pre-established behavioral trait that is solely dictated by genes or an asymmetrical brain. Rather, early hand preference can be highly malleable and sensitive to the sensory-motor experiences and reorganizations of the postural system as infants learn to sit, crawl and walk, and brain asymmetries may be subject to similar malleability. Finally, this work shows that even highly practiced skills, such as reaching at the end of the first year, which is performed with a goal in mind and with prior knowledge of the actions involved and characteristics of the object target, can be sensitive to other patterns that are developed and practiced by the body over the same developmental period. Here again, the brain and the mind do not simply command the body, but are constantly influenced by new patterns of activity and sensations received from the body.

Discussion

The work reviewed above shows that the body is not simply an instrument at the service of the mind or merely responding to the commands of the brain. The body possesses its own physical and biomechanical characteristics which need to be assembled, integrated, and mastered as part of the process of development and interaction with the environment. Discovering one's own body properties is a fundamental aspect of behavioral development as infants, children, and even adults engage in and learn new tasks. The brain is not foreign to this process, neither in greater command, but involved in this process of behavioral learning as it fully partakes, in concert with the body and the environment, to the goal of achieving a particular task. In that sense, body, brain, and environment are continuously meshed together in time and it is through changes in their mutual coupling that behavior and mind emerge and evolve in development.

All the examples I used in this chapter were on perceptual and motor development and relied on tasks that required specific actions from the body to be solved. These examples were chosen because they were more amenable to demonstrate how the body plays a role in the formation of behavior and learning. But what if the task did not involve bodily actions? One could argue that thinking does not always require the active participation of the body and that some forms of representations, mental imaging, or mental manipulations of concept and symbols can be evoked and performed without needing to move any limb or part of the body. But is there such a thing as a pure form of thinking that is not connected to the body in which it resides? My inclination is that there isn't. The brain uninterruptedly "listens" to the body and is continuously tuned to sensations and feelings coming from the body even when the body lays still. Thus, even when we assume that the only brain activity actually taking place is happening in our mind, we still think in concert with our body. Some examples of self-generated mind activities, such as meditating or concentrating hard on a problem while sitting, are not independent from bodily sensations. In appearance, these mind activities do not require active body involvement, yet, the brain, occupied by the thinking or concentrating process, still continues to receive information about posture, body orientation, or even background sounds from the surrounding. Any sound from the environment, any feeling or sensation arising from our body can suffice to disrupt the level of concentration or stream of thoughts during mental activity. The fact is that we learn to think with our body and within our body. If an uncomfortable posture is disrupting our strain of thoughts or meditative activity, we learn to find a better seat, better position, or better location so the constraints arising from our own body won't interrupt our mind efforts. What we practice and consider as "pure" mental activities, are in reality, mind activities intimately coupled with our body. Other, more individually defined forms of mind-body coupling during the so-called "pure" thinking process can be seen in the frowning of the forehead, shaking of the foot, tapping of the pencil, or any other forms of activities that people engage into with their body to stimulate their thinking. Thinking is a form of body-mind coupling, a particular form of coupling that we learn and develop through infancy and early childhood as we are involved in problem solving. And this leads to my last point which deals with development, the origins of thought, and embodied cognition.

Where does the mind come from? How are thoughts generated? And how do memories or mental representations form? The answers to these questions are profoundly rooted in the developmental history of the individual, as infants, children, and adults act on and interact with their environment. Piaget (1967) already considered mental representations, memories, and problem solving abilities as the product of prior internalized patterns of perceptions and actions. Dynamics systems are compatible with the view that it is through the history of prior perceptions and actions in interaction with the task that mental forms such as memories, cognition, and patterns of behaviors emerge (see Spencer, this book; Smith, 2005). Many cognitive skills are, in fact, the product of a long developmental history involving sustained interactions between brain, body, and the environment through perception and action. The ability to count without needing our fingers or without needing pencil and paper, for instance, is the product of many repeated efforts at mapping the mind with the body pointing at targets or writing numbers. Language also is deeply grounded in the perceptual and motor system. Very early in development, infants fine-tune their hearing to sounds specific to their native language (Werker & Tees, 1984, 1992), and the production of sounds and speech develops as the child learns to control and master the complex speech apparatus progressively (Goffman, Heisler, & Chakraborty, 2006). And spatial representations have also been shown to form from perceiving and acting in the environment from a very early age (Campos, Anderson, Barbu-Roth, Hubbard, Hertenstein, & Witherington, 2000; Kermoian & Campos, 1988). Even highly skilled chess players, to use another example, who can remember the position of each piece on the checkerboard and anticipate several potential moves ahead of time, have reached that level of mental performance after years of practice at the game grounded in perception and action. Highly skilled chess players have developed those forms of mental skills from hours of prior perceptions and actions involving the brain, the body, and the environment as the pieces were moved on the checkerboard and the direct consequences of those particular moves were observed. In sum, it is hard to think of forms of reasoning or representations that are not grounded in their perceptual and motor developmental history and that are completely disconnected from the body. It is even harder to think of a disembodied mind, particularly when we realize that all the senses the mind relies on to connect with the environment are actually deeply embodied. Our senses of sight, touch, smell, hearing, and taste are intimately coupled with specific motor activities such as orienting the head, moving the fingers, activating our mouth and tongue, or turning the eyes to detect, capture, and encode information from the environment. Can we think of behaviors or skills that do not have their origins in perceptual and motor activity?

The ability to think without activating our body has its origins in the continuous body and mind mapping that takes place over developmental time. Just as infants manage to learn over the years to reach for objects in an endless number of varied situations without efforts and without having to think hard of the dynamics of their movement, we have learned to think, reason, count, and generate solutions in our mind without needing to activate or rely on our body. The fact that we can reach for objects seemingly “automatically” or think allegedly without using our body are not evidence of a separation between mind and body, but rather the indication that brain, body, and environment have reached a different level of coupling. In that sense, acquiring knowledge, becoming competent, remembering or developing social skills are not simply the product of a cognitive achievement; they are the product of a developmental real-time process of mutual interactions between brain, body, and environment.

References

Adolph, K. E., & Berger, S. E. (2006). Motor development. In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology, Vol. 2: Cognition, perception, and language* (pp. 161- 202). New York: Wiley.

Andres, F.G., Mima, T., Schulman, A.E., Dichgans, J., Hallett, M., & Gerloff, C. (1999). Functional coupling of human cortical sensorimotor areas during bimanual skill acquisition. *Brain, 122*, 855-870.

Annett, M. (1985). *Left, right, hand and brain: The right shift theory*. Hillsdale, NJ: Erlbaum.

Bell, M. A., & Fox, N. (1996). Crawling experience is related to changes in cortical organization during infancy: Evidence from EEG coherence. *Developmental Psychobiology, 29*, 551-561.

Bushnell, E. W. (1985). The decline of visually guided reaching during infancy. *Infant Behavior and Development, 8,* 139-155.

Butterworth, G. E. (1977). Object disappearance and error in Piaget's stage IV task. *Journal of Experimental Child Psychology, 23,* 391-401.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences, 20,* 553-557.

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., & Berthier, N. E. (1994). Multimodal perception in the control of infant reaching. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 876-886.

Clearfield, M. W., & Thelen, E. (2001). Stability and flexibility in the acquisition of skilled movements. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 253-266). Cambridge, MA: The MIT Press.

Corballis, M.C. (1991). *The lopsided ape: Evolution of the generative mind.* New York: Oxford University Press.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1*, 149-219.

Corbetta, D., & Thelen, E. (1996). The developmental origins of bimanual coordination: A dynamic perspective. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 502-522.

Corbetta, D., & Thelen, E. (1999). Lateral biases and fluctuations in infants' spontaneous arm movements and reaching. *Developmental Psychobiology, 34*, 237-255.

Corbetta, D., & Thelen, E. (2002). Behavioral fluctuations and the development of manual asymmetries in infancy: Contribution of the dynamic systems approach. In S. J. Segalowitz & I. Rapin (Eds.), *Handbook of neuropsychology (Vol. 8): Child neuropsychology, Part I* (pp. 309-328). Amsterdam, the Netherlands: Elsevier Science Publishing Company.

Corbetta, D., & Bojczyk, K. E. (2002). Infants return to two-handed reaching when they are learning to walk. *Journal of Motor Behavior, 34,* 83-95.

Corbetta, D., Williams, J., & Snapp-Childs, W. (in press). Plasticity in the development of handedness: Evidence from normal development and early asymmetric brain injury. *Developmental Psychobiology*.

Coryell, J. (1985). Infant rightward asymmetries predict right-handedness in childhood. *Neuropsychologia, 23,* 269-271.

Diamond, A. (1990). The development and neural bases of memory functions as indexed by the A-not-B and delayed response tasks in human infants and infant monkeys. In A. Diamond (Ed.), *The development and neural bases of higher cognitive functions* (pp. 267-309). New York: The New York Academy of Sciences.

Diedrich, F. J., Thelen, E., Smith, L. B., & Corbetta, D. (2000). Motor memory is a factor in infant perseverative errors. *Developmental Science, 3*, 479-494.

Dorris, M. C., Paré, M., & Munoz, D. P. (2000). Immediate neural plasticity shapes motor performance. *The Journal of Neuroscience, 20*, RC52, 1-5.

Fetters, L., & Todd, J. (1987). Quantitative assessment of infant reaching movements. *Journal of Motor Behavior, 19,* 147-166.

Gesell, A. (1956). The ontogenesis of infant behavior. In L. Carmichael (Ed.), *Manual of child psychology* (pp. 295-331). New York: Wiley.

Gesell, A., & Ames, L. B. (1947). The development of handedness. *Journal of Genetic Psychology, 70,* 155-175.

Goffman, L., Heisler, L., & Chakraborty, R. (2006). Mapping of prosodic structure onto words and phrases in children's and adults' speech production. *Language and Cognitive Processes, 21,* 25-47.

Hellige, J. B. (1993). *Hemispheric asymmetry: What's right and what's left.* Cambridge, MA: Harvard University Press.

Jenkins, W. M., Merzenich, M. M., Ochs, M. T., Allard, E., & Guic-Robles, T. (1990). Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *Journal of Neurophysiology, 63,* 82-104.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Science, 95,* 861-868.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: A Bradford Book, MIT Press.

Kermoian, R., & Campos, J. J., (1988). Locomotor experience: A facilitator of spatial cognitive development. *Child Development, 59,* 908-917.

Kleim, J. A., Barbay, S., & Nudo, R. J. (1998). Functional reorganization of the rat motor cortex following motor skill learning. *Journal of Neurophysiology, 80,* 3321-3325.

Mathew, A., & Cook, M. (1990). The control of reaching movements by young infants. *Child Development, 61*, 1238-1257.

McManus, I.C. (1985). Handedness, language dominance and aphasia: A genetic model. *Psychological Medicine* - *Monograph Supplement, 8,* 1- 40.

Merzenich, M. M., Allard, T. T., & Jenkins, W. M. (1990). Neural ontogeny of higher brain functions: Implications of some recent neurophysiological findings. In O. Franz & P. Westman (Eds.), *Information processing in the somatosensory system* (pp. 193-209). London: MacMillan Press.

Merzenich, M. M., & Jenkins, W. M. (1993). Reorganization of cortical representations of the hand following alterations of skin inputs induced by nerve injury, skin island transfers, and experience. *Journal of Hand Therapy, 6,* 89-104.

Michel, G. F. (1981). Right handedness: A consequence of infant supine head-orientation preference? *Science, 212,* 685-687.

Munakata, Y. (1997). Perseverative reaching in infancy: The role of hidden toys and motor history in the A-not-B task. *Infant Behavior and Development, 20,* 405-416.

Nudo, R. J., Milliken, G. W., Jenkins, W. M., Merzenich, M. M. (1996). Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *Journal of Neuroscience, 16,* 785-807.

Perris, E., & Clifton, R. K. (1988). Reaching in the dark toward sound as a measure of auditory localization in infants. *Infant Behavior and Development, 11,* 473-491.

Petersen, S. E., Mier, H., van, Fiez, J. A., & Raichle, M. E. (1998). The effects of practice on the functional anatomy of task performance. *Journal of Neurophysiology, 80,* 3321-3325.

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

Piaget, J. (1954). *The construction of reality in the child.* Oxford, England: Basic Books.

Piaget, J. (1967). *La psychologie de l'intelligence* [The psychology of intelligence]. Paris: Armand Colin.

Provins, K. A. (1997). Handedness and speech: A critical re-appraisal of the role of genetic and environmental factors in the cerebral lateralization of function. *Psychological Review, 104,* 554-571.

Ramsay, D. S. (1984). Onset of duplicated syllable babbling and unimanual handedness in infancy: Evidence for the developmental change in hemispheric specialization? *Developmental Psychology, 20,* 64-71.

Ramsay, D. S. (1985). Fluctuations in unimanual hand preference in infants following the onset of duplicated syllable babbling. *Developmental Psychology, 21,* 318-324.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review, 25,* 278-298.

Smith, L. B., Clearfield, M. W., Diedrich, F. J., & Thelen, E. (2006). Piaget was right: Thoughts close to the sensory-motor surface. Manuscript in preparation.

Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B Error. *Psychological Review, 106,* 235-260.

Thelen (2000) Grounded in the world: Developmental origins of the embodied mind. *Infancy, 1,* 3-28.

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

Thelen, E., Corbetta, D., & Spencer, J. P. (1996). The development of reaching during the first year: The role of movement speed. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1059-1076.

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*, 1-86.

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

Streri, A. (1991). *Seeing, reaching, touching.* Cambridge, MA: The MIT Press.

Viviani, J., Turkewitz, G., & Karp, E. (1978). A relationship between laterality of functioning at 2 days and 7 years of age. *Bulletin of Psychonomic Society, 12,* 189-192.

von Hofsten, C. (1979). Development of visually directed reaching: The approach phase. *Journal of Human Movement Studies, 5,* 160-178.

von Hofsten, C. (1991). Structuring of early reaching movements: A longitudinal study. *Journal of Motor Behavior*, *23*, 280-292.

Wellman, H.M., Cross, D., & Bartsch, K. (1986). Infant search and object permanence: A meta-analysis of the A-not-B error. *Monographs of the Society for Research in Child Development*, *51 (Serial No. 214).*

Werker, J. F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.

Werker, J. F., & Tees, R.C. (1992). The organization and reorganization of human speech perception. *Annual Review of Neuroscience, 15,* 377-402.

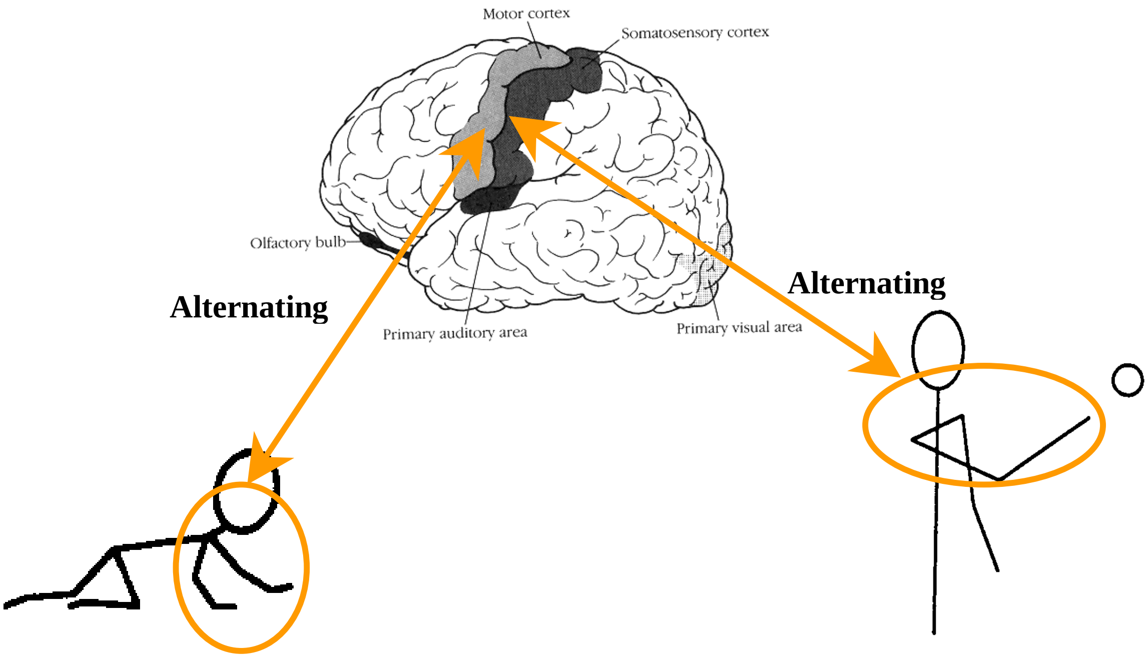


Figure 1: Process of transfer of learning related to the emergence of hands-and-knees crawling. When infants begin to self-locomote on their four limbs they learn to alternate arms and legs to move forward. This new alternated functional use of the arms for locomotion transfers to reaching, via brain functional reorganization, yielding a decline in preferred hand use.

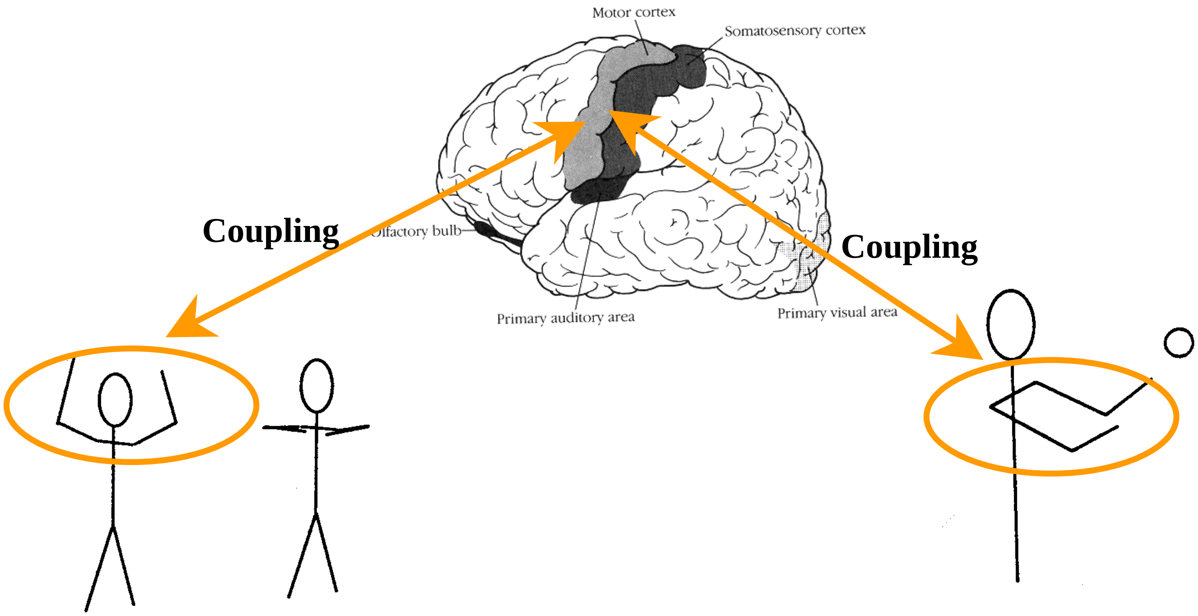


Figure 2: Process of transfer of learning related to the emergence of upright locomotion. When infants begin to walk upright, they keep their upper arms coupled in a high guard position. This upper arm coupling performed during walking temporarily transfers to reaching, via brain functional reorganization, yielding an increase in coupled, two-handed responses during reaching.

Transitions in cognitive development: prospects and limitations of a neural dynamic approach

Han L. J. van der Maas and Maartje E.J. Raijmakers

University of Amsterdam

Introduction

In this chapter we discuss old-fashioned cognitive development from the point of view of two modern approaches, connectionism, and non-linear dynamical system theory. Old-fashioned cognitive development refers to the development of typical human cognitive capabilities, such as reasoning, planning, and abstract thinking. These abilities are distinctively human, in contrast to abilities usually studied in connectionism and the dynamic system approach, which generally focus on perception, pattern recognition, association, motor action, and simple memory. Such abilities are of interest, of course, but do not constitute the greatest challenge for researchers in cognitive development. Yet, connectionists and dynamic system modelers do claim that their models are relevant to higher cognitive functioning. This is apparent in their rejection of other approaches, such as the Piagetian and the information processing accounts of development. The main assertion in both the connectionist and the dynamic system approaches is that higher cognitive functioning is largely based on non-symbolic, graded, and dynamic properties, of which these same approaches provide the best account.

In this chapter, we criticize this claim. We focus on sudden transitions in cognitive development. There is ample evidence for domain specific transitions, e.g., in the Piagetian concrete operational and formal operational tasks (van der Maas & Hopkins, 1998). Our key example will be the famous balance scale task as applied to access proportional reasoning. A typical item is shown in figure 1.

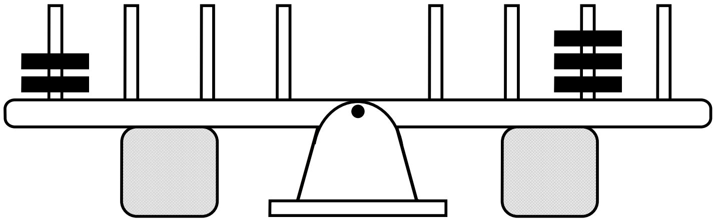


Figure 1: A balance scale item. What happens when the blocks are removed?

Children have to predict the movement of the scale when the blocks, as depicted in figure 1, are removed. It is well known that children use different rules or strategies to solve such items (Jansen & van der Maas, 2002). Young children will focus on the weights and ignore the distance information. Older children and many adults understand that the weight and distance cues conflict in the item that is displayed in figure 1, but only a small minority understands the correct torque rule that provides the solution to this conflict. The majority guesses or uses alternative rules, like the compensation rule, which is based on the sums of weight and distance on each side of the fulcrum. Siegler (1976, 1981) showed that children’s progression on this task takes place by the use of increasingly more complex rules. This is evident in the children’s responses, and in their verbal explanations.

There are three reasons that this task constitutes an interesting case for our present discussion. First, the balance scale task appeals to proportional reasoning. Proportional reasoning that is required in the balance scale task clearly classifies as higher cognitive functioning. Interestingly, the data on this task shows evidence for both graded continuous processing and for symbolic discrete processing. Second, there exists a large body of empirical results on this task, starting with the work of

Siegler (1976, 1981). A serious model for this task should explain the various well-established results, such as the diversity of rules, the order of rules, some deviations of rules, response time patterns, and transitions. Third, the balance scale task is used as a benchmark task in computational modeling. Specifically, in the last 25 years a number of symbolic and connectionist models have been proposed to explain the empirical results.

The main computational models for balance scale task performance are the connectionist model of McClelland (1989, 1995), the cascade correlation network model of Shultz, Mareshal and Smidt (1994), and the ACT-R model of van Rijn, van Someren and van der Maas (2003). The latter is based on a (sub) symbolic architecture. In view of the topics of this book, we concentrate on the connectionist models, and especially on their potential to explain rules and transition between rules in development on the balance scale task. This will provide us with the means to evaluate the general claim that higher cognitive function can be understood in terms of graded, non-symbolic, and dynamic properties of the brain.

However, there is one issue that has to be resolved first. We presented the connectionist and the dynamic system approach as unitary, because they make the same general claim about cognition, and both reject the traditional symbolic approaches to cognition. In our view the two approaches are very similar indeed. Because not everyone will subscribe to this view (e.g., Beer, 2000), we first explain our position with respect to this.

The relation between dynamic systems and connectionist models

A narrow definition of connectionism refers mainly to the typical multi-layered feed-forward back-propagation networks, as they are proposed in McClelland and Rumelhart (1986). We could broaden this view to include cascade correlation networks (Fahlman, 1988), recurrent networks (cf. Elman, 1995), and networks based on Adaptive Resonance Theory (Grossberg, 1976). In fact, at present the field of neural network modeling includes a large number of network types, from the very low level single neuron processes to the very high level neuronal fields. The only common aspect of these models is that they are somehow models of the neural substrate of the brain.

A narrow definition of the dynamic system approach refers to the well-known work of researchers, such as Thelen and Smith (1994), and the dynamic field approach of Spencer and Schöner (2003). In this definition, the dynamic system approach is a theory of development, in which actions in real time are modeled by means of differential equations. Most applications focus on infant behavioral development These researchers present this theory as a new theory of development processes, which provides an alternative to both the standard theories in developmental psychology, and to connectionism (Thelen & Smith, 1994). However, alternative accounts have been given. For instance, Van Geert (1991) proposed nonlinear growth models to explain several developmental phenomena. Van der Maas & Molenaar (1992) proposed the use of catastrophe theory to model and test transition in development. For other examples, we refer to Lewis (2000). These accounts do not make the same theoretical claims. Our own work is inspired by a broad definition of the dynamic system approach (for recent examples see, van der Maas & Raijmakers, 2000; Wagenmakers et al., 2005; Jansen & van der Maas, 2001; Ploeger, van der Maas and Raijmakers; in press).

In the broad sense, any application of concepts, methods, and models from the mathematical study of non-linear dynamical systems to developmental psychology falls under the dynamical approach. Any dynamical model with nonlinear interactions, developed from whatever theoretical perspective, is a nonlinear dynamical model. Any research on phase transitions, or chaos, is dynamical system research.

In our view there is no reason to prefer the narrow definitions of connectionism and the dynamical system approach. The close similarity between the dynamical field model (Schutte, Spencer, & Schöner, 2003) and the connectionist model (Munakata, 1998) of the A not B task supports our view. Of course the choice of definitions makes a large difference in thinking about the relations between the two approaches.

Neural networks are dynamical systems

An important and general conclusion that can be stated from the start is that connectionist models are a subset of the dynamical models, and that, therefore, conflict or competition between the approaches should be absent. Non-linear dynamical models describe systems in which the new state depends nonlinearly on its old state. In addition, the state can depend on states of other processes (such as other neurons), and on parameters describing external influences. Such models can be formulated either in discrete time (with difference equations), or in continuous time (with differential equations). Mathematical study of such models has revealed very rich and complicated behaviors, such as catastrophic changes, chaotic attractors, and self-organization. It has been shown that many physical, chemical, and biological processes can be modeled this way (Wiggins, 1991).

Neural network models are dynamical models by definition. The state of neurons or neural systems depends on its earlier states, and most models account for this explicitly. These dependencies are generally also non-linear. Activations of neurons are usually a nonlinear function of the activity of connected neurons. This clearly applies to connectionist models applied in developmental psychology.

In this broad definition of nonlinear dynamical systems it is not really possible to view the dynamic approach as a new theory of development. In the broad definition, the dynamic approach amounts to an application of concepts, techniques, and models from this field to developmental psychology. Such work can be carried out from any theoretical perspective.

We have argued in the past that many concepts from dynamical system theory are surprisingly consistent with the old Piagetian theory of development (van der Maas, 1995). Application of nonlinear dynamical system theory may lead to a formalization of Piagetian concepts, such as disequilibrium and equilibration, and not lead to a new theory of development. It might also be fruitful to use nonlinear models to formalize Vygotskian concepts, such as the zone of proximal development, or information processing concepts, such as short-term memory. We do not see this consistency as a disadvantage. New tools, to formalize and test developmental theories, should be more than welcome. We propose to view the dynamic system account as theory-free, in the same way that statistical methods are largely theory-free.

Nonlinear phenomena

Three phenomena of nonlinear dynamical systems are highly relevant to developmental psychology. A well-known phenomenon is deterministic chaos. Deterministic nonlinear dynamical systems may show unpredictable chaotic behavior. One of the paradoxes of developmental psychology is that virtually all developmental psychologist agree that development during infancy is important for later development, but that strong and convincing relationships between developmental measures from infancy and measures of later development are hard to demonstrate (McCall & Carriger, 1993). In theory, deterministic chaos may explain this paradox. It may explain why development in certain respects is so irregular and unpredictable. Unfortunately, it is very difficult to prove this conjecture (certainly the alternative, namely that our measurements are simply too noisy, remains compelling). Developmental data are relatively sparse and often quite unreliable. Attempts to demonstrate chaos in much more frequent and more reliable psycho-physiological time series data have generally failed. The reason is that detection of chaos in time series with noisy measurements turned out to be extremely difficult. Demonstrating chaos in developmental data is therefore hardly feasible.

Another famous phenomenon in nonlinear dynamical systems is self-organization. This refers to spontaneous spatiotemporal coherent behavior of the parts of system in the absence of some pre-specified plan or algorithm that controls this behavior. Such behavior is purely a function of the dynamics of (non-linear) interactions between the system's components. Self-organization has been demonstrated in a variety of physical, chemical, and biological processes. The prototypical example is the Beluzhov-Zabotinsky (BZ) reaction, which displays temporal or spatial self-organization depending on whether the reactors are stirred (Winifree, 1980). As we decribe elsewhere (van der Maas & Hopkins, 1998), we view self-organization as the most promising aspect of nonlinear dynamical system theory. The reason is that self-organization could be a mechanism for structural developmental change, as it is proposed in the Piagetian theory, but also in many other developmental accounts. Self-organization might be the definite resolution of Fodor's famous learning paradox (Raijmakers & Molenaar, 2004).

However, like with chaos, to empirical investigate self-organization is no simple matter. Since neural networks are nonlinear dynamical systems, it should be possible to demonstrate self-organization in the learning process of certain types of neural networks. Demonstrating functional self-organization (Raijmakers & Molenaar, 2004) appears to more difficult than many of us expected. However, phase transitions are critical in self-organizing processes, and thus present an indirect way to study self-organization.

Below we focus on catastrophes or phase transitions. There is a good deal of evidence for phase transitions in psychological processes, including development. We discuss this evidence, and assess how well various computational models, including connectionist models, can account for such data. We also look at alternative neural network models, in which phase transition can be modeled. As it turns out, it is not simple to construct neural models that really show interesting nonlinear properties.

Phase transitions

Developmental psychologists entertain different definitions of discontinuous development. The intuitively simple definition focuses on a relatively large change in a relatively short period. This change can be either an acceleration or a sudden jump. In the latter case the large change occurs instantaneously. We have argued in the past (van der Maas & Molenaar, 1992) that the impasse in the research on discontinuous development (see Brainerd, 1978) is due to this informal definition.

First, the difference between an acceleration and a sudden jump is not trivial. The definition should be clear with respect to this distinction. Accelerations in development are interesting phenomena, but, as explained below, are not discontinuous. Second, limiting the definition to sudden jumps helps, but an empirical problem remains. How can we detect sudden jumps? The fundamental measurement problem is that we can never be sure about what happens between measurement sessions. Suppose we tested children every week, and found that some children changed suddenly from one state to another. We can not exclude the possibility that within this week an acceleration took place. This always remains a possibility, irrespective of the measurement frequency.

Here, the dynamic system approach can help. As we explained above, phase transitions (e.g., bifurcations, catastrophes) are very important in the study of nonlinear dynamical systems. Phase transitions are well understood in this field. They are classified in distinct types (bifurcation types), and many techniques for the analysis and detection of these phenomena exist.

One branch of non-linear dynamical system theory is especially relevant for psychology, and for the social sciences, in general, viz. catastrophe theory (Thom, 1975). From a mathematical point of view, the main ideas of catastrophe theory are now incorporated within bifurcation theory, but from our perspective, this approach is still very relevant. The reason is that catastrophe theory provides possible applications in cases, where one does not have at one’s disposal a mathematical model of a process. These applications involve archetypical models for phase transitions, and strong criteria for phase transitions.

Cusp model

In catastrophe theory discontinuities occur when system states become unstable. A catastrophe is a large sudden jump to a new stable state (attractor), as function of smooth and small changes in independent control variables. Such catastrophe can have different forms. The simplest form is the cusp catastrophe. Figure 2 illustrates the cusp catastrophe with a simple example.

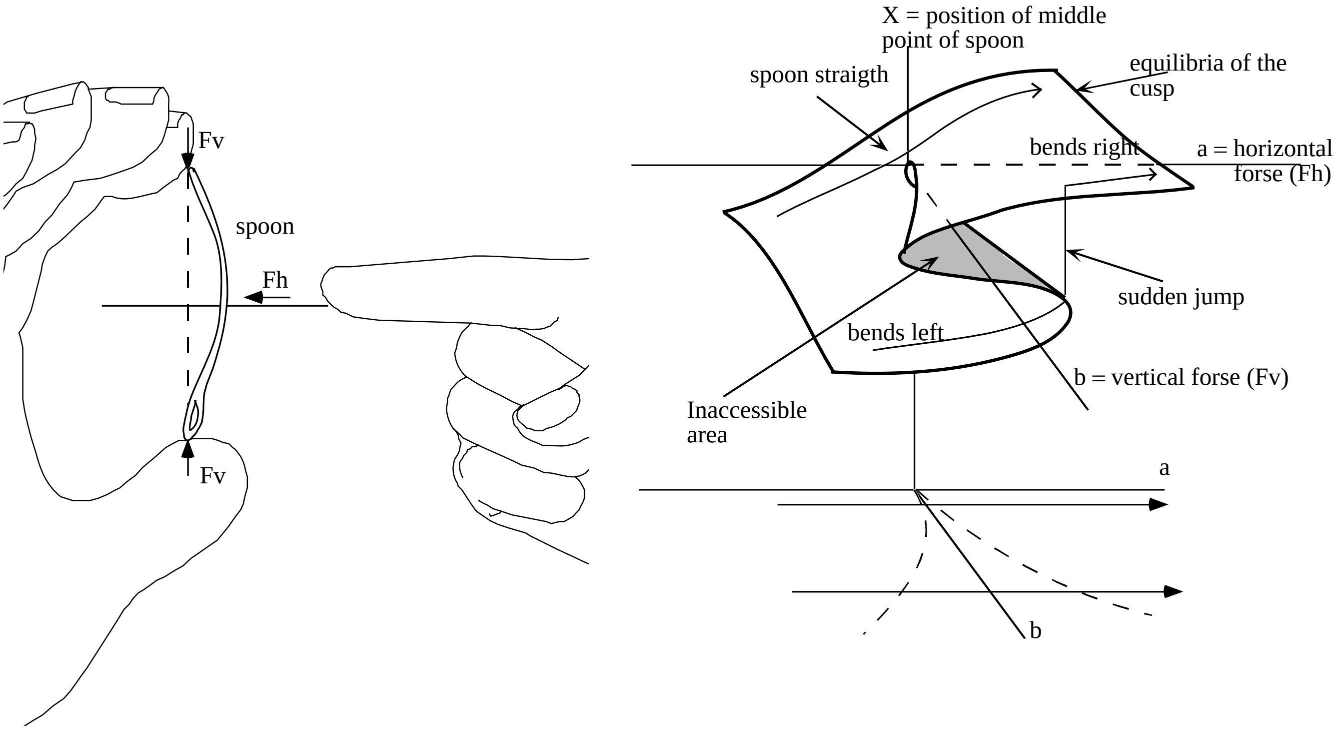


figure 2: The cusp catastrophe models change in X (the position of the middle point of the spoon or stirring stick) as function of changes in the normal factor a (horizontal force, Fh) and the splitting factor b (the vertical pressure on the spoon, Fv). For certain values of a and b, the midpoint of the spoon will suddenly change form left to right, and visa versa. When a is zero and b is high (e.g., only vertical

pressure on the spoon), two attractors exist, separated by an inaccessible unstable state.

A more detailed description can be found in Zeeman (1976), van der Maas, and Molenaar (1992), and Wagenmakers, van der Maas, and Molenaar (2005).

Flags

What makes the cusp model so useful is that we can derive a number of unique predictions of this model in terms of behavioral properties. Gilmore (1981) called these behavioral properties, of which he distinguished eight, catastrophe flags. We will explain five of them with the example in figure 2. First, sudden jumps occur when we exert sufficient vertical pressure on the spoon, while increasing the horizontal force smoothly. At a given moment, the present stable state will become unstable, and the spoon will switch abruptly to the other side. In switching, it will pass the inaccessible state in between the two stable states. This inaccessibility, the second flag, makes this model fundamentally different from models that define discontinuities as accelerations. In such models, all intermediate states are stable. Third, for certain control values, two stable states exists, implying bimodality, which is another important flag. Fourth, when increasing the splitting factor starting at zero (for a = 0), the system has to “choose” between the two emerging states. This is called divergence. Finally, we have hysteresis. This refers to the fact that the sudden jumps from left to right, and visa versa, do not take place at the same values of the horizontal pressure. You have to press relatively hard to force the spoon to jump. A well-known example is the transition from water to ice and visa versa. In shock free condition water freezes at -4^0 Celsius, whereas ice melts at 0^0 Celsius.

Case study: the balance scale

We used the catastrophe flags to study the transition between rules that children apply in for the balance scale task (Jansen and van der Maas, 2001). We focused on the transition from rule I to rule II, as some prior support for this transition had been established. Children, who apply either rule I or rule II respond differently to so-called distance items. These are items, where the weights are the same, and only the distances to the midpoint of the fulcrum differ. Rule I children think that balance scales with such a configuration of weights will remain in balance, whereas Rule II users correctly use the distance information. Rule II users, like rule I children, fail to do so when the number of weights are unequal. The initial evidence for a sudden transition concerns bimodality: scores on sets of distance items are usually strongly bimodally distributed.

Based on the work of Siegler (1976) and Siegler and Chen (1998), we hypothesized that encoding distance information is the relevant normal factor in this case. By changing the distance difference step by step, we tried to find sudden switches between incorrect and correct answer, and possibly also hysteresis. In our experiment, we varied the distance difference from 1 to 5, and back from 5 to 1. In control experiments we reversed these series (with similar results), and checked whether irrelevant changes in the stimuli (color) also caused jumps and hysteresis (however, this was not the case).

The results were as follows. The large majority of children are stable rule I or rule II users. They give the same answer to all items. About 5 % of the children show hysteresis patterns. We tested this in several ways, and although small, this percentage could not be ascribed to chance, i.e., it is a statistical significant effect. In combination with the finding of bimodality, this results support the hypothesis of discontinuous development at the balance scale. However, we did not find divergence. Also a larger percentage of hysteresis patterns would make our point stronger.

Other recent results in balance scale research

To evaluate the connectionist models of the balance scale task, we review some other recent results. In van der Maas, Jansen, and Raijmakers (2004) we give a more comprehensive review. Some doubts have been raised about the rules proposed by Siegler (1976, 1981). Although this does not pose a fundamental problem, additional rules may exist. Many children and adults seem to use compensation rules. They fail to understand the torque rule, and instead compare the sums of weights and distances. More troublesome is that the classification of children into rules with the rule assessment method of Siegler, might be less robust than originally thought. For the rules assessment method, it is essential that only the type of the balance scale item matters, and not the actual number of weights and the distances. Clearly the latter may vary within given item types. Rule I users should fail all conflict distance items, and rule II users should succeed on all distance items, irrespective of the actual values of the weights and distances.

Ferretti and Butterfield (1986) varied the torque differences in balance scale items, and showed that these difference within item types influences rule classification. This finding greatly undermines the rule model, and was used by the connectionist to distinguish their models (that show the TD effect) from symbol-oriented models, and to argue for a more graded non-symbolic interpretation of the balance scale rules. However Jansen and van der Maas (1997) offered two counter arguments. First they re-analyzed the data of Ferretti and Butterfield. Ferretti and

Butterfield used four levels of torque difference (TD). The fourth level is very extreme, and Jansen and van der Maas demonstrated that the TD effect was solely due to this very extreme level, which never actually occurs in any known balance scale test (see figure 3). Second, they improved the rule assessment method by using latent class analysis. This statistical technique makes it possible to test the use of rules statistically. Jansen and van der Maas (1997, 2002) analyzed responses for each item type separately to test the homogeneity of items with item types. They found that in most cases responses appeared to be homogeneous.

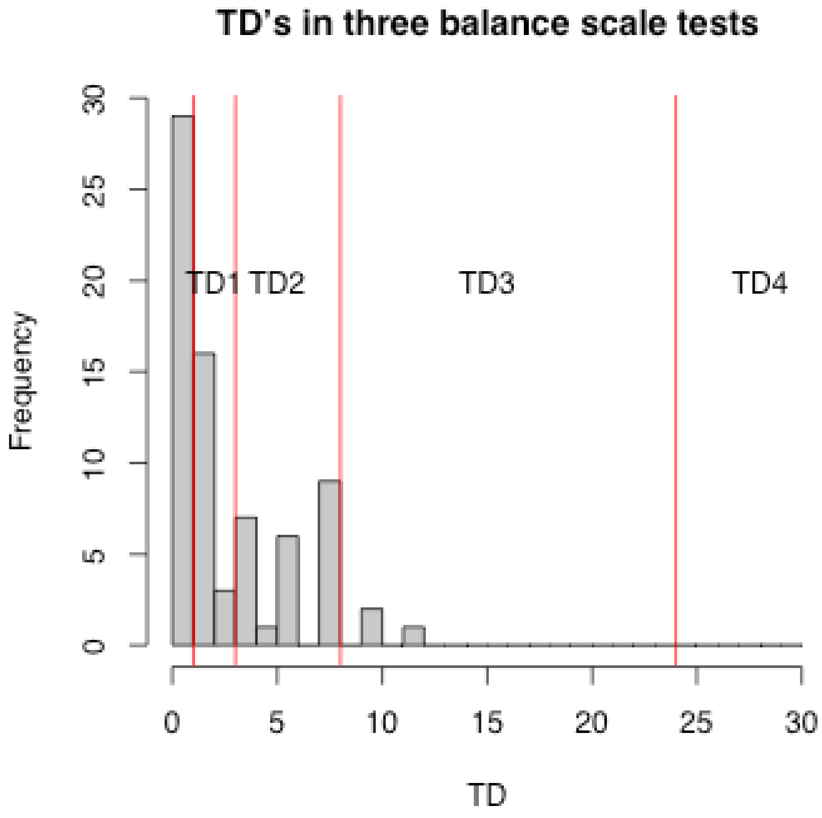


Figure 3: Torque difference effect in the items of three popular different balance scale test (van Maanen, Sijstsma, & Been, 1989; Jansen & van der Maas, 2002;

Siegler, 1981). The vertical lines show the TD levels used by Ferretti and Butterfield. Only for level 4 a TD effect is shown.

Thus, as van Rijn, van Someren, and van der Maas (2003) argue, any account of the data must address the question why children's behavior is homogenous for moderate torque difference levels, and heterogeneous for extreme torque differences.

Finally, van der Maas and Jansen (2003) analyzed response times to balance scale items. They collected data with a computerized test, and analyzed the response times of each rule with linear (mixed effects) models. They showed that the response time results are very well predicted using the rule model of Siegler. With respect to this, we give one example. We used one rather odd item type, which we call the weight distance items. These are items with a larger number of weights on the side with the larger distance. They are odd in that all rules should lead to the correct answer, and so these items have no discriminatory value. However, if we look at response times these items are of interest. According to the rule model of Siegler, subjects using rule I or II will only look at the weights. Since these differ, they immediately choose a response. Older subjects (including under graduates in our sample), using rule III and higher, are supposed to makes additional processing steps. They also first compare the number of weights, but then they compare the distances, and finally they check whether the largest number of weights is on the side with the larger distance. So they make two additional steps, and are thus expected to be slower than the younger children, who use rule I or II. As shown in figure 4, younger children, in spite of the fact that the process information slower, are much faster than the older subjects.

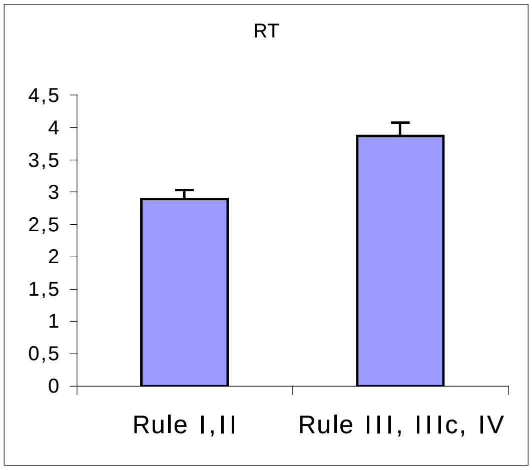


Figure 4: Mean response times to balance items in which the side with the larger distance has more weights (from van der Maas and Jansen, 2003).

Do connectionist models explain the development of proportional reasoning

Van Rijn et al. (2003) give an overview of computational models of the balance scale task. Between 1978 and 1996 several symbolic models were proposed, in the nineties they were followed by connectionist models, and recently van Rijn et al. (2003) proposed an Act-R model. We will not review the symbolic models, but rather focus on the connectionist models, the feedforward backpropagation (PDP) model of McClelland (1989, 1995), and the cascade correlation (CC) model of Shultz, Mareschal and Schmidt (1994).

The first issue concerns the rules. The connectionists employ two arguments. First, when they simulated data with their models, and subjected these data to the rule assessment method of Siegler, they roughly found the same rules, and also the same progression through rules. Second, to explain the discrepancy between perfect rule use and the simulations, they used the torque difference (TD) finding of Ferretti and Butterfield. They argued that only the connectionist models were able to explain the

TD effect. In addition, they use the TD effect to argue that information processing on this task is much more gradual than expected. But as we explained above, the TD effect is less convincing than suggested. It only exists for very extreme torque differences. Models should explain the absence of TD effect for low and moderate TD levels, and the presence of a TD effect for very high levels of TD.

Jansen and van der Maas (1997) simulated data with the PDP model, and analyzed the data with latent class analysis. They compared the results to data of children on exactly the same item set. They showed that the results were very different. Although the latent class models for children did show deviations from the rules, they generally confirm the use of rules. The latent class analysis of the PDP model data was a failure. There were no clear classes, and often fitting latent class model could not be found. Also the data of children appeared to be rather homogeneous within item types (no TD effects), whereas the PDP data were very heterogeneous.

Recently, Quinlan, van der Maas, Jansen, Booij, and Rendall (2006) reported a similar analysis of the CC network. The results of the CC model resemble more closely the empirical results. Given data generated by these models, well fitting latent class models were found, and some of the expected rules seem to be present. Support for rules I and II and weighted addition rules was established. Human subjects, in contrast to the CC model, use un-weighted addition and rule IV, i.e., the correct rule. In latent class analysis, the CC model was found to employ rules (e.g., always right side of the balance) that are not observed in empirical data. Especially troublesome is that even after extensive learning, the CC model, like the PDP model, fails to use the correct torque rule. [1]

The second issue is discontinuous development. Raijmakers, van Koten, and Molenaar (1996) showed that the PDP model develops continuously. In the original analysis of the PDP (and the CC) model, response patterns were first classified with the rule assessment method, and then these rule were plotted against time (epochs). Such a plot strongly suggests transitions and stages. Raijmakers, van Koten, and Molenaar (1996) demonstrated that this is an artifact of the data-analytic procedure. The raw score on each item type should show discontinuities, but these turn out to be very gradual. Also no evidence for bimodality could be found.

The CC model does a better job. In his book, Shultz (2003) showed that the CC model exhibits strong accelerations in performance. Since this model also showed evidence for rule I and II in the latent class analysis, it will also exhibit bimodality on distance items. We did not did perform a hysteresis experiment with either the PDP model or the CC model. One problem is that in this experiment items are administered without feedback. But without feedback, CC and PDP models will be subject to any change in their weights. Learning only takes place by feedback in these connectionist models. For that reason we do not see how connectionist networks could show hysteresis. Furthermore, these networks do not show real phase transition or self-organization. The accelerations in the CC model are due to the addition of hidden nodes when learning falters. Such accelerations are unrelated to non-linear self-

---

[1] The reason for this failure probably lies in the fact that the activation rules of these networks are additive. The networks have many free parameters so that in a limited item set they may mimic multiplication with weighted addition, but they will never generalize to all balance scale items. See also Dawson and Zimmerman, 2003.

organization. The ACT-R model of van Rijn et al. (2003) does provide an account for the hysteresis effect, although not in terms of self-organization.

New simulations might be in order to test hysteresis in connectionist networks. This is certainly the case with the response time data reported by van der Maas & Jansen (2003). Both the ACT-R model and the connectionist models are able to make predictions about response times. These repose time results constitute an new challenge for computational modeling of proportional reasoning.

Our present conclusion about the connectionist models is that they are largely unconvincing as models of proportional reasoning, and more generally as models of higher cognitive functioning. Connectionists downplay the role of rules, reasoning, and representations, but at the same time demonstrate that they can approximately simulate the main empirical phenomena. Indeed, children are not perfect rule users, and some aspect of the reasoning (rule III, for instance) may be best described with statistical rules, as implemented in connectionist models. But overall these connectionist models are just too simple, they miss essential computational properties, and the evidence for genuine rules use by children is stronger than expected.

Can neural networks model higher cognitive processes?

As we stated in the introduction the real challenge for psychology lies in typical human abilities. The main successes of the connectionist approach do not lie here, but rather they can be found in other fields, such as memory and perception. As clear from the above, we are not impressed with the connectionist models of higher reasoning processes. Does this mean that we should abandon this approach?

We do not think so. First, neuronal processes evidently form the basis for all our information processing, including formal abstract thinking. Second, the connectionist models have been subject to development over a period of 10 to 15 years old. The field of neural network modeling has expanded in many directions, and at present there certainly are better types of networks, which can be considered.

Below we briefly discuss two important general lines of research. The first concerns the coupling of neural models with symbolic models, and the second concerns the utilization of the nonlinear properties of networks.

Hybrid models

The ACT-R model proposed by van Rijn, van der Maas, and van Someren (2003) has proved to be quite successful in simulating the main empirical phenomena in the area of the balance scale. It applies the ACT-R architecture, which combines symbolic and non-symbolic aspects of information processing. More specifically, production rules are accorded activation values, which change in a gradual fashion, like weights of network connections. The most important idea in the ACT-R model is that all balance scale rules can be based on one more general rule: search for a difference between the right and left side of the scale. If no difference can be found, respond “balance”. The only difference young children apperceive is a difference in weights. They simply do not see the difference on distance items (i.e., items, in which only the distances differ). Somewhat later in development children also apperceive the difference in distance, a difference in sum, and, finally proceed to a difference in product of weight and distance. In the ACT-R model changes in knowledge outside the domain of proportional reasoning play a role (like learning to multiply, and increases in capacity of short term memory) as well as certain ACT-R mechanisms (production rule compilation).

The ACT-R model is not really a hybrid model. Such models explicitly combine neural and symbolic properties. Examples can be found in Wermter and Sun (2000). A related line of research is reviewed in Jacobsson (2005), which discusses techniques for rule extraction form neural networks. Another interesting line is the work on symbolic computation in stochastic neural networks (Tabor, 2002)

Neural networks with functional self-organization

In the second line of research, the focus is on constructive neural networks. Raijmakers and Molenaar (2004) discuss the learning paradox, put forward by Fodor (1980). The point of this paradox is that learning new concepts necessarily involves composites of innate primitive concepts, since learning is a form of inductive inference involving hypothesis formation and confirmation. Learning fundamentally new concepts seems impossible. The addition of conceptual resources is the only way that a system can become more powerful. If we want to present neural networks as viable models of higher cognition, we will have to show how such models can solve this paradox, and thus how such models can acquire more powerful concepts

Generative neural networks, such as cascade correlation networks, are proposed as constructive neural models that escape the learning paradox (Quartz, 1993). Raijmakers (1997) criticized this claim. She argues that additions of new hidden units are discrete events, i.e., qualitative changes in the network, which represent additions of conceptual resources. It would be more convincing if arbitrary small quantitative changes in the network (say, a small change in learning rate) led to qualitative changes in the network.

As Raijmakers and Molenaar (2004) argue, this requires phase transitions in network behavior. Ideally we would observe continuous changes in model parameters (learning rate, connections strength, range of connections, etc.) leading to sudden transitions to new learning regimes, in which more powerful concepts are mastered (Molenaar & Raijmakers, 2000). Examples of work in this line can be found in Pollack (1991), Rodriguez, Wiles, and Elman (1999), and van Ooyen et al. (1995).

Raijmakers and Molenaar (2004), and Raijmakers, van der Maas, and Molenaar (1997) present evidence for phase transitions in a specific ART (Adaptive Resonance Theory, Grossberg, 1976) network. In this network, Exact ART, phase transitions occur between two learning regimes, based on localized and distributed representation, as function of small changes in the range of inhibitory and excitatory connections. These two types of representation in neural networks form the basis of very different types of computation (Page, 2000).

This is just one example of work in this line of research. As neural networks are nonlinear dynamical systems, they will display all kinds of nonlinear phenomena (such as chaos, bifurcations and self-organization). The construction of networks, in which such properties have a function in learning more powerful concepts, is not easy (Tabor, 2002), but, in our view, it is the most promising way to proceed.

Connectionist and Dynamic Systems Approaches to Development: On the Cusp of a New Grand Theory or Still Too Distributed?

The title of this section is the title of the symposium that led to this book. In this section we summarize our point of view. We adopt the broad definitions of the connectionist and the dynamic approach. Connectionist models are dynamic system models, and dynamical system models are applicable to all kinds of theories and models.

In our view, the most promising part of dynamic system theory concerns phase transitions and self-organization, which we study empirically and in simulation research. We are not impressed with the classical connectionist networks, because they fail to model higher cognitive processes, and because they do not show interesting dynamical behavior. The ultimate challenge is to construct neural networks with nonlinear properties that allow for functional self-organization via phase transitions. Clearly, the networks we propose also fail as model of higher cognitive processes, such as proportional reasoning. We can only admit this, and we do not exclude the possibility that at least in the short run, more yields can be expected from (sub-) symbolic approaches, such as ACT-R and hybrid models.

We argued that there is no solution in downplaying the importance of higher cognitive functioning. Our studies of the balance scale showed us that children often use rules that strongly resemble the traditional production rules (if ‘equal weight’ then ‘balance’). At the same time, there is evidence for graded processing (e.g., extreme TD effects, violations of rules, rule switching). The real challenge is to develop neural network models that capture both phenomena. We choose for a very fundamental research line, in which we focus on self-organizing phase transitions in neural networks.

References

Beer, N. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Science, 4,* 91-99.

Brainerd, C.J. (1978) The stage question in cognitive-developmental theory. *Behavioral and Brain Sciences, 2*, 173-213.

Dawson, M. R. W., & Zimmerman, C. (2003). Interpreting the internal structure of a connectionist model of the balance scale task. *Brain and Mind, 4*, 129-149.

Elman, J. L. (1990) Finding structure in time. *Cognitive Science* 14:179–211.

Fahlman, S. E. (1988). *An empirical study of learning speed in back-propagation networks*. (Tech. Rep. CMU-CS-88-162). Pittsburgh, USA: Carnegie-Mellon University, Department of Computer Science.

Ferretti, R. P., & Butterfield, E. C. (1986). Are childrens' rule-assessment classifications invariant across instances of problem types? *Child Development*, *57*, 1419-1428.

Fodor, J.A. (1980). On the impossibility of acquiring more powerful structures. In M. Piattelli-Palmarini, J. Piaget & N. Chomsky (Eds.), *Language and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 142–162). Cambridge, MA: Harvard University Press.

Gilmore, R. (1981) Catastrophe theory for scientists and engineers. John Wiley & Sons, New York.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: feedback, expectation, olfaction and illusions. *Biological Cybernetics*, **23**, 187–202.

Jansen, B. R. J., & van der Maas, H. L. J. (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review*, *17*, 321-357.

Jansen, B. R. J., & van der Maas, H. L. J. (2001). Evidence for the phase transition from Rule I to Rule II on the balance scale task. *Developmental Review, 21,* 450-494.

Jansen, B. R. J., & van der Maas, H. L. J. (2002). The development of children’s rule use on the balance scale task. *Journal of Experimental Child Psychology, 81*, 383-416.

Jansen, B.R.J., & Van der Maas, H.L.J. (2001). Evidence for the phase transition from rule I to rule II on the balance scale task. *Developmental Review*, **21**, 450–494.

Lewis, M.D. (2000). The promise of dynamic systems approaches for an integrated account of human development. Child Development, 71, 36–43.

Mareschal, D., & Shultz, T.R. (1996). Generative connectionist networks and constructivist cognitive development. *Cognitive Development*, **11**, 571–603.

McCall, R. B., & Carriger, M. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development, 64*, 57-79.

McClelland, J. L. (1989). Parallel distributed processing: implications for cognition and development. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 8-45). Oxford: Clarendon Press.

McClelland, J. L. (1995). A connectionist perspective on knowledge and development. In T. J. Simon & G. S. Halford (Eds.), Developing cognitive competence: New approaches to process modelling (pp. 157-204). Hillsdale, NJ: LEA.

McClelland, J. L. & Rumelhart, D. E., eds. (1986) Parallel distributed processing: Explorations in the microstructure of cognition, vol. 2: Psychological and biological models. MIT Press.

Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A PDP model of the AB task. Developmental Science, 1, 161-184.

Page, M. (2000) Connectionist modelling in psychology: A localist manifesto. Behavioral and Brain Sciences23(3).19:394–98.

Ploeger, A., van der Maas, H. L. J., & Raijmakers, M. E. J. (2006). Can evolutionary psychology contribute to the basic issues in developmental psychology? A view from evolutionary developmental biology. Psychological Bulletin, in revision.

Pollack, J. B. (1991) The induction of dynamical recognizers. *Machine Learning,* 7:227–52.

Quartz, S.R. (1993). Neural networks, nativism, and the plausibility of constructivism. *Cognition*, *48*, 223–242.

Raijmakers, M. E. J. (1997). Is the learning paradox resolved? *Behavioral and Brain Sciences, 20,* 573-574.

Raijmakers, M. E. J., & Molenaar, P. C. M. (2004). Modelling developmental transitions in adaptive resonance theory. *Developmental Science, 7,* 149-157.

Raijmakers, M.E.J., van der Maas, H.L.J., & Molenaar, P.C.M. (1997). Numerical bifurcation analysis of distancedependent on-center off-surround shunting neural networks. *Biological Cybernetics*, **75**, 495–507.

Raijmakers, M.E.J., van Koten, S., & Molenaar, P.C.M. (1996). On the validity of simulating stagewise development by means of PDP networks: application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science*, **20**, 101–136.

Rodriguez, P. Wiles, J. & Elman. J. L. (1999) A RNN that learns to count. *Connection Science*, 11, 4–40, 1999.

Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. Child Development, 74(5), 1393-1417.

Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Shultz, T. R., Mareschal, D., & Schmidt, W. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning, 16*, 57-86.

Shultz, T.R., Mareschal, D., & Schmidt, C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning*, **16**, 57–86.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8,* 481-520.

Siegler, R. S. (1981). Developmental sequences between and within concepts. *Monographs of the Society for Research in Child Development, 46,* Whole No. 189.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: a microgenetic analysis. *Cognitive Psychology*, *36,* 273-310.

Spencer, J. P. & Schöner, G. (2003) Bridging the representational gap in the dynamic systems approach to development. *Developmental Science* 6(4):392–412.

Tabor, W. (2002). The Value of Symbolic Computation. *Ecological Psychology, 14*(1/2): 21-51.

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

Thom, R. (1975). *Structural stability and morphogenesis*. Reading, MA: Benjamin.

van der Maas, H. L. J.(1995). Beyond the metaphor? Cognitive Development, 10, 621-642.

van der Maas, H. L. J., & Raijmakers, M. E. J. (2000). A phase transition model for mother-child interaction: Comment on Olthof et al., 2000. *Infant and Child Development*, *9*(2), 75-83.

Van der Maas, H. L. J. & Jansen, B. R. J. (2003). What response times tell of children's behavior on the balance scale task. *Journal of Experimental Child Psychology, 85*, 141-177.

Van der Maas, H. L. J., & Hopkins, B. (1998). Developmental transitions: So what's new? *British Journal of Developmental Psychology*, 16, 1-13.

Van der Maas, H.L., & Molenaar, P.C. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological Review*, **99**, 395–417.

Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98*, 3-53.

Van Maanen, L., Been, P., & Sijtsma, K. (1989). The linear logistic test model and heterogeneity of cognitive strategies. In E. E. Rosram (Ed.), *Mathematical psychology in progression* (pp. 267-287). Berlin: Springer-Verlag.

Van Ooyen, A., Van Pelt, J., & Corner, M. A. (1995). Implications of activity dependent neurite outgrowth for neuronal morphology and network development. *Journal of Theoretical Biology*, **172**, 63–82.

Van Rijn, H., van Someren, M., & van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, *27*, 227-257.

Wagenmakers, E. J., Molenaar, P. C. M., Grasman, R. P. P. P., Hartelman, P. A. I., & van der Maas, H. L. J. (2005). Transformation invariant stochastic catastrophe theory. *Physica D, 211*, 263-276.

Wagenmakers, E.-J., van der Maas, H. L. J., & Molenaar, P. C. M. (2005). Fitting the cusp catastrophe model. In B. Everitt & D. Howel (Eds.), Encyclopedia of Statistics in Behavioral Science, Vol. 1, pp. 234–239. New York: Wiley.

Wiggins, S. (1990). Introduction to Applied Nonlinear Dynamical Systems and Chaos. NY: Springer-Verlag.

Wermter, S., & Sun, R, (2000). Hybrid neural systems, Heidelberg: Springer

Winifree, A.(1980) The Geometry of Biological Time. Wiley, New York.

Zeeman, E.C. (1976) Catastrophe Theory. Scientific American, 234(4), 65-83.

Connectionist and Dynamic Systems models of development:

The case of analogical completion

Denis Mareschal, Robert Leech & Richard P. Cooper

School of Psychology

Birkbeck University of London

Abstract

Connectionist and dynamic systems are two of the most recent causal accounts of development. Both frameworks provide mechanistic frameworks for considering how behaviour emerges over time. The dynamic systems framework emphasises the immediate online emergence of behaviour whereas the connectionist framework emphasis emergence of behaviour across longer time scales. A close examination of the mathematics used to describe the mechanistic theories that underlie these frameworks shows that they are in fact two complementary sides of the same coin. We illustrate this by describing a simple connectionist model of the development of analogical completing in young children that builds on the connectionist notion of gradual accumulation of knowledge and the dynamic systems notion of online instabilities to explain how analogy completion develops. We argue that this model provides a better account of a broad range of children's analogical completion than other existing models.

There are two sides to any coin

In 2002 the journal Developmental Science devoted an entire special issue to exploring two new theoretical frameworks for understanding development (*Developmental Science, 5*, 2002: Dynamic Systems and Connectionism). What makes these two frameworks stand out from other contemporary theories of development in the literature is the central role that mechanisms of change play. In other words, both approaches are fundamentally interested in is understanding how change occurs in developmentally relevant phenomena.

To many outsiders, both approaches might appear to share a number of at least superficial similarities. They both emphasise the use of computer modelling as a research tool and both use mathematics to describe the internal processes of their models. Although such methods are far more common in psychology now than they used to be even 15 years ago (especially in the study of adult cognition and perception), they are still relatively rare in the study of developmental psychology.

Both approaches also appear to deny the sequential processing assumptions of cognition that underlie many of the early information processing theories. That is, instead of viewing cognition as the outcome of a sequential series of syntactic steps operating on symbolic representations (e.g., Pylyshyn, 1984; Sternberg, 1969) the emphasis is on parallel processing of multiple sources of information (e.g., Rumelhart & McClelland, 1986). Finally, both approaches also reject the fundamental notions of the Language of Thought hypothesis (Fodor, 1975). That is, they reject the idea that cognition consists in the syntactic manipulation of symbolic tokens.

However, a closer look suggests that there are also some apparently fundamental differences between the two approaches. Connectionist systems are deeply committed to the notion of representations, and emphasise how « knowledge » that is represented in active or latent form (e.g., Munakata & McClelland, 2003) constrains emergent behaviours. A look back at the history of connectionist and neural network modelling shows how much effort has gone into exploring how information about the world is internally represented in such systems (e.g., Feldman & Ballard, 1982). This may be because the first applications of neural network that would later become the mainstay of the connectionist approaches were intended to model memory systems (e.g., Rosenblatt, 1958; Hinton & Anderson 1981; see also Anderson & Rosenfeld 1989 for a detailed history of connectionist neural computation research). It is precisely this emphasises on the acquisition of task appropriate representations that drew developmental psychologists to connectionism (e.g., Plunkett & Sinha, 1991). Because of their history, connectionist systems have also emphasised the importance of stable states. A memory is said to have been retrieved when the network had stabilised in a state that was consistent with the information contained in the cues. Dynamic processes were involved in the retrieval of the memory but what was really important was the end state.

In contrast, Dynamic Systems approaches have their origin in the study of control, and in particular how neural networks could be used as control systems for locomotion (e.g., Barto, Sutton & Anderson, 1983). Through the 1970s and 1980s, the problem of motor control was approached in the way that many other cognitive problems were approached: that is, through the use of symbolic serial processing (Boden, 1988). However, it soon became clear that at least in the domain of motor control serial processing was in fact an illusion and that stable states were only fleetingly present if ever. In reality, the sensory motor system is a continuously adapting system, closely coupled to the environment (Ballard et al., 1997). This emphasis on a close coupling between control systems and environment was immensely successful in producing models of sensorimotor responses that worked both as engineering projects (e.g., could control a robot in the real world) and as explanatory frameworks for explaining human locomotion (e.g. Jordan, 1989). These successes have led developmentalists in the Dynamic Systems camp to deny a role for representations and instead to focus on process as the key factor in explaining behaviour (e.g., Smith & Samuelson, 2003; Thelen & Smith, 1994).

So, while Connectionism and Dynamic Systems share some fundamental similarities when viewed from a distance, there are at least two core elements of the frameworks that seem to put them in opposition. First, Connectionism emphasises representations whereas Dynamic Systems deny that representations play any role in the generation or control of behaviour. Second, Connectionism emphasises the stable states of a system whereas Dynamic Systems emphasises the transition states of the system (even suggesting that stable states are illusory). The central theme of this volume it to explore these oppositions.

In fact, we believe that while these are theoretically fundamental differences, a closer look at the equations that underlie the formal descriptions of each approach suggests that these differences may be differences in emphasis rather than differences in kind, and that, at least in the context of cognitive development, Connectionism and Dynamic Systems are in fact two sides of the same coin. To illustrate this consider the equations used to implement the common dynamical field theory version of Dynamic Systems (e.g., Thelen et al, 2001). These are generalised continuous versions of the equations used to describe interactive activation networks (Amari, 1977, 1980; Grossberg, 1980; Rumelhart & McClelland, 1986). Thus, while the conceptual emphasis is placed differently, the two approaches are in fact speaking of the same thing.

It is therefore very possible that the visions are complementary and that both visions can contribute to a better understanding of behaviour as it unfolds through instantaneous time and developmental time. To illustrate this, we will describe a recent connectionist model of the development of analogical completion in early childhood which draws on both the gradual learning and distributed representations of connectionist networks, and the notion of instantaneous settling into transient attractors of dynamic systems to capture the fine-grained performance of children at different ages.

The case of analogical completion

Analogical reasoning has traditionally been the preserve of high-level models using symbolic representations and explicit structure mapping (French, 2002). Such models have tended to view development as something of an afterthought. However, the more developmentally-inspired model presented here serves not only to emphasize the importance of development as a major constraint on cognitive theories, but it also suggests the compatibility of connectionist and dynamic systems accounts of cognitive development. Analogy arises both from how connection weights change over training but importantly how this interacts with how the network settles over time into attractor states, and equally importantly, how these attractor states subsequently change when perturbed by a different input.

Our model is designed to simulate young children's performance on *a:b::c:d* analogies involving simple causal domains, e.g., *apple is to cut apple as bread is to what?* (*cut bread*) (see Goswami and Brown, 1989; 1990; Rattermann and Gentner, 1998).

In the model, analogy is viewed as a by-product of relational priming in a simple memory system. Relations are represented implicitly as transformations between states (e.g., the relation *cutting* equates to the transformation from *apple* to *cut apple*) (Thomas and Mareschal, 1997; Rogers and McClelland, 2004). Under our account, analogical completion occurs though the *a* and *b* terms of the analogy priming a relation which then biases the *c* term so that the pattern completion properties of the network produces the analogically appropriate *d* term. The model demonstrates the complementarity of connectionist and dynamic systems, simultaneously employing both gradual connection-weight-based acquisition of relational knowledge as well as instantaneous settling into attractor states leading to analogical completion. The model stands up well to comparison with a set of key behavioral phenomena from the developmental analogical reasoning literature, as listed below. Some of these phenomena are currently beyond the scope of many non-developmental accounts.

1) One of the central findings in previous experimental research is the strong relationship between the development of relational knowledge and the acquisition of analogical reasoning (e.g., Goswami and Brown, 1989; Rattermann and Gentner, 1998). That is, children can only form and comprehend analogies involving relations they are, to some degree, familiar with.

2) One corollary of the importance of relational knowledge accretion is that acquisition of analogical reasoning develops at a different pace in different domains, and does not result from a domain general change (Goswami and Brown, 1989).

3) The existing developmental literature suggests that the ability to reason by analogy develops spontaneously. Children regularly form analogies in many different domains without the need for any formal tuition (Pauen and Wilkening, 1997; Ingaki and Hatano, 1987; Tunteler and Resing, 2002).

4) Gentner and colleagues (Gentner and Toupin, 1986; Gentner, 1988; Rattermann and Gentner, 1998) have repeatedly found evidence that a relational shift occurs over development. Children gradually change from initially making analogical judgements on the basis of object similarity (i.e., superficial or perceptual features of an object) to increasingly relying on relational similarity.

5) Young children can solve so-called cross-mapping analogies, where there is a direct conflict between relational and object responses (Gentner et al, 1995; Kotovsky and Gentner, 1994). In a cross-mapped analogy the same object appears in both the base and the target but with a different role, so the correct response involves disregarding the perceptually identical object. Children find cross-mapped analogies harder and solve them later in development than other analogies.

6) The developmental literature (e.g., Kotovsky and Gentner, 1994; Rattermann et al, 1994; Lowenstein and Gentner, 2005) has repeatedly shown that the analogical abilities of children can be improved by prior training with or exposure to relational labels.

7) Hosenfeld et al (1997) have directly compared the development of analogical reasoning directly with predictions derived from dynamic systems theory. They found

evidence consistent with three markers of a discontinuous change: (i) a sudden jump in accuracy; (ii) a critical slowing down in the neighborhood of the sudden jump; and (iii) an increase in inconsistency, also accompanying the sudden jump.

A developmental model

Figure 1 shows the general architecture of the network used to model both the acquisition of relational information as well as analogical completion (Leech, Mareschal, & Cooper, 2001). The network weights are bi-directional and symmetrical. The bottom layer (roughly corresponding to the input layer) is split into two banks of units, representing the presentation of two different objects in a ‘before’, or pre-transformation state. Similarly, the upper layer (corresponding to the output layer) is split into two banks, representing the same two objects in their ‘after’ or post-transformation state. The ‘before’ and ‘after’ representations can be thought of as two temporally contiguous states of the world. A relation is conceived of as the transformation from the ‘before’ to the ‘after’ state. The Object (t1) layer corresponds to some object (e.g., *apple*) and the CA (t1) layer corresponds to a causal agent (e.g., *knife*), which is the object that causes the transformation (in this case *cutting*) to occur (another useful way to conceptualize CA (t1) and CA (t2) is as context layers). The activation of any unit varies according to a sigmoidal activation function from 0 to 1. The initial weights are uniformly randomized between ±0.5.

[INSERT FIGURE 1]

Because of the bi-directional connections, input activation cycles through the network before settling into an attractor state. During training, contrastive Hebbian learning is used to change the connection weights so the attractors on the output units approximated target output states of the network (O'Reilly, 1996; O'Reilly and Munakata, 2000). The learning rate is uniformly set to 0.1, and activations are updated for 5 cycles between each weight update. Additionally, the bi-directional connections are central to the priming mechanism used in the model. Priming occurs because the bi-directional connections allow units which are not clamped to a specified activation value to maintain activity resulting from an earlier event. The prior activity that is maintained across the network impacts how subsequent input is then processed (see below).

The network is trained on input patterns produced on-the-fly by adding Gaussian noise (mean = 0.0, variance = 0.1) to 20 different prototype Object (t1) patterns (these are randomly generated input vectors within the range [0, 1]) as well as 4 different CA (t1) (causal agent) prototype patterns. This was intended to capture the fact that although two instances of cutting an apple with a knife are similar, they are never identical.

Four transformation vectors were also randomly generated. The transformation vectors encode the relation between the pre- and post transformation states of the object. In fact, the transformed state of the object (Object (t2)) is obtained by adding a transformation vector to an object (Object (t1))[1]. For example, Object (t1) (e.g., *apple*): [0.5 0.2 0.8 0.2 0.4], might be transformed by the vector (e.g., *cut*ting): [-0.4 0.0 0.0 0.0.7 0.0] resulting in Object (t2) (*cut apple*): [0.1 0.2 0.8 0.9 0.4]. Note that the transformation vector is only implicit in the training regime, only Object (t1) and a transformed Object

---

[1] CA (t2) is always identical to CA (t1).

(t2) are presented to the network, not the actual transformation vector itself. Different objects (e.g. *bread* or *apple*) transformed by the same relation (e.g. *cutting*) are transformed by the same vector. Thus, the network generalizes across objects on the basis of shared transformations tied to particular casual agents.

Training the network

Training consists in randomly selecting an object (Object (t1)) and a causal agent (CA (t1)), clamping Object (t1) and CA (t1) to these patterns and then letting activation cycle round the network 5 times (this is called the minus phase). The process is then repeated with the additional Object (t2) and CA (t2) target patterns also clamped on (this is the plus phase). The weights are then updated according to the difference in Hebbian terms corresponding to the plus and minus phases. Because both the pre and post transformation states of the network could be obtained by direct observation of the world, learning of relational information constitutes a form of self-supervised learning (Japkowicz, 2001).

In our model, each of the 20 Object (t1) objects is affected by 2 of the 4 causal agents and so 2 of the 4 transformations. Thus, there are 360 potential analogies for the network to be tested on. However, when an object is presented in conjunction with one of the remaining 2 causal agents, the target Object (t2) is the same as the untransformed Object (t1) (i.e., there is no transformation). Thus, the presence of the causal agent alone does not predict that a transformation will occur.

Testing the network

Analogical completion is a separate process from the learning of relation information, arising from the different way information is presented to the network. For analogical completion the activation pattern corresponding to *apple* is clamped on at Object (t1) and *cut apple* at Object (t2) while CA (t1) and CA (t2) are initially set to 0.5, the resting value. This is analogous to presenting the network with *apple*:*cut apple* (i.e., the first half of an *a:b::c:d* analogy). The network subsequently settles into an attractor by filling in CA (t1) and CA (t2) and arriving at hidden unit activations consistent with the transformation *cutting*. Following this, the Object (t1) and Object (t2) units are unclamped and a second pattern, corresponding to *bread*, is presented to Object (t1) and nothing presented to Object (t2). This is analogous to presenting the network with the activation pattern *bread:?* (i.e., the second probe-half of the *a:b::c:d* analogy). Thus, the network is no longer in equilibrium and settles into a new attractor state. However, the prior priming of the transformation *cutting* in the context of *apple* biases the network to settle into the attractor state corresponding to the transformation *cutting*, resulting in the *cut bread* pattern at Object (t2) (i.e., the analogically appropriate response).

How the network arrives at a response is an indication of the important role dynamic attractors play in analogical completion in the model. Figure 2 shows, at three different stages of training, the sum-of-squared distance (SSD) between the actual output of the network when tested on the *bread:cut bread:apple:..?* analogy and four possible trained Object (t2) patterns as activation cycles 5 times. The target object patterns correspond to those used in Goswami and Brown (1989). The lower the y-axis value, the closer the actual activation is to that possible output pattern. Consequently, Figure 2 shows which of the four objects that the network has been trained on is closest to the network's actual response after different amounts of training.

[INSERT FIGURE 2]

Performance of the model

After 100 epochs of training (Figure 2a), the network fails to complete the analogy appropriately. Instead its output is closest to that for an object (i.e., perceptual) similarity match (i.e., *apple*). After 1,500 epochs of training (Figure 2b), the network's response is ambiguous, whereas after 5,000 epochs of training (Figure 2c), the network settles into the appropriate state to complete the analogy (i.e., *cut apple*). However, to infer correct analogical completion, it is also important that the network does not produce an analogical response when it is *not* appropriate. Consistent with the results in Figure 2(a-c), it could be the case that the attractor basin corresponding to *cut apple* is so wide that *cut apple* is always the network's response. However, this is not the case as Figure 2(d) indicates. Here, the network is presented with a non-analogy and produces, appropriately, a non-analogical response.

The model is assessed against the seven key phenomena from the developmental literature listed earlier. These are considered in turn:

(1) *Relational knowledge accretion.* Figure 3 contrasts the percent of analogies completed appropriately by the network over training with the mean sum of squared error at Object (t2) (a measure of how well the network has mastered the causal domain on which it is trained). The proportion of analogies correct and sum of squared error are strongly negatively correlated ($r = -0.97$, $p < 0.001$). A similar demonstration of the interrelatedness between analogical completion and relational knowledge occurs by contrasting successful analogy completion with how well the network 'extracts' the causal agent (i.e., presenting the network with *apple* and *cut apple* and seeing what causal agent the network produces at

CA (t1) and CA (t2)).This ability is important because it demonstrates that the analogical completion observed in the network is not simply a matter of forming a simple input-output (or stimulus-response) link. Successful production of the appropriate causal agent correlates strongly with analogical completion ($r = 0.93$, $p < 0.001$). This correlation mirrors results obtained from a similar experimental comparison by Goswami and Brown (1989) with young school-age children.

[INSERT FIGURE 3]

(2) *Domain specificity*. In a related vein, the performance of the network also demonstrates that analogical reasoning depends on underlying domain specific relational knowledge. This again mirrors the domain specific – as opposed to domain general – change over development observed in children's ability to form analogies (Goswami and Brown, 1989).

(3) *Spontaneity*. Since the network is never trained on analogical completion, the analogical ability of the network is consistent with the reported spontaneous use of analogy by children (Pauen and Wilkening, 1997; Ingaki and Hatano, 1987; Tunteler and Resing, 2002).

(4) *Relational shift*. The network, like children, shows a 'relational shift' (changing from object to relational responding). Table 1 compares the response profile of children of different ages (Rattermann and Gentner, 1998) on the target *a:b::c:d* task with the network at distinct stages in training. We see that there is a close parallel between network and child error rates. This profile suggests that the network, like the children, undergoes a change from responding on the basis of

object similarity to relational similarity (i.e., correct responses and wrong transformation responses).

[INSERT TABLE 1]

(5) *Cross-mapped analogies*. Finally, both children (Gentner et al., 1995), and the network, can solve cross-mapping analogies where there is a strong conflict between object and relational similarity. The network requires more training before solving these analogies, reflecting the finding that these analogies are harder for children than analogies where object similarity aids the relational analogy (Gentner et al., 1995).

(6) *Effect of relational labels.* We contrasted analogical completion in our original model with a modified version which included an additional orthogonal label cue, presented to the network in addition to the causal agent. The analogical completion performance of the modified model follows a very similar developmental trajectory to the original model, just shifted earlier in training. Thus the model mirrors the observed augmenting effect of training with relational labels on children's analogical development (Kotovsky and Gentner, 1994; Rattermann et al., 1994; Lowenstein and Gentner, 2005).

(7) *Indicators of a discontinuous change.* Figures 4(a-c) indicate that our model also demonstrates developmental markers of discontinuous change (Hosenfeld et al, 1997) similar to those witnessed in children's development. Figure 4(a) shows that a sudden jump occurs at around 2,400 epochs of training. Figures 4(b) suggests that at the same point there is a peak in inconsistent responses (as

measured by presenting the same network twice with noisy versions of the same input). Figure 4(c) similarly suggests that alongside the sudden jump the network takes longer (i.e., slows down, as measured in mean activation cycles) to produce its final response.

[INSERT FIGURE 4]

Discussion

We have seen that the development of simple analogical abilities, as demonstrated by children's changing performance on *a:b::c:d* type analogies, may be simulated by a model that employs relational priming to arrive at the *d* term of an analogy given the *a:b::c:* part of the analogy. The simulations capture seven key phenomena apparent in children's performance over time. The model (and underlying theory) differs substantial from other models and theories of analogical reasoning. We therefore begin the discussion with a more detailed consideration of the relation of our theory and model to other theories and models of analogical reasoning. In the second part of the discussion we turn our attention to six themes raised by the comparison of Connectionism and Dynamic Systems.

Relational priming and other approaches to analogical reasoning

Traditionally, accounts of analogical reasoning have been grounded in symbolic approaches to cognitive processing. Thus, according to Structure Mapping Theory (SMT: Gentner, 1983, 1988), one of the dominant approaches, analogical reasoning involves first selecting base and target domains for an analogy and then constructing a structural mapping between the domains in which objects in the base domain are mapped to objects in the target domain. Predicates or relations between objects are preserved across this structural mapping, so that if a predicate holds of an object in the base domain, then by analogy the predicate will hold of the corresponding object in the target domain. Many other accounts of analogical reasoning adopt a similar basic assumption: that the core to analogical reasoning lies in mapping or aligning the structural relations between elements in the base domain with corresponding elements in the target domain. This is even the case in hybrid connectionist/symbolic accounts such as ACME (Holyoak and Thagard, 1989) and LISA (Hummel and Holyoak, 1997). The relational priming account of analogy differs substantially from all of these accounts in that the fundamental process is held not to be one of structural alignment, but one of priming. Relational priming therefore constitutes a genuine distinct alternative theory of analogical reasoning.

The difference in mechanism between relational priming and structural alignment accounts of analogical reasoning reflects differences in origins of the theories and differences in the types of analogical reasoning that the theories take as core data. While SMT has been applied to developmental data (e.g., Gentner, 1988), the theory was originally conceived in the context of adult analogical competence (Gentner, 1983). Other structural alignment accounts of analogical reasoning have largely tended to follow this lead. Correspondingly, the data that such theories have been primarily concerned with is that of the adult literature (e.g., Gick and Holyoak, 1980). Complex, multi-relational analogies, such as that between the solar system and the Rutherford atom, are central within this literature. In contrast, in presenting the relational priming account of analogical reasoning we take the primary data to be that involving children's developing abilities at relatively simple analogies involving a single relation (i.e., the seven key phenomena described in our review of the behavioural data). Relational priming, and its instantiation in the model presented above, is successful to the extent that it can account for each of these phenomena.

At the same time, it remains to be demonstrated that relational priming can account for the more complex analogical abilities of adults, particular those involving multi-relational analogies. In our view such analogies are not beyond the scope of relational priming. Rather, they require either the iterative application of relational priming or the application of relational priming with more complex relations (e.g., ternary relations or relations between relations).

Cross-cutting themes

The seven key phenomena highlighted in our review of the behavioural data include some that might be regarded as characteristic of Connectionism (e.g., relational knowledge accretion) and others that might be regarded as characteristic of Dynamic Systems (e.g., Hosenfeld et al.'s indicators of discontinuous change). Crucially, our model exhibits key features of both Connectionism and Dynamic Systems approaches in capturing these effects. We now consider the implications of this for the relation between Connectionism and Dynamic Systems.

Development

We have argued that the development of early analogical ability is characterised by seven key phenomena. These phenomena arise, on our account, from increasing relational knowledge and the effects of this knowledge as expressed through relational priming. That is, development within our model is equated with increasing relational knowledge resulting from continuous exposure to exemplar transformations and not, for example, biophysiological maturation. Yet knowledge accretion and relational priming are not sufficient to account for all of the key phenomena. In particular, Hosenfeld et al.'s (1997) indicators of discontinuous change (phenomenon 7) would not be predicted from a theoretical account couched purely in terms of continuous knowledge accretion and relational priming.

At the cognitive level, what changes with development in our model is knowledge of relations. This knowledge is accessible in the sense that clamping the external nodes of the network to any two of, for example, *apple*, *knife* and *cut apple*, and allowing the network to settle will result in the network filling in the remaining external nodes with its best guess at the third element in the relation. The accuracy of that guess indicates the extent to which the network has acquired knowledge of the object and the causal agent.

At the computational level what changes with development is the attractor structure of the network. Initially, this attractor structure is, if present at all, entirely random. Early in training the network learns that, to a first approximation, features tend to be preserved in transformations resulting from the application of causal agents (i.e., while cutting transforms some features of apple, it preserves others, while the features that it does transform are preserved by interactions with yet other causal agents such as bruising or washing). This knowledge is encoded in the initial weak attractor structure where transformations between before and after states of the object are dominated by perceptual similarity. Only through additional learning does the attractor structure change to reflect the relational nature of causal agents (so that *apple*, *knife* and *cut apple* correspond to a stable state of the network).

The model also demonstrates that, while Hosenfeld et al.'s (1997) indicators of discontinuous change may initially appear inconsistent with an account in which the only change over time results from knowledge accretion, this is not in fact the case.

Knowledge accretion sits comfortably with the Dynamic Systems approach, and in concert they account for the apparent discontinuities in behaviour. Moreover, the discontinuities were emergent features of our model – their existence played no role in the model's design. There is no sense in which they were "programmed in". To develop this argument further, in our opinion the Dynamic Systems approach is essential in accounting for Hosenfeld et al.'s data. We believe that the data could not be captured by a model based on purely feed-forward connectionist techniques, or one based on serial symbolic processing.

Core concepts

As should be clear from the above, core concepts of both Connectionism and Dynamic Systems play central roles in our account of the development of analogical reasoning. From a representational perspective, the account (and subsequent model) is committed to featural representations of objects in the before and after states, and of causal agents at the CA units. During analogical completion the representation of the causal agent and a representation of the transformation it implies is derived from the transformation between the before and after states (i.e., the *a* and *b* terms of the *a:b::c:d* analogy). These representations are the product of constraint satisfaction implemented in the dynamics of the settling network. At this level the model is committed to the core concept of settling to an attractor state as exhibited by Dynamic Systems.

The representations established at the hidden units during settling when presented with the *a* and *b* terms of an *a:b::c:d* type analogy, and the representations established at the CA units, will, once the network has been trained, correspond to the representations established when the network is trained on the relevant transformation by which *a* has been seen (during training) to be transformed to *b*. Thus, in the *apple:cut apple* case the representation established at the CA units corresponds to the representation of knife used in the training set. The activation pattern of the hidden units established following settling after presentation of the *a* and *b* terms of an analogy are more interesting. They effectively represent the relation as a qualified transformation of the before state. The transformation is qualified because *cutting* does not transform all objects: the representations of *stone*, *knife* and *stone* corresponds to a stable state of the fully trained network. At the same time hidden unit representations are the product of a non-symbolic learning algorithm and are not semantically transparent. The model thus shares with many connectionist models the characteristic that there is no way in which individual hidden units may be mapped to macro-level features.

Solving *a:b::c:d* type analogies within the model actually involves two attractor states: one relating *a*, *b* and an unspecified causal agent, and the other involving *c*, the unspecified causal agent and *d*. The transition from one attractor to another is typical of Dynamic Systems, and within our model analogically correct transitions are the result of priming, a concept which again is most naturally related to Dynamic Systems. But within our model it is a commitment to a further core concept from Connectionism – of knowledge accretion over time via exposure to exemplars – that results in the development of hidden unit representations that support priming. More specifically, it is this exposure in conjunction with the contrastive Hebbian learning algorithm that results in the semantically non-transparent representations at the hidden units which prime the network so that it transitions from the attractor state generated by the *a:b* terms of the analogy to an analogically appropriate attractor state given the *c* term of the analogy.

Time scales

While Connectionism and Dynamic Systems work together in our account of the development of analogical reasoning, they work on different time scales. Principles of Dynamic Systems function during the solving of individual analogies in terms of settling to attractor states and the priming of transitions between attractor states. Connectionist principles, on the other hand, operate on developmental time scales, with learning and the accretion of knowledge shaping the structure of attractors over timescales corresponding to months and years. From this perspective the model might be viewed as a Connectionist model of development superimposed upon a Dynamic Systems model of analogical reasoning. However, while such a view has some intuitive appeal, it does not do full justice to the interaction between the two approaches, for the attractors on which the Dynamic System is based are developed from representational and learning principles of Connectionism. Therefore neither view can be adopted in isolation.

Levels of explanation

We have argued that our relational priming model may be viewed as either a dynamic system or a connectionist model. Given this, it is tempting to think that the difference in views reflects a simple difference in levels of explanation, with, for example, the Dynamic Systems view being more implementational and the Connectionism view being somehow higher level. This is not an accurate depiction of the relation between the two.

The fundamental philosophical claim behind levels of explanation is that while a system's behaviour may be explained at multiple levels, explanations within a level are in some way complete. Thus an algorithm can be described independent of its implementation, and a computational function can be specified independent of its method of evaluation (i.e., of an algorithm for the function). This kind of separation cannot be given for the dynamic and connectionist elements of our model. They are highly inter-related, and regardless of whether one adopts Marr's (1982) traditional tri-partite approach or a more fine-grained approaches such as Newell's (1990), higher level explanations require principles from neither Connectionism or Dynamic Systems, while lower level explanations make reference to principles from both. Given this, the relation between dynamic system principles and connectionist principles within our model is better characterised in terms of differing time scales (as described above) than in terms of differing levels of explanation.

Constraints

The theoretical assumption of relational priming as the core process underlying simple analogical reasoning requires some computational mechanism to establish and hold a representation during the several seconds taken to solve an analogy. Either Dynamic Systems or Connectionism might provide such a mechanism via recurrent connections within a network of nodes. At the same time, empirical constraints (particularly those suggesting discontinuous change) point to the involvement of other principles grounded in Dynamic Systems (e.g., descriptions in terms of phase changes). Yet models founded purely on Dynamic Systems principles cannot easily account for the empirical facts of change over developmental timescales (i.e., of learning). Learning, particularly as it relates to relational knowledge accretion and domain specificity, argues against a purely Dynamic Systems (anti-representationalist) approach. The theoretical and empirical constraints therefore point to a system that combines principles of both Dynamic Systems and Connectionism. The model presented above does just this.

With regard to the neural implementation of the model, there are no strong constraints from physiology. At best one might argue that the neural evidence points to the distributed representation of semantic knowledge in the temporal cortices (see, e.g., Shelton and Caramazza, 2001). The model is consistent with this in that it involves distributed representations of semantic knowledge. Beyond this the model makes no neural claims.

Assumptions

As in many models, assumptions may be classified as relating to architecture, representation, processing, or learning. A critical feature of our architecture is the simultaneous representation of the *a* and *b* terms of an analogy at the before state and after state nodes of the model. This is critical in allowing the model to derive a representation of the relation between the terms. Bi-directional connectivity and the processing assumptions of a settling architecture are critical in ensuring that the model derives a stable representation of the relation and relevant causal agent that may prime subsequent processing of the *c* term of the analogy. Finally, the training set in conjunction with contrastive Hebbian learning ensures that the model develops attractor states in which two related objects (corresponding to both *a* and *b* terms and *c* and *d* terms of an analogy) and a relevant causal agent may be simultaneously and stably represented at the appropriate nodes in the model. In the context of the debate between Connectionism and Dynamic Systems, the assumptions concerning representation, processing and learning are most critical.

That objects must be represented in some way in “before” and “after” states is relatively uncontroversial: information must be supplied to and extracted from the model, and these representations serve that purpose. Perhaps more interesting is the assumption that the causal agent is explicitly represented. After all, in an analogy such as *apple:cut apple::bread:cut bread* there is no explicit mention of a causal agent. It is indeed possible that the CA nodes could be excised from the model without impacting the model's functioning, but children are able to identify relevant causal agents when given the *a* and *b* terms of an analogy. Furthermore in their experience of the world we assume that children do not simply observe objects in before and after states. Rather they typically observe objects being transformed by causal agents.

We assume, following the development literature (Pauen and Wilkening, 1997; Ingaki and Hatano, 1987; Tunteler and Resing, 2002), that analogical reasoning is an emergent ability that develops spontaneously rather than a skill at which children receive specific training. Furthermore, we assume that changes in analogical abilities reflect the acquisition of domain knowledge. Our assumptions regarding learning therefore concern how objects are transformed, and not how to solve analogies. The training set therefore includes instances of objects being transformed by causal agents (corresponding to, for example, the interaction of an apple with a knife) and objects remaining unchanged by causal agents (corresponding to, for example, the interaction of a stone with a knife). Beyond the existence of both types of interaction, there are no special biases in the training set, though of course if bias were introduced by increased interaction with one kind of object and decreased interaction with another, then the network will exhibit an influence of domain knowledge and achieve mastery of analogies relating to the objects it has more experience with (as in the behavioural data: Goswami and Brown, 1989).

The use of contrastive Hebbian learning allows the model to form appropriate attractor structures, but other learning algorithms with similar characteristics are likely to yield similar results. In particular, we believe that the model's behaviour is robust to incorporation of other assumptions from the LEABRA framework (O'Reilly, 1996;

O’Reilly and Munukata, 2000). The use of contrastive Hebbian learning demonstrates that these specific assumptions are sufficient to produce the necessary attractor structure in the trained model.

Conclusion

We have presented a model of the development of simple analogical reasoning and shown how the model accounts for seven characteristics of children’s developing abilities to reason analogically. We have further argued that the model’s success is critically dependent upon principles of both Connectionism and Dynamic Systems. Thus, our model demonstrates complementarity between the approaches. Connectionism and Dynamic Systems need not be viewed as mutually exclusive alternative conceptions of cognitive developmental. Rather, they may by viewed as alternative, but mutually supportive, perspectives on a single system. While this argument has been made with a specific model, there is nothing unique about the model’s relation to Connectionism and Dynamic Systems. The connectionist approach to representation and the dynamic approach to settling and constraint satisfaction can and have been combined in a range of domains, and it is our opinion that this combination of approaches is likely to continue to yield advances in our understanding of the computational foundations of cognitive development.

Acknowledgements

This work was supported by an ESRC studentship awarded to Robert Leech, and by EC grant 29088 (NEST).

References

Amari, S. I. (1977) Neural theory of association and concept formation. *Biological Cybernetics*, *26*, 175–185.

Amari, S. I. (1980) Topographic organization of nerve fields. *Bulletin of Mathematical Biology*, *42*, 339–364.

Anderson, J. A. & Rosenfeld, E. (1989) *Neurocomputing: Foundations of research*. Cambridge, MA: MIT press.

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

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuron-like adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*, 834–846.

Boden, M. A. (1988) *Computer models of mind*. Cambridge, MA; Cambridge University Press.

Feldman, J.A. & Ballard, D.H. (1982). Connectionist models and their properties. *Cognitive Science 6*, 205–254

Fodor, J. (1975) *The language of thought*. Cambridge, MA: MIT Press

French, R.M. (2002). The Computational Modeling of Analogy-Making. *Trends in Cognitive Sciences*, *6*, 200–205.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*, 47–59.

Gentner, D., Rattermann, M. J., Markman, A. B., & Kotovsky, L. (1995). Two forces in the development of relational similarity. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling,* (pp. 263–313). Hillsdale, NJ: LEA.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*, 277–300.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355.

Goswami, U., & Brown, A.L. (1989). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, *35*, 69–95

Grossberg, S. (1980) How does the brain build a cognitive code? *Psychological Review*, *87*, 1–51.

Hinton, G. E., & Anderson, J. A. (1981). Parallel models of associative memory. Hillsdale, NJ: Erlbaum.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295–355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427–466.

Inagaki, K., & Hatano, G. (1987). Young children’s spontaneous personifications as analogy. *Child Development*, *58*, 1013–1020.

Japkowicz, N. (2001). Supervised and unsupervised binary learning by feedforward neural networks. *Machine Learning, 42*, 97–122.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797–2822.

Leech, R., Mareschal, D. & Cooper, R. (2003) A connectionist account of analogical development. In R. Altman & D. Kirsh (eds.) *Proceedings of the twenty-fifth annual conference of the Cognitive Science Society* (pp. 710-715). London: LEA .

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*, 315–353.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Munakata, Y. & McClelland, J.L. (2003). Connectionist models of development. *Developmental Science, 6*, 413–429

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

O'Reilly, R.C. (1996). Biologically plausible error-driven learning using local activation differences. *Neural Computation, 8*, 895–938.

O'Reilly, R.C. & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press.

Pauen, S., & Wilkening, F. (1997). Children's analogical reasoning about natural phenomena. *Journal of Experimental Child Psychology, 67*, 90–113.

Plunkett, K. & Sinha, C. (1992). Connectionism and developmental theory. *British Journal of Developmental Psychology*, *10*, 209–254.

Pylyshyn, Z. W. (1984) *Computation and cognition: towards a foundation for cognitive science.* Cambridge, MA: MIT press.

Rattermann, M. J., Gentner, D., & DeLoache, J. (1990). The effects of familiar labels on young children's performance in an analogical mapping task. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, (pp. 22–29). Hillsdale, NJ: LEA.

Rattermann, M.J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, *13,* 453–478.

Rogers, T. T., & McClelland, J.L. (2004). *Semantic Cognition*. MIT Books.

Rosenblatt, F. (1958) The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65,* 386–408.

Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel Distributed Processing, Vol. 1*. Cambridge, MA: MIT Press,

Shelton, J. R. & Caramazza, A. (2001). The organization of semantic memory. In B. Rapp (Ed.) *The Handbook of Cognitive Neuropsychology,* (pp. 423–443). Philadelphia, PA: Psychology Press.

Smith, L. B., & Samuelson, L. (2003) Connectionism and Dynamic Systems. *Developmental Science, 6,* 434–439.

Sternberg, S. (1969) Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist, 57*, 421–457.

Thelen, E Schöner, G, Scheier, & Smith, L. B. (2001) The dynamics of embodiment: A field theory of infant perseverative reaching. *Brain and Behavioral Sciences, 24*: 1-33.

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

Thomas, M.S.C., & Mareschal, D. (1997). Connectionism and psychological notions of similarity. In *Proceedings of the nineteenth annual conference of the Cognitive Science Society*, (pp. 757–762). London: Erlbaum.

Tunteler E., & Resing W.C. (2002). Spontaneous analogical transfer in 4-year-olds: a microgenetic study. *Journal of Experimental Child Psychology, 83*,149–66.

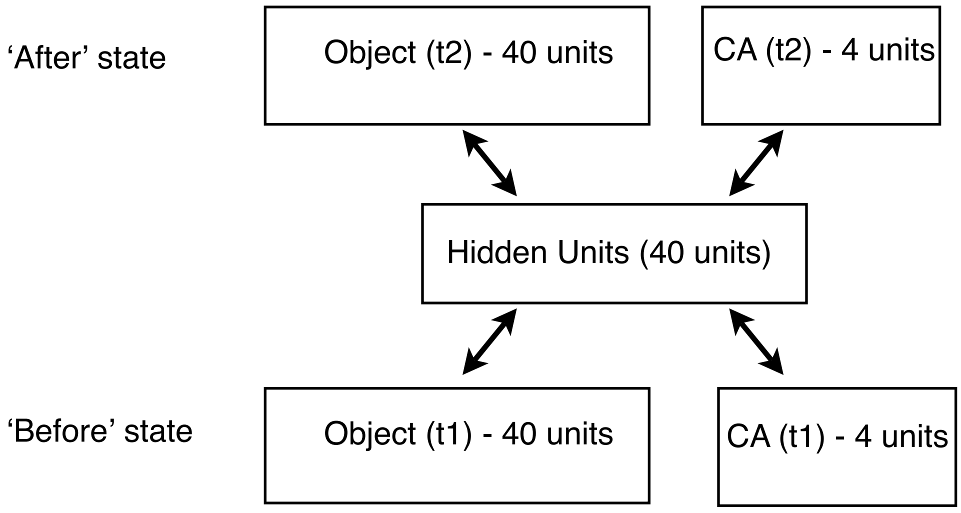


Figure 1: Schema of the architecture of the model.

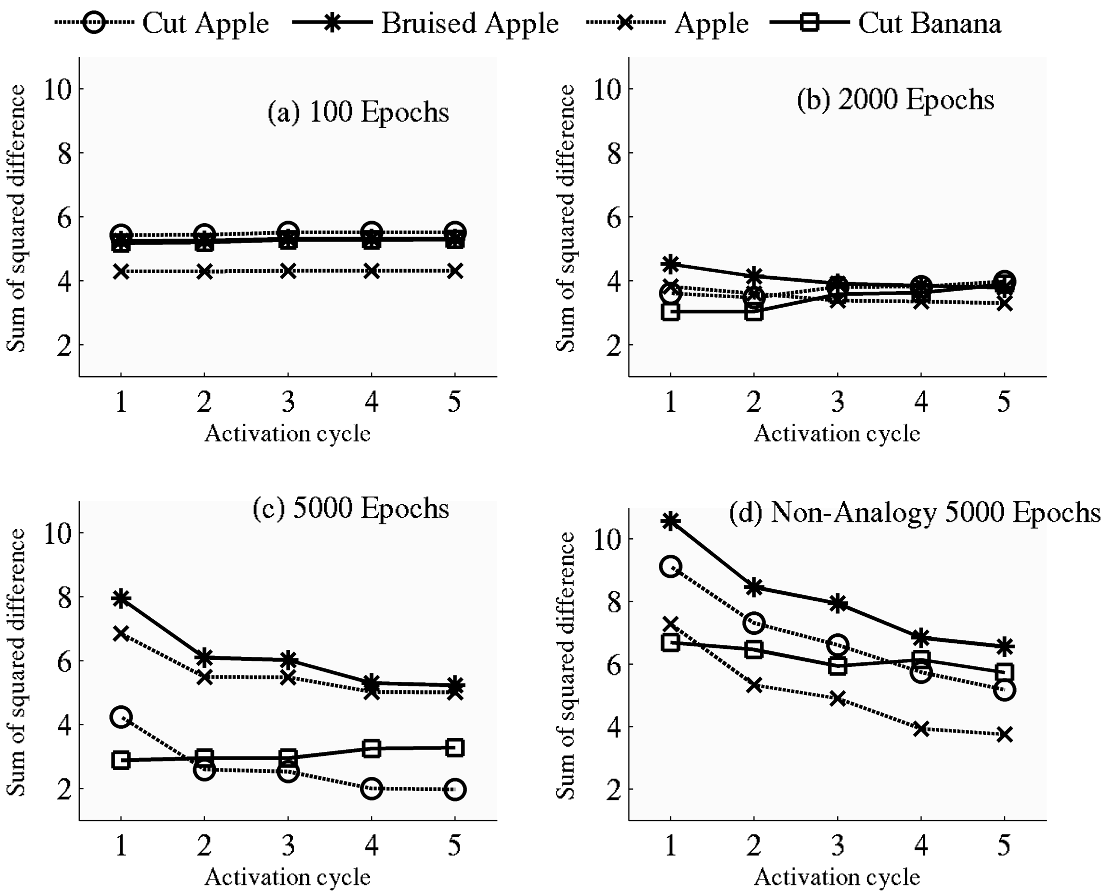


Figure 2: Analogical completion over development. The y-axis shows the sum-of-squared distance calculated between the actual output of the network and four target patterns (the four different lines). The lower the y-axis value, the closer the actual activation is to that target pattern. Figures (a-c) show the network's response to an analogy (e.g., bread is to cut bread as apple is to what?) and (d) shows an example of the network's response to a non-analogy (e.g., bread is to bread as apple is to what?).

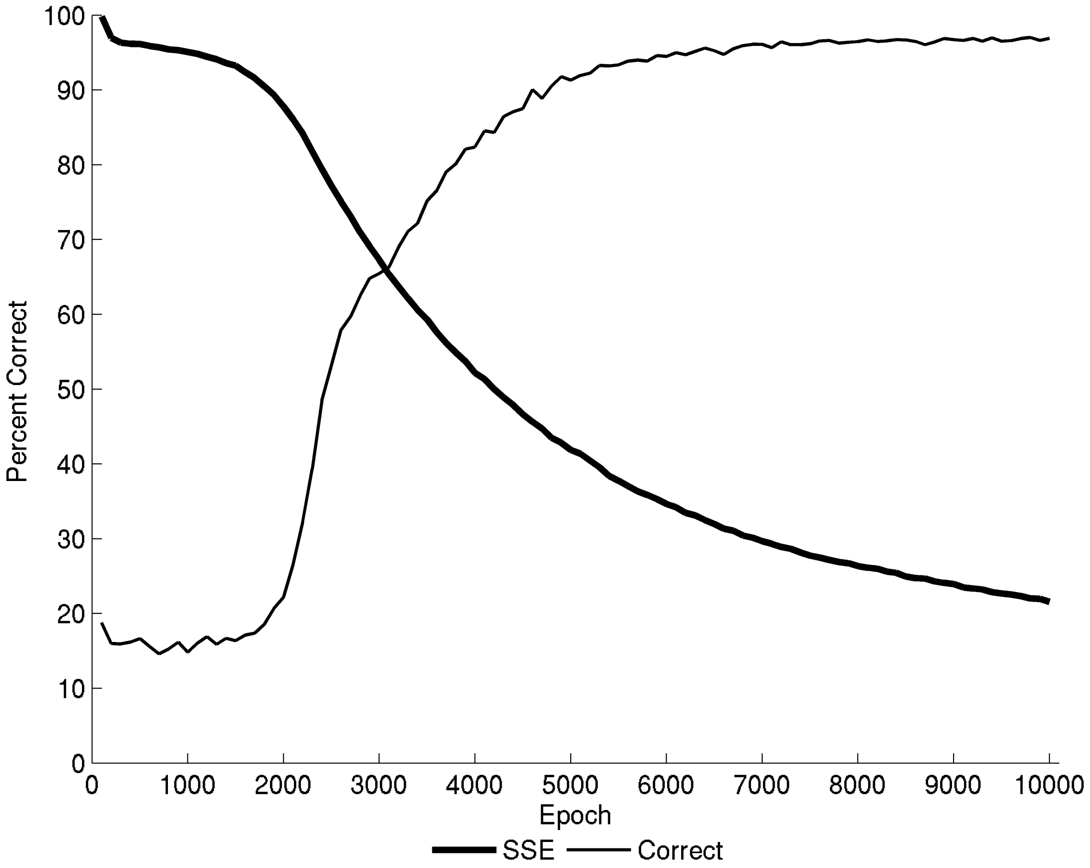


Figure 3. Percentage of analogically appropriate responses at Object(t2) (thin line) and the normalized mean sum of squared error (thick line) over training. These results are the average of 50 replications.

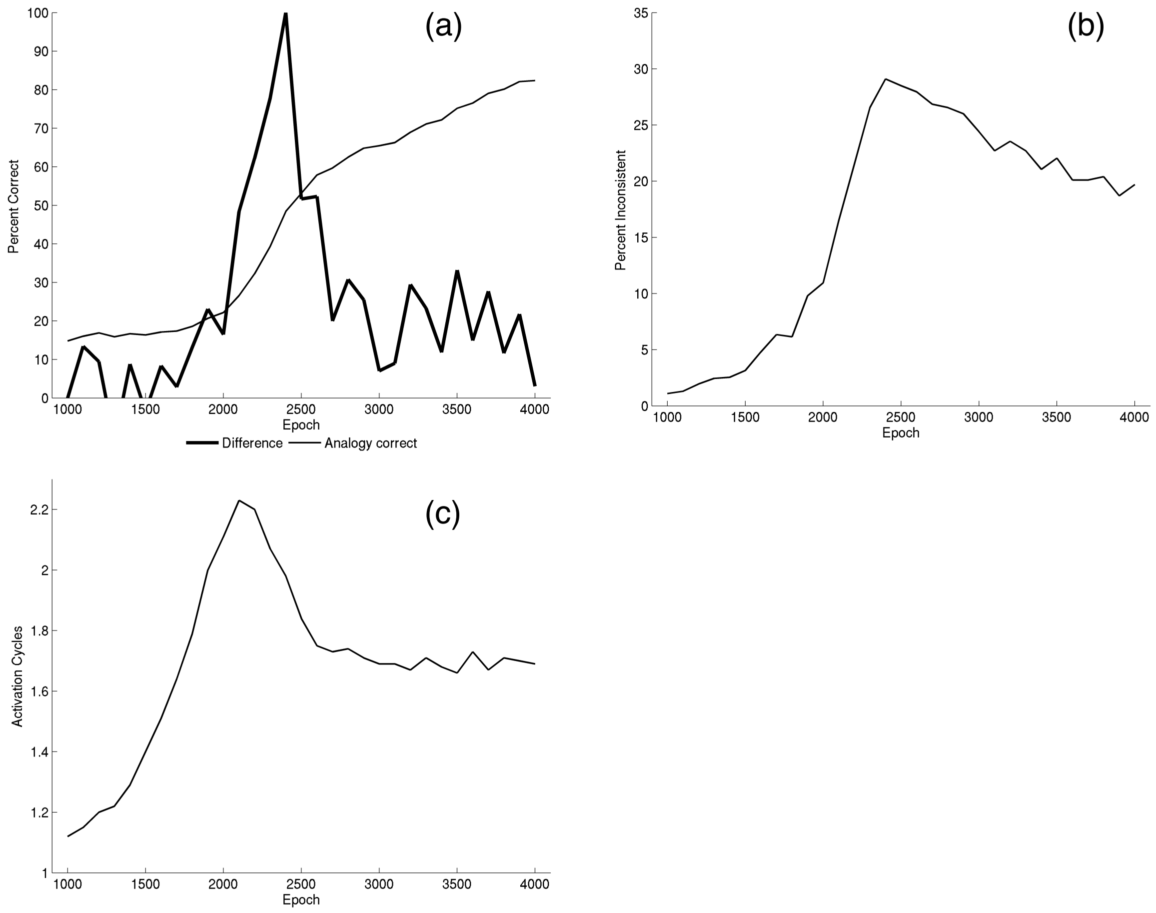


Figure 4: a) Evidence for a ‘sudden jump’. The thin line is the percentage of analogies completed correctly. The thicker line is the normalized difference in correct analogies across time, demonstrating a ‘sudden jump’ centered around 2400 epochs of training(the results are the average of 50 replications). b) Inconsistency across training (the results are the average of 50 replications). c) Number of activation cycles before a unique response, across training (the results are the average of 100 replications.)

	Children		**Networks**	
Response type	**4-year-olds**	**5-year-olds**	**2300 epochs**	**2800 epochs**
Correct analogy	35%	67%	39.3%	62.5%
Wrong transformation	35%	28%	16.1%	8.9%
Object-similarity error	22%	3%	22.4%	2.5%

Table 1: The profile of responses made by children and the network (the network’s performance is the average of 50 replications. The children’s data is taken from Rattermann and Gentner, 1998).

Towards an integration of connectionist learning and dynamical systems processing: case studies in speech and lexical development.

Bob McMurray
Jessica Horst
Joseph Toscano
and
Larissa K. Samuelson

Dept. of Psychology
University of Iowa

Corresponding Author:
Bob McMurray
bob-mcmurray@uiowa.edu
Dept. of Psychology
E11 SSH
University of Iowa
Iowa City, IA 52242
319-335-2408 (office)
319-335-0191 (fax)

Introduction

The goal of experimental psychology is to use observable behavior to understand mental and biological processes, a challenging task that is doubly hard for the developmental scientist. Developmental science is tasked with not only understanding these unobservable processes, but also with understanding the (typically unobservable) processes behind their change. This study of *developmental mechanism* often appears intractable—there are too many degrees of freedom, to many points of inference, and too little data.

In recent years, however, computational theory has begun to make the search for developmental mechanism more tractable. Computational approaches provide a check on the logical consistency of theories, a method for examining the prediction space of a theory, a way to determine what factors of an adult processing model are susceptible to developmental change, and a set of biological, mathematical and physical constraints on theories. In particular, connectionism (Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996) and dynamical systems theory (Thelen & Smith, 1996) both offer compelling paradigms for modeling development. However, both differ on the constraints they place on possible mechanisms.

This chapter examines the interplay of these two frameworks at the level of both theory and computation with two case studies from the field of language acquisition. We start with a discussion of developmental mechanism and developmental process and the contributions of connectionist and dynamical systems theories. The first case study—a model of speech category learning—will illustrate how a dynamical systems perspective may inform a classic connectionist mechanism (statistical learning). The second case study examines a hybrid model of early word learning which combines connectionist and dynamical systems techniques. Both of examples suggest that we can no longer treat these two paradigms as independent.

Developmental Mechanism, Connectionism and Dynamical Systems Theory

As a rough working definition, we define *developmental mechanism* as an explanation of the fundamental processes or capabilities within the developing system that account for long-term, "developmental" change (but may also account for short-term processes). This explanation must bridge multiple levels of description (in a developmental systems sense). For example, relating development to interactions between the environment and the biological makeup. It must also explain change at an appropriate level of description. For example, explaining language development at a molecular level is a stretch, but explaining it via interactions with behaviorally-relevant inputs (e.g. words) is relevant.

While developmental mechanism often refers to specific singular mechanisms (e.g. accommodation), development itself is clearly multiply determined. True Developmental Mechanism or Process represents the totality of these many causes of change, but for ease of analysis, developmental science has historically peeled off components of this larger process for study. Ultimately, however, it is not appropriate to assume that these unary mechanisms are independent—interactivity should be the norm rather than the exception. The purpose of this chapter is explicitly to examine the interaction of two such mechanisms.

Connectionism and dynamical systems each offer a number of candidate developmental mechanisms. For the purposes of illustration, the present manuscript will discuss two: the statistical learning mechanisms proposed by connectionism, and the online "in-the-moment" processing (or constraint satisfaction) proposed by dynamical systems. Moreover connectionism and dynamical systems each ask different sorts of questions about developmental change. While connectionism focuses on the process of learning (e.g. which aspects of structure in the environment are encoded and how are they encoded), dynamical systems theory focuses on the dynamics of learning: how the system unfolds as a temporal process.

We believe that these fundamental mechanisms may be profoundly interrelated, and that both approaches to developmental change provide useful views on the same problems. The purpose of this chapter, then, is to start to move towards such interaction. In the special issue of *Developmental Science* (Spencer & Thelen, 2003) that preceded this volume and the conference that inspired it, Smith and Samuelson (2003) asked:

> *Are these in-principle differences between connectionist and dynamic systems theories? That is, could you track the trajectory of change in the weights of a connectionist model and gain insight into the process of learning or development in the modeled system? Or, could a dynamical system learn the statistical regularities inherent in the world and use them to direct and change its behavior? Surely both are possible.*

This chapter will (modestly) do both. A dynamical systems approach has a lot to offer the study of learning. The first case study demonstrates that not only can this help us understand learning processes, but that learning may be involved in moment-by-moment processing. Moreover, in simple localist architectures, dynamical systems can be integrated with connectionist learning. The second case study presents a model that has done this and shows how dynamical processes benefit from statistical learning, and, more importantly, how they buttress these learning processes to yield in the moment behavior that exceeds what would be provided by learning alone. First we discuss the developmental mechanisms offered by each paradigm.

Connectionism and Statistical Learning

Connectionism rests in large part on the demonstration that simple networks of neuron-like units can behave in interesting ways, and that complex behavior can emerge out of simple local interactions. This emergence is the cornerstone of *developmental connectionism*—by studying the *ongoing* emergence of structure and the ways in which the environment influences it, we can apply the simple principles of connectionism to development. Because these local interactions and their emergent properties arise out of environmentally tuned learning, development represents the ultimate test of this framework: a comparison of learning and development in these simple systems to that of the child is critical evidence for this framework. Connectionist networks are models of cognitive and perceptual processes that transduce some input into some output, usually in a goal-specific way. This represents a simplification for two reasons. First, the real-time, moment-by-moment dynamics are often left out, resulting in a model that transforms input to output in a small number of discrete steps (though see McClelland & Elman, 1986). Second, connectionist networks rarely directly model the raw sensory inputs or motoric outputs (though see Nakisa & Plunket, 1998; Schlessinger, Parisi & Langer, 2000). These simplifications create a resemblance to information processing models, but this is only superficial: connectionism's commitment to task-specific processing means that intermediate processing states (hidden units) are simply that, not abstract domain independent symbolic representations.

These two simplifications are made largely because of an underlying commitment to a single core developmental mechanism: statistical learning. Processing in a network is fundamentally a matter of connection weight. Activation is passed from input to hidden to output layers along variable-strength weights, and the specificity of processing of a network is largely due to its particular weights. Most learning rules alter these weights, encoding statistical relationships between features of the input, and between those same features and the output.

Under this view, the environment contains rich statistical structure that, once encoded in weight matrices, enables the system to perform tasks such as word segmentation, edge detection, or face identification with minimal domain-specific architecture. Thus, the same simple processing units and learning rules can handle many different types of inputs, and differences between domains arise because of the statistical structure of the task, not the underlying neural hardware. This simple approach has been surprisingly successful. Networks can learn to use transitional probabilities between phonemes (Elman, 1990), eigenvectors of face images (Turk and Pentland, 1991), surface regularities among irregular past tense verbs (Rumelhart & McClelland, 1986), and regularities in edges and their orientation in visual scenes (Olshausen and Field, 1996) to perform common cognitive and perceptual tasks. These examples differ in terms of learning principles and architecture, but they make the common points that the information needed to solve a particular problem is available in the world, and that development can be accounted for by the accumulation of this information as biases on processing (connection weights).

While connectionist networks are *capable of* solving many sorts of problems, an individual network only learns the specific statistical structure that is relevant for the particulars of the task that it is performing. Given the same input and a different goal, a different structure may be extracted. In fact, the actual statistics that are directly encoded in a network's weights are the relationships between inputs and outputs—in order to understand the relationship between different inputs (structure in the environment), one must analyze the relationship between weights and hidden or output units. More importantly, this structure is not encoded as abstract knowledge, but rather, it is encoded directly within the processing capacity.

A substantial body of work on goal-directed statistical learning has asked examined the sorts of statistics can be learned and the properties of various learning schemes. However, a small number of studies have taken a more dynamical systems approach. Here, if the statistical structure in the environment is stable, and the learning rules are based on gradient descent, then, weight change can be seen as movement towards an attractor state in weight-space along the dynamics specified by the learning rule.

Thus, we can ask about the stability and dynamics of learning itself. For example, back propagation is susceptible to catastrophic interference (McCloskey & Cohen, 1986): training a network on one batch of data followed by switching to a different training set results in the loss of any learning on the first training set. Put another way, manipulating the stability of the input results in the model's gravitating towards a new attractor state, causing it to "forget" the prior input set. In contrast, Hebbian learning models that have adopted an attractor state (in weight-space) for a given input are often unable to adapt to new statistical structure—forcing new data into the old hole (see McClelland, this volume). Thus, Hebbian learning, in many cases, does not have the ability to leave an attractor state, once established.

While these analyses of learning dynamics in connectionist systems are revealing, they also are somewhat cursory (did the model learn or not)—clearly more microstructural analyses are possible. Nonetheless, both analyses also suggest a solution: slow learning. This prevents the network from settling into an incorrect attractor state, and maximizes the quantity of data received by the model prior to settling, thus making it more likely that a model would be exposed to both datasets prior to stabilizing. This constraint, while a good solution to these problems, will be challenged by the present paper in three ways. First, slow learning implicitly hypothesizes that learning does not take part in short-term phenomena such as short-term perceptual learning (e.g. Kraljic & Samuel, 2005; Norris, McQueen & Cutler, 2003). The first case study will examine the limits of this hypothesis in speech perception. Second, occasionally children's ability to acquire knowledge appears rapid and complete with only a few exposures. The second case study will demonstrate how integrating online constraint satisfaction with slow learning yields overt behavior that appears as if the model learned new words rapidly with minimal exposure.

Dynamical Systems Theory and Online Processing

As a theory of development, dynamical systems theory offers two important additions to our cachet of mechanisms: embodiment and online processing. Under the embodiment view (Samuelson & Smith, 2002; Thelen, 2000), the mechanics of the body, sensory and perceptual processes, and the notion of children as active learners and creators of their environments play pivotal roles in shaping motor behavior, and underpin cognitive and linguistic behavior.

We agree on the importance of embodiment. It is important to note, however, that embodiment is distinct from the fundamental *computational* issues that divide connectionism and dynamical systems. There are dynamical systems models that make no reference to embodiment (e.g. van Geert, 1998) and connectionist approaches that embrace it (e.g. Bechtel, 1997). Nonetheless, for many accounts, this factor is indeed a key difference: connectionists tend to lean towards explanations based on information, and dynamical systems theorists lean towards those based on embodiment.

The second major contribution is that dynamical systems paradigm views time as continuous and fundamental, using the simple, but powerful formalism:

$$S_t = F(S_{t-1}) \qquad (1)$$

The state of the system at the current time (S_t) is some function F of the state at the previous time (S_{t-1}). F(S) then represents development itself—how one describes the change that is unfolding over time. Of course, the more radical (and realistic) instantiation of this formalism

$$S_t = F(S_{t-dt}) \qquad (2)$$

treats time as continuous. The deep implication here is that there is no principled place to divide time for measurement. It makes no sense to assume the momentary state of a six-month-old is a function of the momentary state of the three-month-old—that would require causality to leap into the future. Instead, to understand development we must look at adjacent times and examine moment-by-moment processing. That is, the state of the six-month-old is the product of many moment-by-moment processes leading from birth (and before) to later states.

This formalism admits no clear division between the long time-spans of developmental time and the short time-spans of in-the-moment behavior and action. Dynamical systems theory argues that the proper study of development must be concerned with in-the-moment behavior or, on the flip side, that in-the-moment behavior is a crucial determinant of development. This latter idea has become a hallmark of developmental dynamical systems thinking: behavior is not learned or acquired in any kind of static sense, but rather, it is assembled online on the basis of constraints provided by context, the developing systems' abilities, and developmental history.

This, in a sense, takes the connectionist emphasis on goal-directed processing "to task" by pointing out that any behavioral task is the product of multiple constraints and influences, over and above what may be learned by the system. The connectionist concept of "process", under this view, is too simple, ignores the behavioral context and active nature of the child in the environment and in the system, and reveals a perhaps misplaced focus on the underlying information processing. While this may be true, it does not negate the power of the connectionist argument: clearly for many perceptual, cognitive and linguistic problems the statistical structure latent in the world and in the sorts of tasks that a child encounters can be a powerful source of information for generating behavior.

Towards an Integration

Given the description above, connectionism and dynamical systems provide complementary views of development. While connectionism focuses on slower learning processes, dynamical systems argues that moment-by-moment behavior cannot be ignored. Moreover, while connectionist models use learning to argue that there is sufficient information in the input to accomplish some task, dynamical systems that examine learning focus more on the temporal dynamics of learning rather than the information and mechanisms behind it.

Both paradigms are dominated by computational formalism as a key component of understanding developmental mechanism. Moreover, connectionist systems are a special class of dynamical systems (see Smolensky, Mozer & Rumelhart, 1996) and many dynamical systems adopt connectionist-like architectures (e.g. dynamic field theory: Thelen, Schöner, Scheier & Smith, 2001). Despite this, there have been few attempts to integrate both developmental mechanisms into a single computational framework. Connectionist networks that embrace moment-by-moment processing (e.g. McClelland & Elman, 1986) do so by abandoning learning altogether. Other models (e.g. Tabor and Tanenhaus, 1999) implement a dynamical system on top of a connectionist network—in a sense treating the moment-by-moment processing as a separate module from the connectionist learning. On the other hand, dynamical systems that embrace learning (e.g. van Geert, 1998) do so by focusing only on the slow developmental time-course, and by simplifying the nature of the learning and the relevant information.

Integrating these approaches has proven to be mathematically difficult, because both learning and online processing represent dynamical systems on their own, with their own sources of instability. Including both in a single model multiplies the sources of instability. This should ultimately be a central question of concern to developmental scientists: how are the brain and the body constrained to avoid these multiple sources of instability? However, in the short term, this difficulty has slowed the integration of both developmental mechanisms in a single model.

Consider the classic logistic dynamical system (e.g. Spivey, in press; Ward, 2002):

$$y_{t+1} = \eta * y_t * (1 - y_t) \tag{3}$$

This simple dynamical system is stable, with a tendency to trend (over some number of iterations) towards a value between 0 and 1 (its attractor states) regardless of the initial values of y. However, this stability only exists when η is between 0 and 3—at values between 3 and 4 it oscillates, and at higher values it behaves chaotically.

If one were to add learning to this simple system, the obvious place would be to use a learning rule to optimize η. However, the behavior of the system is dependent on η—if the learning rule resulted in η greater than 4, chaos will result. Thus, the learning rule must be stable by itself, and have a stable attractor between 0 and 4. While this seems simple, the sorts of dynamical systems (and learning rules) one might use to model actual development will be much more complex and contain many more parameters, quickly making the problem intractable.

These problems have led to an assumption of additivity. That is, while many connectionists acknowledge the importance of online dynamics, they assume that they either live "on top" of an underlying connectionist network (e.g. Tabor & Tanenhaus, 1999), or that the connectionist network itself is a distillation of a much more complex underlying dynamical system (but that the simplification doesn't matter). While this may be reasonable from the standpoint of computation, connectionism and dynamical systems have both become scientific paradigms for the study of development. Both offer hypotheses concerning possible *developmental mechanisms* (statistical learning, online processing and embodiment). Thus, from the standpoint of developmental science, it is not safe to assume that these developmental mechanisms are additive—their interactions may have crucial implications for development. Thus, we turn to our investigation of these interactions.

Case Study 1: Statistical Learning and Speech Category Acquisition

The acquisition of speech categories presents an ideal domain in which to study the dynamic properties of statistical learning processes because the input is well understood, the developmental timecourse has been clearly mapped out over many studies, and we have empirical evidence for statistical learning.

Empirical Results

Speech input is fundamentally continuous and can be characterized by a variety of continuous acoustic cues, such as formant frequencies, the amplitude of components such as vocal chord vibration or aspiration, and the relative timing of such components. This continuous acoustic/phonetic cue-space is carved into many categories by a given language (e.g. phonemes), and the number of categories and their locations (in cue-space) differ cross-linguistically. One of the first tasks that infants are faced with, then, is determining the relevant categories of speech sounds.

The developmental story that has emerged suggests that infants are initially capable of discriminating a range of phonetic contrasts, many of which will not be used in their native languages. Werker and Tees (1984), for example, found that 6-month-old infants raised in an English-speaking environment (as well as those raised in a Hindi-speaking environment) could discriminate the /d/ and /d°/ (a dental stop used in Hindi), as well as /b/ and /d/, and /k’/ and /q’/ (ejective palatals and uvulars used in Thompson). By 12 months, however, these infants were only able to discriminate sounds from their own language. These findings have been extended to a range of phonetic contrasts and the pattern that has emerged is clear: infants initially are able to discriminate more sounds than they will need, and this ability is lost over the first year of development (Werker & Curtin, 2005 for a review).

Statistical learning has been proposed as a fundamental mechanism for this change (Maye, Werker & Gerken, 2002), and is supported by three major arguments. First, a growing body of evidence in phonetics has shown that speech cues are not distributed uniformly, but rather, that they statistically cluster in a way that reflects underlying categories. Lisker and Abramson (1964) measured Voice Onset Times (VOTs) in a large number of languages and found that VOT was distributed in a multi-modal Gaussian distribution with categories clustered as Gaussian distributions. This clustering has also been found for vowels and approximants (Peterson & Barney, 1951; Espy-Wilson, 1992). Thus, many continuous speech cues form Gaussian statistical clusters—perfect fodder for statistical learning.

Second, speech categories reflect a corresponding Gaussian structure. A number of studies by Joanne Miller and her colleagues (reviewed in Miller, 1997) have examined adults’ goodness ratings for wide ranges of VOTs and found that these ratings form a roughly Gaussian function with the best exemplars located at that most statistically frequent exemplars. These graded categories have also been seen in infants. Miller and Eimas (1996) showed that infants who were habituated to a non-prototypical /k/ (with an extreme VOT of 200 ms) dishabituated when exposed to a more prototypical one (110 ms). Finally, McMurray and Aslin (2005) demonstrated that the voicing categories of 8-month-old infants exhibit a gradient pattern which was visible over differences as small as 10 ms of VOT.

The third argument, and the linchpin in the case for statistical learning, comes from a study by Maye et al (2002). Infants were exposed to a series of tokens from a /d/-/t/ VOT continuum. One group of infants heard tokens whose frequencies formed a bimodal distribution (two categories) and a second group heard tokens from a unimodal distribution (one category). After a brief 2.5 minute exposure to these statistical distributions, both groups were tested for discrimination. Infants in the bimodal group discriminated pairs of sounds consistently with their having learned two categories. However, infants in the unimodal group did not discriminate these sounds.

Thus, speech category learning is an ideal domain in which to study connectionist learning processes: we understand the statistics of the input, we have a clear picture of the eventual end-state, and there is empirical evidence for statistical learning as a mechanism. A number of connectionist models have been proposed to account for this process, but despite their ability to learn the statistics, each model embodies a number of additional components that may contribute to development (e.g. topographic architectures and competition: Guenther & Gjaja. 1996; logistic activation functions and supervised learning: Elman & Zipser, 1988; unique architectures created by a genetic algorithm: Nakisa & Plunket, 1998). These factors affect the developmental trajectory of the model while not necessarily being related to statistical learning.

Since the goal of the present chapter is to understand the dynamics of learning itself, it is necessary to distill this learning process to the simplest possible learning device, using the fewest number of architectural assumptions. Thus, we developed a model of statistical speech category learning based on the Mixture of Gaussians (MOG) architecture (McMurray, Aslin & Toscano, in preparation). This model uses a classic statistical model that is well understood and learning is based on simple gradient descent. The simplicity and lack of theoretical assumptions of this model enable a deeper understanding of the underlying dynamics of learning.

Architecture of the Model

Each speech category (e.g., voiced or voiceless) forms a Gaussian distribution (Box 1, Equation 4) along a continuous phonetic cue (e.g. VOT). A phonetic dimension (e.g. voicing), then, is modeled as the sum of the probabilities that a given VOT arose from some number (K) of Gaussians (Equation 5). For example, voicing in English is the sum of two (K=2) Gaussians with equal prior likelihoods (ϕ), means (μ) at 0 ms (voiced) and 50 ms (voiceless), and standard deviations (σ) of around 10 ms (voiced) and 20 ms (voiceless). Given a particular parameter-set, this model categorizes a given VOT by computing the likelihood of that VOT for each of the component Gaussians, with the most likely Gaussian representing its voicing category.

To model the learning process, model is exposed to a series of VOTs (one at a time). After each one, three sets of parameters are adjusted to maximize the likelihood of the entire model (the sum of individual Gaussians). The learning rules (Equations 6-8) for each parameter are the derivative of the log-likelihood with respect to the parameter being updated (e.g. μ), given the current input (VOT). They are derived from standard gradient descent assumptions. These learning rules are unsupervised—there is no error signal, only a set of rules that push the model to find the most likely parameter set. Most mixture models are trained on batches of data (not a single datapoint at a time), using more powerful methods such as expectation maximization or supervised techniques (e.g. discriminant analysis). However, supervised methods lack the psychological plausibility to study development, and batch-mode methods (like EM) do not permit a moment-by-moment analysis of learning. Gradient descent allows the learning rules to be applied iteratively (one input at a time) and without a teacher.

Because children do not know in advance how many categories may exist on a given dimension (languages like English have two voicing categories, while Hindi has three), the model should not have that knowledge either, but should be capable of learning any distribution. Thus, prior to training, the model is initialized with a large number of Gaussians (more than will be necessary). Each of these Gaussians starts with equal likelihoods ($\phi = 1/K$) and standard deviations (around 2). The locations (μ's) of each category in VOT-space are randomized to provide some variation for the learning rules to act on. The model is then exposed to a single VOT, all of the parameters are updated, and the next VOT is presented. These inputs are sampled from a dataset that is generated from the actual distribution of VOTs in a given language.

Testing the model: Competition

Initial tests were conducted to verify that the model could successfully learn the distribution of speech sounds. One hundred models were initialized with $K=20$, random μ's, and $\sigma=5$. These models were exposed to a random distribution of VOTs (taken from data on English reported by Lisker and Abramson [1964]). These models were initially tested to see if they learned the correct number of categories (the number of Gaussians for which ϕ_i was greater than .01). Surprisingly, none of the models learned two categories—most had substantially more ($M=10.25$). In fact, each category was typically represented by multiple overlapping Gaussians (Figure 1a). While one could argue for a distributed representation in which a given speech category is represented by a number of overlapping Gaussians, this model did not even approximate that well. On average, 67% of the Gaussians were 5 ms (of VOT) or more away from a correct μ for the dataset, and the number of Gaussians that were closest to a given category did not reflect the true base rate (ϕ): the voiced category was heavily overrepresented (M=52% of the ϕ-space) compared to the voiceless (M=17% of the ϕ-space).

This lack of success was perhaps not surprising—fitting a mixture of Gaussians with an unknown K and priors on each Gaussian (ϕ) has been a historically difficult problem. To solve it, we adopted a competition scheme loosely based on the competitive learning network of Rumelhart and Zipser (1986) (see McMurray, Aslin & Toscano, in preparation). Winner-take-all learning was adopted as a component of the learning rule for ϕ: only the most likely Gaussian in the mixture would alter its ϕ. Since the effective likelihood of a Gaussian is normalized ($\phi^2 / \Sigma\phi^2$) an increase in the winning ϕ resulted in an effective decrease of the losing ϕ's. Winner-take-all is also psychologically plausible: ϕ is analogous to frequency, so updating the frequency of only one category after a single exposure to a speech sound makes sense.

This simple change to the learning procedure resulted in dramatically improved performance. On a new batch of 100 models, 95% of them learned the correct (two category) solution, and 100% of these successful models learned the correct μ's and σ's. While this may seem a minor detail, the simplicity of this model highlights its importance: statistical learning mechanisms alone lack the power to learn multimodal distributions of speech cues; an additional competition mechanism is needed. While this has long been implicit in the connectionist literature, this approach makes it explicit. Thus, even the simplest statistical learning problems require a minimal amount of processing to be successful. As we shall demonstrate shortly, a dynamical systems approach may offer a method of implementing this more realistically.

The dynamics of development

This more successful version of the model was next examined developmentally. The first goal was to determine if it displayed the same developmental time-course as has been demonstrated empirically. The second, more important goal was to examine learning from a dynamical systems perspective, asking what happens moment-by-moment.

In terms of the developmental timecourse we first point out that the initial random distribution of μ's (Figure 2) ensures that newborn models will be able to discriminate a large number of phonetic contrasts—very few pairs of VOTs are assigned to the same category. Of course, as this model learns the training distribution, it eliminates unneeded Gaussians, and tunes the remaining ones to the parameters found in the language it is being exposed to—exactly as infants are seen to do between 6-12 months (Werker & Curtin, 2005 for a review).

To test this, we evaluated 100 models on discrimination of pairs of VOTs over development (see McMurray, Aslin & Toscano, in preparation). Discrimination was determined by a simple metric: if the same Gaussian was most likely for both VOTs, the model responded "same," otherwise it responded "different". During exposure, within-category VOT pairs (-40 and 0) were initially categorized as belonging to different categories and later as a single category (Table 1). The between-category (0 and 40) pair was classified as different throughout development, and a last comparison between a prototypical voiceless sound (40 ms) and a non prototypical one (80 ms) yielded moderate discrimination throughout training (analogous to Miller & Eimas, 1996).

This attunement should be part of any model that learns the statistics of the input. However, a richer empirical test will require a subtler computational analysis of the developmental timecourse. To probe this statistical learning mechanism more deeply, we conducted a trial-by-trial analysis of the trajectory of learning for a new set of 150 models. These models were initialized with K=5, random μ's, and starting σ's of 5, 10 or 15 (50 models for each starting σ). These models were exposed to a simple distribution of VOTs (μ's at [0, 50], σ's at [10, 10] and ϕ's at [.5 .5]) and their parameters extracted every 50 training cycles. Models were run until they had stabilized (either successfully or not)—usually around 6000 training cycles. Overall, only 79% of these models succeeded in learning the distribution, although this was due to small K's and larger starting σ's[1]. A trial-by-trial analysis of changes in μ for individual models is shown in Figure 3a. Not unexpectedly, there is a large amount of variance in μ initially (since they were set to random values), and rather quickly they approach the values in the training distribution (0 and 50) following a characteristic hyperbolic curve (also not surprising given the formula for $\Delta\mu$). The analysis of σ was more intriguing. Figure 3b displays σ as a function of training-cycles for models with a starting σ of 5 (smaller than that of the training set), 10 (the same size) or 15 (larger). In all three cases, σ very quickly grew to be much larger than would eventually be required (even if it already was larger), and then gradually settled towards the correct solution. This was a natural consequence of the learning rules. Any mismatch between the current data-point and μ resulted in both a shift in μ and also an increase in σ. Thus, even if the model started with the correct σ, σ increased temporarily until μ's settled.

This predicts a significant increase in the size (in phonetic space) of speech categories (followed by a slower decrease), particularly during the 6-12 month period when speech categories are changing the most. While this cannot be tested with head-turn and habituation methods (because they do not allow enough repeated measurements), the two-alternative anticipatory eye-movement paradigm (McMurray & Aslin, 2004) may offer a solution.

Perceptual Learning

An oft neglected facet to connectionist learning is that learning is always happening—every input forces a model to evaluate the relationship between its current parameters and the input, and adjust them if necessary. While this is obvious, it has important implications for current work on perceptual learning. A number of recent studies demonstrate that adult listeners are capable of shifting category boundaries of speech sounds with very short exposures (Kraljic & Samuel, 2005; Norris, McQueen & Cutler, 2003). Without exception, these studies have hypothesized a secondary mechanism to account for short-term plasticity. Thus, we asked whether the same developmental mechanisms (statistical learning) that account for initial speech category learning could also yield the sort of short-term perceptual plasticity shown empirically.

Twenty-five models were exposed to a bimodal distribution (μ=[0 50]; σ=[10 10]; ϕ=[.5 .5]). After 10,000 training cycles (sufficient for the model to settle), a new dataset was introduced with rightward-shifted categories (μ=[15 65]). Training alternated between these two datasets every 300 cycles, while μ and σ were recorded, for 6 alternations. This was repeated with 5 different values for η_μ, the rate of change in μ. Figure 4a shows μ (for the rightmost category) as a function of time and η_μ. It is clear that the model can quickly adapt to the new means—by the third alternation, some models adapt within 30-40 training cycles. This was particularly true for the models with a high η_μ. Models with a low η_μ were not able to shift their means all the way to either end and tended to drift towards the average of the two values.

To assess this short-term plasticity, we computed the time it took for the model to switch from within 5 ms of the correct mean to within 5 ms of the incorrect one. A hierarchical regression analysis examined switch-time as a function of η_μ, target (original μ's or shifted), and alternation number. It confirmed that this "switch time" was highly correlated with η_μ ($R^2_{partial}$-=.51, β=-.696, t(756)=28.1, p<.0001). The alternation number (1^{st}, 2^{nd}, 3^{rd}) was not correlated with switch time ($R^2_{partial}$=.001)—the model's plasticity did not change over the alternations. However, the model was significantly faster at shifting towards the original μ's ([0 50]) than the novel ones ($R^2_{partial}$=.08, β=-.21, t(756)=8.5, p<.0001), suggesting that the model may have not shifted entirely to the new values. This was particularly the case for low η_μ's, shown by an interaction of η_μ and target ($R^2_{partial}$=.012, β=-.27, t(753)=3.011, p=.003).

At least in principle, then, the same learning mechanism that accounts for initial category learning could operate quickly enough (upon established categories) to account for short term plasticity. However, it is possible that we have accomplished this by artificially increasing η_μ. Since slower learning is generally thought to yield a more stable model and a greater likelihood of success, increasing the learning rate to aid in short-term processing may hinder basic developmental processes. An analysis of these simulations revealed this not to be the case. Overall, 88% of the simulations correctly learned the two-category solutions and a logistic regression confirmed that success was not correlated with η_μ (Wald(1)=.096, p>.2). While increasing η_μ beyond about 20 led to instability and failure at learning, switch time appears to flatten at η_μ's above 16. Thus, there is no disadvantage to high η_μ's, and no additional plasticity advantage for very high η_μ's. The same statistical learning mechanism that accounts for developmental change can also account for short-term perceptual learning, without any loss of efficacy in the basic developmental process.

A connectionist approach

The MOG approach provides a distillation of the statistical learning processes. In doing so, it reveals the dynamics of learning and also links short-term plasticity and developmental mechanism. It is also points out that statistical learning alone is not sufficient for successful learning: competition among categories is required. The one-step, discrete winter-take-all competition, however, represents a simplified, implausible form of competition. A more realistic competition framework may provide an opportunity to implement statistical learning in a more dynamical framework. This could simulate millisecond-by-millisecond processing and illuminate the relationship between learning and online processing.

Thus, we developed a Hebbian Normalized Recurrence (HNR) network (McMurray & Spivey, 1999) for speech categorization (Figure 5). The 130 input nodes formed a topographic map of the input dimension (e.g. VOT). Each node had a Gaussian tuning curve whose mean (characteristic VOT) was organized from low (-40 ms of VOT) to high (80 ms). Input nodes were initially randomly connected to output nodes. After receiving an input, activation was weighted and added to the current output (Box 2, Equation 9). Output nodes compete with each other with a squared-normalization rule (Equation 10). This formalism approximates winner-take-all competition in the limit (if it was run repeatedly) but also allows for partial (not complete) competition and creates intermediate states leading up to winner-take-all. Output nodes pass activation back to the input (Equation 11) via a set of feedback connections. Importantly, feedback activation only affects *active* input units (multiplicatively), which introduces some nonlinearity into the system and prevents hallucination. Activation cycles back and forth in this way until the output nodes settle (stop changing). Importantly, at every cycle a small amount of Hebbian learning takes place (Equation 13) strengthening the weights connecting active inputs and outputs, and decaying weights for which either are inactive.

Preliminary tests of this architecture showed that 48 of 50 models trained on the distribution of VOTs from Lisker and Abramson (1964) learned two categories. Moreover the model showed a similar developmental pattern to the empirical data and the MOG model: displaying multiple stable categories early in development (M=4.2 stable categories halfway to the end of learning). Finally, the quadratic normalized competition in the output layer was required for learning. Without it, none of the 50 models learned the distributions.

Thus, this model has largely the same features as the MOG in terms of learning and competition. However, the recurrent feedback/competition loop also enables us to examine the time-course of processing. Figure 6a displays the model's activation of voiced and voiceless output nodes as a function of time after receiving a voiced input. It's clear that the model is temporarily considering both alternatives, but eventually settles on the correct target. Figure 6b illustrates the graded nature of this model's speech categories—the activation for the competitor systematically increases as VOT approaches the category boundary.

This provides a surprisingly good fit to eye-movement data from McMurray, Tanenhaus and Aslin (2002). This study used an eye-movement paradigm to examine subjects' moment-by-moment word recognition as a function of within-category VOT variation. The pattern of fixations over time revealed systematic gradiency (Figure 6c): as VOT approached the category boundary subjects were more likely to fixate a competitor (e.g. *pear* when the target was *bear*)—a good fit to the model activation. This fit suggests that the gradient pattern of lexical activation may be the result of statistical learning of speech cues. Given emerging results that this gradiency may be crucial to normal adult online comprehension mechanisms (McMurray, Tanenhaus & Aslin, in preparation), then, this developmental mechanism may prepare the system for efficient online processing.

The last simulations with the HNR network examined statistical learning more closely to determine if the learning mechanism posited by the MOG differed from the connectionist simulation. Topographically organized Hebbian learning networks of this sort approximate a MOG by summing the Gaussian inputs along weights (priors). However, unlike the MOG, this model can only learn by adjusting the weights (analogous to ϕ)—μ and σ are not learnable. This more constrained learning, and our intuition that connectionist learning is inherently slow, suggested that this network may not be capable of the short-term plasticity shown in the MOG.

This was examined in a series of 200 simulations. Fifty models were run at each of four different learning rates. After 3000 trials of training, means were alternated as before, from [0 50] to [15 65] and the model was tested every 5 training cycles. Results (Figure 7a) suggested that the model exhibited a similar degree of plasticity to the MOG. It is clear that with at least some learning rates, the model can be highly plastic. However, at the highest learning rate (.005), the model was both very plastic and very unstable. In fact the data presented for this condition represents only the 6% of the models: 94% failed to learn the categories.

While the model was able to shift its categories in real-time, an analysis of the dynamics of this shift revealed that the intermediate states (between alternations) showed strikingly different patterns. Figure 7 shows the HNR network (Panel B) and the Mixture of Gaussians (Panel C) shifting from means of [0 50] to [15 95]. Very clear differences emerge—while the MOG maintains a Gaussian shaped category the entire time, the connectionist network adopts a much stranger shape. In fact, because it represents categories as the sum of Gaussians (rather than a single one), it can learn any number of strangely shaped categories. Thus, even within the simple mechanism of statistical learning, substantial differences exist between instantiations.

Case Study 1: Discussion

The first case study examined what appears to be a relatively simple case of statistical learning: the acquisition of speech categories. Preliminary simulations with the MOG approach, however, suggested that statistical learning alone is insufficient—an additional competition mechanism is needed. While connectionist approaches often gloss over such architectural details they are clearly necessary for learning. However, they also offer an opportunity to consider the role that online processing dynamics may play.

Across both models, a consideration of the dynamics of learning itself yielded an opportunity to understand the power and scope of this developmental mechanism. We saw how the width of the categories in the MOG increases early in development—even if the categories are already larger than those of the language-input. In addition, both classes of simulations were able to simulate the short-term plasticity without a secondary mechanism (although their intermediate states suggest they may represent somewhat different forms of statistical learning). Thus, perhaps learning itself is best thought of as an in-the-moment process.

Case Study 2: Word learning

The second case study offers a different view of the relationship between learning and in-the-moment processes. Here we start with what appears to be entirely an online, in-the-moment process, and offer a simulation that demonstrates how connectionist learning principles can benefit from and influence this process. Our discussion centers on early word learning, an ideal domain in which to study this because there are clear components of long-term learning (vocabulary growth) and in-the-moment behavior (identifying referents) that may interact.

Empirical Results

Between the ages of 1 and 5, children acquire new words at an extraordinary rate, learning as many as 50,000 words by their first birthday (Carey, 1978). This immense rate of learning has led many to posit a need for specialized (perhaps innate) mechanisms that are directed towards forming sound/meaning mappings. One such proposed mechanism has become known as fast-mapping. While the term fast-mapping has been used as a catch-phrase to describe any word learning that occurs with minimal exposure, we use it here to describe a specific type of word learning based on mutual exclusivity. Fast-mapping is based on the idea that children make the assumption that every object has only one name. Carey and Bartlett (1978) asked children to retrieve the *chromium* tray from a set containing a chromium (olive-green) tray and a blue one. Children retrieved the correct tray, suggesting that they had inferred that *chromium* referred to the color for which they did not have a name. This form of fast-mapping has been shown to be operative in children as young as 17 months (Halberda, 2003).

It is not clear, however, that this sort of fast-mapping represents learning. While children can identify novel words in this sort of task, this may actually represent in-the-moment problem-solving, not learning. Assessing this requires the researcher to determine if words that are identified in a fast-mapping task are retained. Existing experimental paradigms are not clear on this fact. Some studies have asked this question by presenting children with a subsequent trial of the same sort (the novel words paired with a known word), raising the possibility that children are simply solving the problem again (e.g., Mervis & Bertrand, 1994). Other studies review the name object linkages before testing (e.g., Goodman, McDonough, & Brown, 1998). None of these studies rule out the possibility that fast-mapping represents a form of in-the-moment problem solving, and they do not provide evidence that children are truly learning the words.

A recent study by Horst and Samuelson (submitted), however, provides convincing evidence that children do not retain word-referent mappings formed in fast-mapping contexts. In this study, children were initially given a series of typical fast-mapping trials in which they were presented with two familiar objects and one novel object. Children were asked to retrieve the *fode,* and correctly selected the novel object 68% of the time. Following a five minute break, the children were then presented with the *fode* and two other novel objects. When asked to retrieve the *fode* in this situation, children were at chance, despite their being able to correctly select it in the initial fast-mapping trial (Figure 9a). This suggests that their behavior on the fast-mapping trial did not result in much (if any) learning. Subsequent experiments showed that this was the case even when the number of novel names was reduced to one, when the number of total trials was reduced to three, and when the experimenter provided feedback to the child. This evidence makes the case that fast-mapping may represent only in-the-moment behavior, not a specialized learning mechanism (or a learning mechanism at all).

While fast-mapping does not offer the “one-shot” learning that has been trumpeted in the literature, it is possible that this in-the-moment behavior is still related to word learning. Here we explore a subtle version of this hypothesis. Our simulations demonstrate that in-the-moment problem solving may 1) provide a training signal to permit slow connectionist learning, and 2) allow the system to behave as if it knows a word, despite relatively little knowledge. These simulations illustrate the deep relationship between online processing and learning.

Model Rationale

The HNR architecture is in many ways the ideal architecture for modeling this process. As we've shown, this class of models is capable of learning subtle regularities in the environment using simple unsupervised mechanisms. Thus, it may be able to capitalize on statistical associations (between sounds and visual stimuli) to learn words. We propose a somewhat associationist account: consistent co-occurence between auditory and visual forms permit at least some word learning. For example, every time *dog* is heard, the only visual stimulus that is *consistently* present is also a *dog*. Although other visual stimuli may be available, *dog* will be more consistent over time. This invokes behaviorist explanations of word-learning which were largely ruled out by arguments that unambiguous reference is not possible in most visual scenes without innate constraints (though see Smith, 2000). While this may be true in a single presentation, word learning does not happen in a single presentation—coocurrence is consistent and reliable over the long-term.

However, learning is not the whole story. Children are engaging in a complex constraint-satisfaction process when determining the referent of a novel word given the contents of their lexicons and the visual context. This is the domain of dynamical systems and again the HNR architecture offers a solution. While in the previous case study, the competition/settling algorithm only captured the competitive dynamics, non-learning versions of this model (e.g. Spivey, in press) have been shown to solve a host of problems in which multiple, graded constraints must be integrated probabilistically in real time. These include complex multidimensional categorization and visual search. Thus, this implementation of online constraint-satisfaction should be able to determine the correct word/object mappings, in the moment, given partial input and partial knowledge—much as infants appear to do.

Thus, the current model is built around an associationist statistical learning core which implements referent selection using a simple, dynamic probabilistic constraint satisfaction device. The present simulations will show that these two components interact at multiple time-scales to generate learning and development that exceeds what either could do alone.

Architecture of the Model

The HNR network used for word learning is shown in Figure 8 (see McMurray, Horst & Samuelson, in preparation, for more details). Two banks of localist input units (auditory and visual) were used. In each array, a single node was designated for each possible spoken word, and for each visual referent. Auditory and visual arrays are independent (they cannot directly pass activation to each other). On a given trial, nodes corresponding to the auditory and visual inputs are activated (including nodes for visual competitors). Input arrays then pass activation to the decision layer (equivalent to the output layer in the prior simulation) (Box 3, Equation 14). As before, this decision layer is squared and normalized (Equation 15), and activation is fed back to the two input layers (Equation 16), normalized (Equation 17), and the cycle continues.

As in the prior network, the weights (auditory→decision and visual→decision) were initially random. On each cycle of processing, a small amount of Hebbian Learning is applied to the weights (Equation 18). Note that this learning rule differs slightly from that of the prior network, specifically in the way weights decay. While the previous rule (based on Rumelhart & Zipser, 1986) decayed all of the weights equally, the current rule (adapted from ART, Grossberg, 1976) only decays weights that are connected to either an active input unit or an active decision unit. This restricted weight-decay function maintains plasticity in a large number of weights (those connected to inactive input or output nodes) allowing the system to learn new words.

The Strength of Associative Learning

Early simulations with this architecture revealed that it could easily learn lexica even when the consistency between auditory and visual units was quite low. Two-hundred forty models were trained on a 20 word lexicon. Consistency was manipulated (on each trial) by randomly activating some of the visual competitors with a competitor-likelihood ranging from 0 to 0.95. Table 2 shows the model's performance at each competitor-likelihood. The model performed above 90% until a competitor-likelihood of 0.4 and was close to 70% with competitor-likelihoods of 0.75 (in which ¾ of the lexicon was visually present on any trial during learning). Surprisingly, even at 0.95 competitor-likelihood, the model's performance was better than chance (M=.405, t(19)=3.01, p=.006). Simple statistical learning mechanisms can extract word/object pairings from extremely noisy input, particularly when combined with online constraint-satisfaction principles.

Fast-Mapping and Word Learning

The next simulations tested the model on an explicit fast-mapping task. Twenty models were initialized with 15 auditory and visual nodes and 90 decision units[2]. These units were divided into three classes of five words: familiar words, novel words (for novel fast-mapping trials) and held-out words (which were never activated). The models were initially trained on only the familiar words, to simulate an infant who had only learned a few words. Next, the model was tested in 15 testing trials. These trials consisted of three types, familiar and novel fast-mapping trials, and retention trials (Horst & Samuelson, submitted).

Familiar words were tested by exposing the model to a familiar word (auditorially), along with the visual target, and one familiar and one novel visual competitor. After presentation of the input, the model was cycled until decision units stopped changing. Because the feedback / competition cycle suppressed activation for visual competitors, this allotted more attention to the target. Thus, the visual unit that had the most activation was the object that the model selected. On each novel fast-mapping trials (randomly interspersed among the familiar trials), the model was tested on one of the five novel words paired with the correct (novel) referent, and two visual competitors. Again, selections were based on the visual array. Retention trials were conducted *after* the fast-mapping trials and were designed to determine if the model retained anything about the novel words during the fast-mapping trials. The model was tested on each of the five novel items paired with a second novel item and one of the held-out objects. Here, if the model had learned during the fast-mapping trials, it should be able to select the correct referent. Importantly (and unlike many connectionist models), learning was not turned off during testing, allowing the model to potentially learn during the test period (as infants might do in such an experiment).

The model's performance (Figure 9b) was surprisingly close to behavioral results from Horst & Samuelson (submitted) (Figure 9a). While the model's overall performance was better than that of the infants (M_{model}=96%, $M_{infants}$=72% for familiar words), the pattern of results was identical. On the familiar and fast-mapping trials, the model performed quite well (Familiar: M=.96, SD=.10; Fastmapping: M=.86, SD=.16). These two conditions were barely significantly different (t(19)=2.1, p=.047) and both were significantly better than the chance score of 1/3 (Familiar: t(19)=26.8, p<.001; Fastmapping: t(19)=14.7, p<.001). However, like the infants, the model performed at chance on the retention trials (M=.42, SD=.24, t(19)=1.6, p>.1).

It is tempting to conclude from the literature and from the fast-mapping results that the ability to identify the correct objects means that the infant (or the model) has complete linkages between auditory and visual input. Conversely, it is tempting to conclude from the model's (or infant's) failure to complete the retention trials that no true learning had occurred at all—only in-the-moment constraint satisfaction. However, the subtler conclusion is that some learning occurred, but in-the-moment processing augmented it (with respect to in the moment behavior). To evaluate these hypotheses, we conducted an analysis of the weight-matrix to determine how much learning actually occurred.

The complete weight-matrices had been stored at three points during training for the above simulations: 1) prior to any exposure, 2) after the initial exposure but before test, and 3) after testing on the fast-mapping trials. The amount of learning was computed by taking the RMS difference (the mean of the squared change in each weight) between portions of the weight matrix across these snapshots. This RMS difference was computed for subsets of the weight matrix (e.g., the novel words), not the full matrix, in order to determine the amount of learning for a given class of items (familiar, novel and held-out). The held-out words served as a control, revealing the amount of weight change with no direct input (i.e., weight-decay). The familiar words reveal how much weight change we should expect for a word to go from unknown to known. Thus, a comparison of fast-mapped words to these groups will reveal how much learning has occurred over the fast-mapping trials relative to behaviorally understood benchmarks.

Figure 10a compares the amount of learning that occurred during the initial 3000 training epochs for the familiar(ized), fast-mapped, and held-out words. Not surprisingly, very little learning occurred for the fast-mapped and held-out words (RMS_{novel}=.08; $RMS_{heldout}$=.08). At this point the model had not been exposed to any of these words, and the learning rule only decays weights connected to active nodes. Figure 10b shows the same amount of change over the testing trials (both familiar and fast-mapping). First notice the y-axis scale—there was quite a bit less learning overall here than during training. Nonetheless, a comparison of the familiar and the fast-mapped words to the RMS of the held-out words (the control condition, since RMS [a squared value] cannot be directly compared to 0) revealed that both classes of words did receive statistically significant learning (Familiar: t(19)=13.1, p<.001; Fast-mapped: t(19)=39.0, p<.001).

Thus, the model was learning something about the fast-mapped words during fast-mapping trials. However, this overall small amount of learning (M=1.9e-6), did not even approach what constituted learning on the familiar words (M=.76; t(19)=97.8, p<.001). The amount of learning on the fast-mapping trials does not qualify as word learning (if we classify performance on the familiar words as word-learning). Thus, we can conclude the following. First, online integration mechanisms permit the model to "fast-map" novel words in some situations but this represents largely in-the-moment processing, not learning. Second, this excellent performance in the model accounts for children's performance in this task. Third, because the training regime (presentation of a single auditory token coupled with a number of possible visual referents) was identical to the fast-mapping trials, the presence of some learning on these trials, then, suggests that while fast-mapping is accounting for the in-the-moment, overt behavior, it may also be important for learning over the long haul.

The Interaction of Learning and Online Processing

The final simulations were run in order to understand the relationship between the online processing (fast-mapping) and learning. Eighty networks were trained on a small 20 word lexicon and tested every 25 training epochs for 8000 cycles. In twenty of these networks both the recurrent feedback and competition (in the decision layer) were turned off to see if Hebbian learning alone was sufficient to learn a lexicon (the NULL model). Twenty were run with the competition alone (NORECURR), and twenty were run with the feedback alone (NOCOMP). The remaining twenty were run with both processes, a fully functional model (FULL).

In the previous two simulations, testing was conducted by analysis of which visual units were most active at the end of cycling (simulating the focusing of attention to the correct object). However, without recurrence the visual units did not change in response to the auditory input—

selection based on them would be meaningless. Therefore, rather than analyzing model “performance”, we adopted a more competence-based definition of success. A single auditory input was activated and the decision unit with maximal activation was computed. Next the corresponding visual unit was activated and the decision unit with maximal activation was determined. If these two were the same decision unit, then the model had mapped the visual and auditory units onto the same decision unit—it “knew” the correct association.

The results from this simulation were surprising. In figure 11a, it is clear that, early in training, the fully functional model (FULL) underwent more learning than the other models. However, this initial phase got it to only around 60% correct. While it eventually got quite close to 100%, this later phase of learning was markedly slower—in essence it learned the first 10-11 words quite quickly and then took more time to attain the remaining words. The model without competition (NOCOMP), on the other hand, was slightly delayed and did little learning until around 1000 epochs. However at this point it learned the complete lexicon quite fast. The model without recurrence (but with competition) (NORECURR) was seriously hindered and took much longer than the other models, and the NULL model was unable to learn at all.

It was a bit surprising that, of the two models with recurrence, NOCOMP ultimately outperformed FULL, asymptoting at 100% compared to 96% for FULL. Figure 11b, however, explains the different performance. This figure displays (for just the two classes of models with recurrence) the percent correct using the prior method (based on activation in the visual array). As before, FULL learned faster than NOCOMP. However, this time, it is FULL’s performance that asymptotes at 100%, while NOCOMP never reaches perfect performance.

This suggests a number of findings. First, it is interesting that the NOCOMP model does not reach perfect performance despite apparently perfect weights. However, our measure of competence is only based on the connection strength between the target and the decision units (it ignores the connections of the competitors). Thus, while it may have all of its targets associated properly, weights to the competitors may be quite strong as well and without competition these cannot be suppressed. Competition in the FULL model, on the other hand, forces the decision units to reach a single active unit. This in turn forces competitor weights to decay, and allows the model to ultimately learn a better solution.

Second, the fact that the FULL model never arrives at perfect knowledge (despite perfect performance) is probably the result of the fact that competition allows the network to, in-the-moment, figure out the correct solution—imperfect weights are sufficient. Having slightly imperfect weights may even allow the network to encode correlations between visual units (e.g. fork is often co-present with spoon), although in this case, such correlations would be spurious.

Third, and most importantly, the lack of learning by the NULL model and very slow learning in the NORECURR models suggests that the online processing (fast-mapping) is essential to the learning process. Competition is clearly an asset (and without recurrence, required), but recurrent feedback loops between audition, decision, and visual units play a more critical role. This active comparison of decision-unit representations with available auditory and visual inputs is the fast-mapping process: in-the-moment integration of multiple constraints. This processing (fast-mapping) is, in a sense, serving as a teaching signal for the Hebbian learning. By forcing the model to arrive at the correct solution given the constraints, the model now has a more accurate representation upon which to learn—in-the-moment processing is generating the solution and this is entrained in the weights. Thus, while we have been referring to learning as unsupervised, perhaps it is better thought of (in this case) as self-supervised.

Finally, the amount of time the model spends processing a single word is related to learning. Figure 11c shows the average number of processing cycles it took each model to settle during the training sequence. For both the FULL and NOCOMP models, processing time had peaks at around the time when the greatest change in performance was occurring. These peaks are notable for their strength (upwards of 3500 cycles compared to 10-15 for a fully functional model) and their brief duration. Our interpretation is that learning initially destabilizes the online settling process. In fact, the settling time on the only first trial (in which the weight matrices were random) averaged 1039 cycles (SD=208) with learning and 248 (SD=22.2) without. This greater number processing cycles, however, allows more learning to occur (since learning occurs on every cycle) which ultimately leads to better weight representations and faster processing.

Moreover, although difficult to see in Figure 11c, these architectural considerations have important effects on even the fully trained model. The FULL model averaged 16.6 processing cycles at the end of training, while NORECURR averaged only 12.58. Clearly recurrence delays processing a bit as the model considers both inputs in a more integrated sense, though this ultimately benefits performance. However, the NOCOMP model averaged 46.7 cycles suggesting that competition may be crucial to arriving at the correct solution quickly.

Case Study 2: Discussion

This model contributes an important theoretical explanation to the results of Horst and Samuelson (submitted). It demonstrates that while learning occurs during fast-mapping, this learning is only partial—the seemingly good performance is largely due to online processing dynamics. However, on the flip side, it also demonstrates that online processing dynamics significantly augment slow connectionist-style learning—you can have slow learning and good performance too. At a computational level, this validates the fact that competition (and in this case competition plus recurrent feedback) is essential for learning. In fact, this mechanism is powerful enough that basic associative mechanisms were able to extract word-object mappings even with half of the lexicon visually co-present. In a sense, then, this online, dynamical constraint satisfaction provides a teaching signal for associative Hebbian learning. This competition/recurrence, however, contributes to more than learning. It also provides useful in-the-moment behavior: the task of the child is not simply to learn but also to communicate, behave and act in the moment.

At a broader level, this model suggests some powerful ways in which online processing and learning interact to drive development. Here, online processing forces the model to arrive at a correct referent, a solution which can subsequently be entrained in the weights. Importantly, statistical learning alone cannot learn words effectively, and theoretical approaches based solely on online processing fail to relate fast-mapping behavior to learning. An interactive approach that considers both mechanisms offers a compelling alternative.

Conclusions

We began with a definition of developmental mechanism as an explanation of developmental change that bridges multiple levels or domains and discussed two such mechanisms offered by connectionism and dynamical systems theory: statistical learning and moment-by-moment processing. We have shown that not only are these mechanisms complementary, but that a tighter integration offers a lot to our understanding of development.

Our discussion of the MOG and HNR approaches to speech categorization revealed first that no statistical learning could take place without some form of competition. This was extended with a more dynamic form of competition in the HNR network enabling it to fit reveal the temporal dynamics of online processing, and to explain the underlying gradiency seen there as the product of developmental mechanism. We also demonstrated that a dynamic approach to learning can yield novel insight, particularly the with respect to the MOG in that the parameterization of learning (in terms of parameters of a Gaussian distribution) enabled a clearer understanding of different aspects of categorization change over time. Finally, and most importantly, our analyses of perceptual learning phenomenon reveal that in some cases statistical learning *can be* online processing. The same statistical learning mechanism that accounts for developmental change may be an integral part of short-term phenomena. Thus, an understanding of in-the-moment behavior may require an understanding of developmental mechanism.

Our discussion of word learning extended the HNR architecture in a way that clearly illuminated the complex relationship of in-the-moment behavior and long-term learning. Here we modeled the ability of children to quickly evaluate multiple constraints and select the correct referent, and the seemingly contradictory failure in their ability to retain such knowledge (Horst & Samuelson, submitted). Our simulations captured both findings but provided an explanation that is relevant to understanding the relationship between dynamical and connectionist mechanisms. We demonstrated that even if learning is slow (in a connectionist sense), in-the-moment behaviors can buttress it, allowing models (and children) to behave in a way that exceeds their stored knowledge.

More importantly, we also demonstrated the converse: online processing is crucial for learning. Learning was substantially improved when the competition and recurrent feedback were included as part of the model. In effect, these online processes (the probabilistic integration of auditory and visual input with knowledge stored in the weights) pushed the model to the correct answer and served as a teaching signal for learning—self-supervised learning. It was this powerful learning that likely accounted for the model's ability to learn word-object mappings even when 90% of the lexicon was visually co-present with the target during training. This argues for basic associative learning mechanisms, when supported by simple constraint satisfaction. More importantly, it shows that these processes are not additive, but interactive.

As a package, these simulations provide no unified framework for combining dynamical systems and connectionist frameworks (although the HNR architecture is promising in this regard). However, they do illustrate that a combination of such mechanisms promises to yield a far richer view of development than either alone. In particular, they reveal that, while individually online processing and statistical learning may provide only weak solutions to developmental problems, when combined they introduce a powerful new framework for understanding behavior, both moment-by-moment, and the ways in which it changes over the long haul. As developmental scientists, we must acknowledge a singular developmental process, even as we attempt to divide it into mechanisms like statistical learning and online processing for ease of study. These simulations were conducted in this spirit, but in the whole they support this singularity. Our examination of statistical learning and online processing suggests they are two highly interactive sides of a multifaceted coin. But perhaps they are two regions of a smooth sphere?

Acknowledgements

The authors would like to thank Michael Spivey, Robbie Jacobs and Richard Aslin for discussions leading to the computational architectures presented here. Portions of this work were funded by NIH F31DC006537-01 to BM and NIH 5R01HD045713 to LKS.

References

Bechtel, W. (1997) Embodied connectionism. In Johnson, D., Erneling, C. (Eds). (1997). *The future of the cognitive revolution.* (187-208). New York: Oxford University Press.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan & A. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 264-293). Cambridge, MA: MIT Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15*, 17-29.

Elman, J. (1990) Finding structure in time. *Cognitive Science, 14,* 179-211.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D. and Plunkett, K. (1996) *Rethinking Innateness: a Connectionist Perspective on Development*. Cambridge, MA: The

MIT Press.

Elman, J. and Zipser, D. (1988) Learning the hidden structure of speech. *Journal of the Acoustical Society of America, 83(4),* 1615-1626.

Espy-Wilson, C. (1992) Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *Journal of the Acoustical Society of America, 92(2),* 738-757.

Goodman, J. C., McDonough, L., & Brown, N. B. (1998). The role of semantic context and memory in the acquisition of novel nouns. *Child Development, 69*(5), 1330-1344.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, Part I: Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121–134.

Guenther, F. and Gjaja, M. (1996) The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America, 100,* 1111-1112.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87(1),* B23-B34.

Horst, J.S., and Samuelson, L.S. (submitted) Fast mapping but not fast word learning: referent selection without retention.

Kraljic, T., and Samuel, A. (2005) Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51 (2005),* 141-178.

Lisker, L. and Abramson, A.S. (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word, 20,* 384-422.

Maye, J., Werker, J.F. and Gerken, L. (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* 101-111.

McClelland, J. and Elman, J. (1986) The TRACE model of speech perception. *Cognitive Psychology, 18(1),* 1-86.

McCloskey, M., and Cohen, N. (1989) Catastrophic interference in connectionist networks: The sequential learning problem. In G.H. Bower (Ed.), *The psychology of learning and* *motivation: Advances in research and theory* (Vol. 24, pp. 109-165).

McMurray, B., and Aslin, R.N. (2004) Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6(2),* 203-229.

McMurray, B. and Aslin, R.N. (2005) Infants are sensitive to within-category variation in speech perception. *Cognition, 95(2),* B15-26.

McMurray, B., Aslin, R.N., and Toscano, J. (in preparation) Statistical learning of phonetic categories: computational investigations with a mixture of Gaussians.

McMurray, B., Horst, J., and Samuelson, L. (in preparation) The interaction of learning and online constraint-satisfaction: a computational approach to early word-learning.

McMurray, B. and Spivey, M. (1999) The categorical perception of consonants: The Interaction of Learning and Processing. *Proceedings of the Chicago Linguistics Society, 35,* 205-219.

McMurray, B., Tanenhaus, M.K. and Aslin, R.N. (2002) Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86(2),* B33-B42.

McMurray, B., Tanenhaus, M.K., and Aslin, R.N. (in preparation) One foot on the garden-path, and one foot on the highway: Lexical sensitivity to subphonemic detail facilitates phonetic ambiguity resolution.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the Novel Name Nameless Category (N3c) Principle. *Child Development, 65*(6), 1646-1662.

Miller, J.L. (1997) Internal structure of phonetic categories. *Language and Cognitive Processes, 12,* 865-869.

Miller, J.L. and Eimas, P.D. (1996) Internal structure of voicing categories in early infancy. *Perception & Psychophysics, 58(8),* 1157-1167.

Nakisa, R. and Plunkett, K. (1998) Evolution of a rapidly learned representation for speech. *Language and Cognitive Processes, 13(2&3),* 105-127.

Norris, D., McQueen, J., and Cutler (2003) Perceptual learning in speech. *Cognitive Psychology 47 (2003),* 204-238.

Olshausen BA, Field DJ (1996). Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature, 381,* 607-609.

Peterson, G.E. and Barney, H. (1951) Control methods used in the study of vowels. *Journal of the Acoustical Society of America, 24(2),* 175-184.

Rumelhart, D.E. and McClelland, J.L. (1986) On learning the past tenses of English verbs. In (McClelland, J.L. et al., eds) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol. 2* (pp. 216–271), Cambridge, MA: The MIT Press.

Rumelhart, D.E. and Zipser, D. (1986) Feature discovery by competitive learning. In D. E. Rumelhart & J. McClelland (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1* (pp. 151-193). Cambridge, MA: The MIT Press.

Samuelson, L. K. (2002). Statistical Regularities in Vocabulary Guide Language Acquisition in Connectionist Models and 15-20-Month-Olds. *Developmental Psychology 38,* 1016-1037.

Schlessinger, M., Parisi, D. and Langer, J. (2000) Learning to reach by constraining the movement search space. *Developmental Science, 3(1),* 67-80.

Smith, L. (2000) Avoiding associations when it's Behaviorism you really hate. In R. Golinkoff, et. al (Eds.) *Becoming a Word Learner: A Debate on Lexical Acquisition.* (169-174)

Smith, L., and Samuelson, L. (2003) Different is good: connectionism and dynamic systems theory are complementary emergentist approaches to development. *Developmental Science, 6(4),* 434-439.

Smolensky, P., Mozer, M.C., & Rumelhart, D.E. (Eds.) (1996). *Mathematical perspectives on neural networks.* Mahwah, NJ: Lawrence Erlbaum Associates.

Spencer, J., and Thelen, E. (2003) Connectionist and dynamic systems approaches to development. [Special Issue] *Developmental Science, 6(4).*

Spivey, M. (in press) *The Continuity of Mind.* New York: Oxford University Press.

Tabor, W., and Tanenhaus, M.K. (1999) Dynamical Models of Sentence Processing. *Cognitive Science*, 23(4), 491-515.

Thelen, Esther. (2000) Grounded in the world: Developmental origins of the embodied mind. *Infancy, 1(1)*, 3-28.

Thelen, E. and Smith, L.B. (1996) *A Dynamic Systems Approach to the Development of Cognition and Action.* Cambridge, MA: The MIT Press.

Thelen, E., Schöner, G., Scheier, C., and Smith, L. (2001) The dynamics of embodiment: A dynamic field theory of infant perseverative reaching errors. *Behavioral and Brain Sciences, 24*, 1-86.

Turk, M., and Pentland, A. (1991) Eigenfaces for recognition. *Journal of Cognitive Neuroscience. 3(1)*, 71-86.

Van Geert, P. (1998) A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review, 105(4),* 634-677.

Ward, L.M. (2002). *Dynamical Cognitive Science*. Cambridge, MA: MIT Press.

Werker, J.F. and Tees, R. (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Werker, J.F., and Curtin, S. (2005) PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development, 1(2)*, 197-234.

Footnotes

[1] Interestingly, this success rate was also related to the starting σ. While models starting with σ ‘s of 5 or 15 did fairly well (σ=5: 42/50 succeeded; σ=15: 44/50 succeeded), the models starting

with $\sigma=10$ (the correct σ) were largely responsible for the low success rate (33/50 suceeded). This was reliable ($\chi^2(2)=8.37$, $p=.015$).

---

[2] This was far more decision units than the model would eventually use. However, by using a large number, we ensure that initially (when the weights are random) the model is more likely to activate different decision units for different auditory and visual inputs. This helps minimize the chance of overgeneralization (in which a single decision unit maximally responds to two words).

Tables

Comparison (ms VOT)	**Proportion discrimination after X training cycles.**					**Description**
	1000	***2000***	***3000***	***4000***	***5000***	
-40 vs. 0	0.88	0.81	0.6	0.52	0.37	Spanish prevoiced vs. voiced.
0 vs. 40	0.99	1	0.99	1	1	English voiced vs. voiceless.
40 vs. 80	0.76	0.71	0.65	0.61	0.7	English voiceless vs. extreme.

Table 1: Discrimination of within-category (-40 vs. 0), between-category (0 vs. 40) and extreme (40 vs. 80) VOTs as a function of training cycles.

Competitor-likelihood	% Correct
0	1
0.05	0.998
0.1	0.993
0.15	0.995
0.2	1
0.25	0.988
0.3	0.963
0.35	0.918
0.4	0.900
.45	0.865
0.5	0.798
0.55	0.793
0.6	0.758
0.65	0.75
0.7	0.703
0.75	0.698
0.8	0.588
0.85	0.523
0.9	0.495
.95	0.405

Table 2: Average performance of models as a function of competitor-likelihood. Competitor-likelihood refers to the likelihood that each visual competitor is visually present (with the target) during training. A competitor-likelihood of .5, for example means that, on average, half of the words in the lexicon will be copresent with the target on any given training trial.

Figures

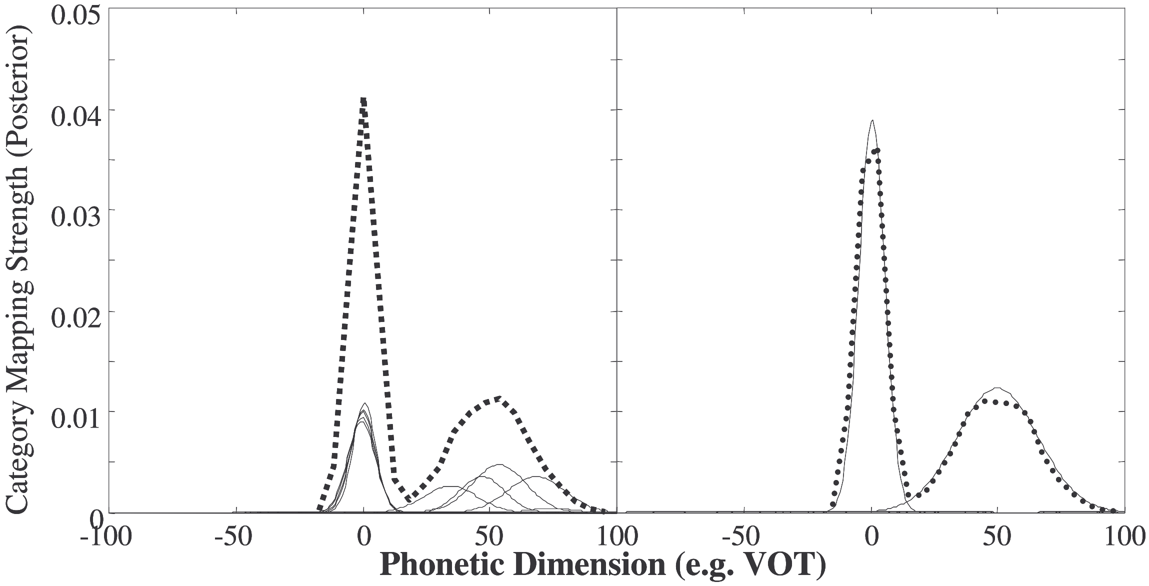


Figure 1: Comparison of models after 10,000 learning trials with and without winner-take-all competition. A) Model without winner-take-all learning. Multiple, overlapping, Gaussians are present for each category of the input (dashed line). B) Model with winner-take-all learning. Only one Gaussian remains for each category.

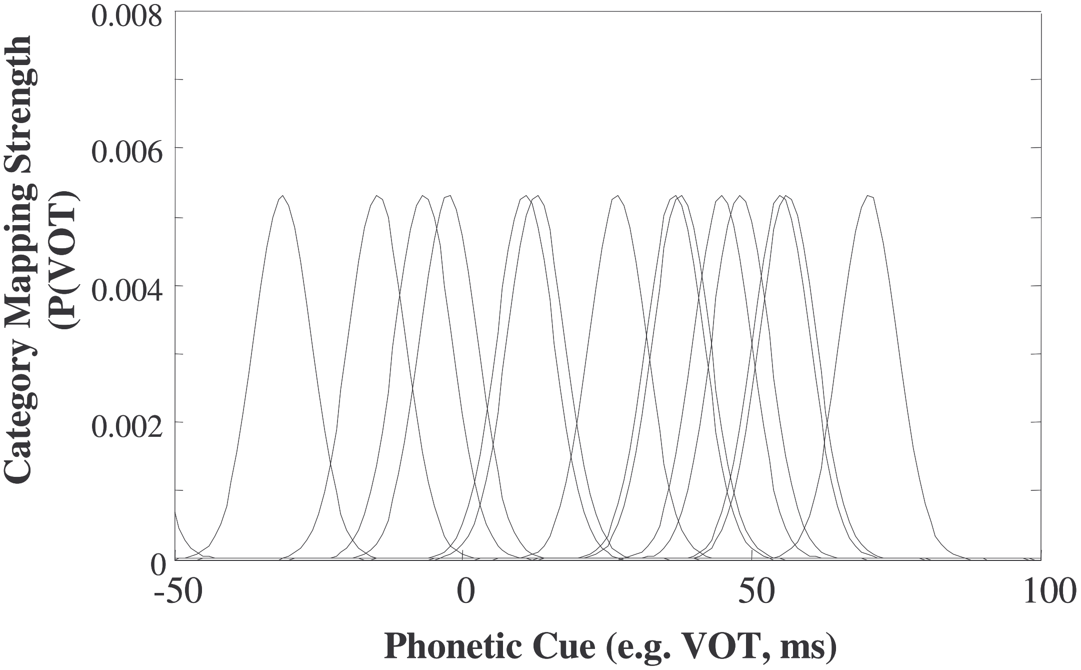


Figure 2: One initialization for a K=15 model. ϕ and σ are constant across Gaussians (.067 and 5, respectively) while μ's are randomly distributed.

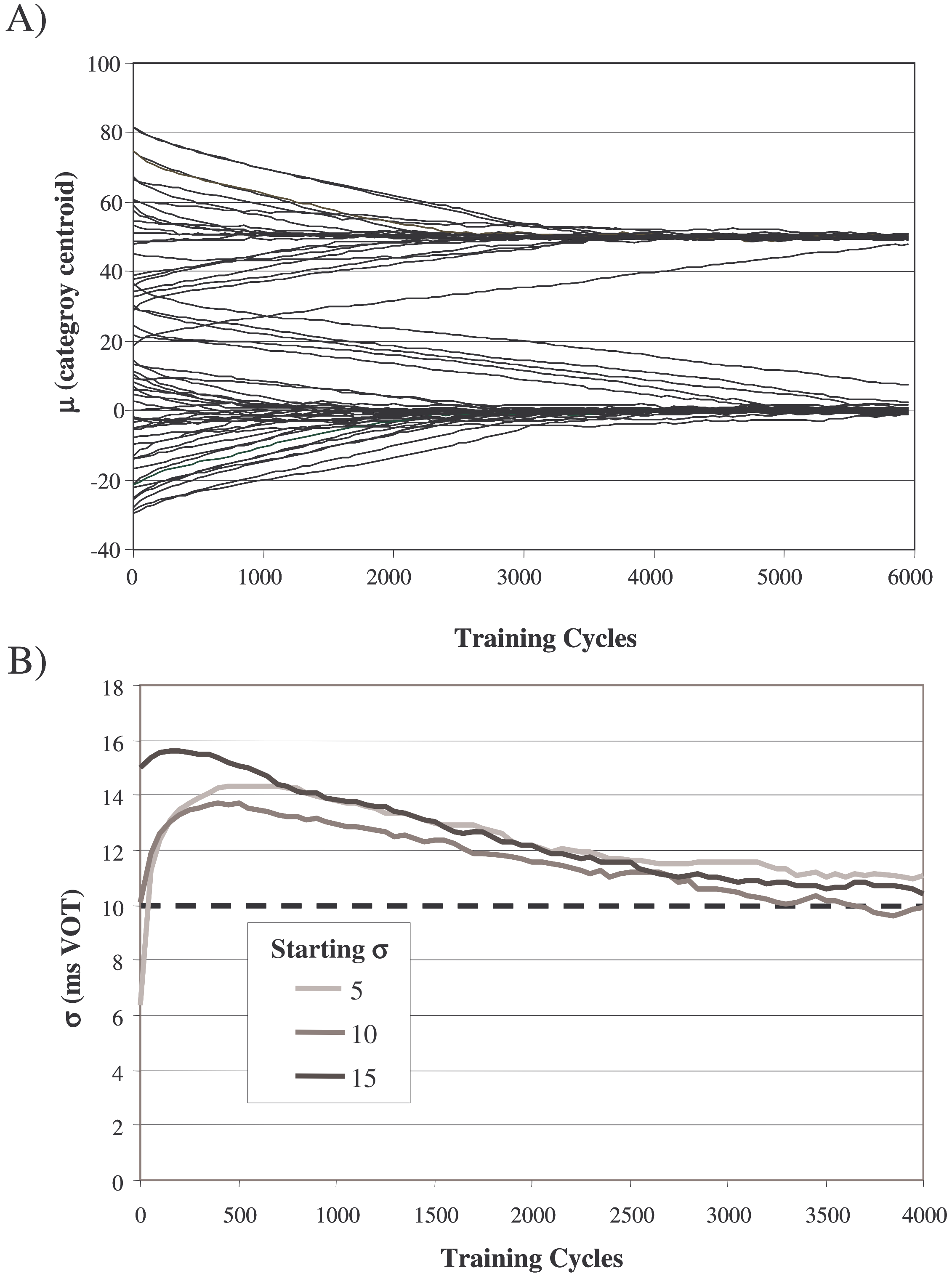


Figure 3: Results of a dynamical analysis of learning in the Mixture of Gaussians framework. Panel A) Trajectories of μ's for individual models. Panel B) Mean σ as a function of training cycles and starting σ. The dashed line represents the σ of the dataset used in training.

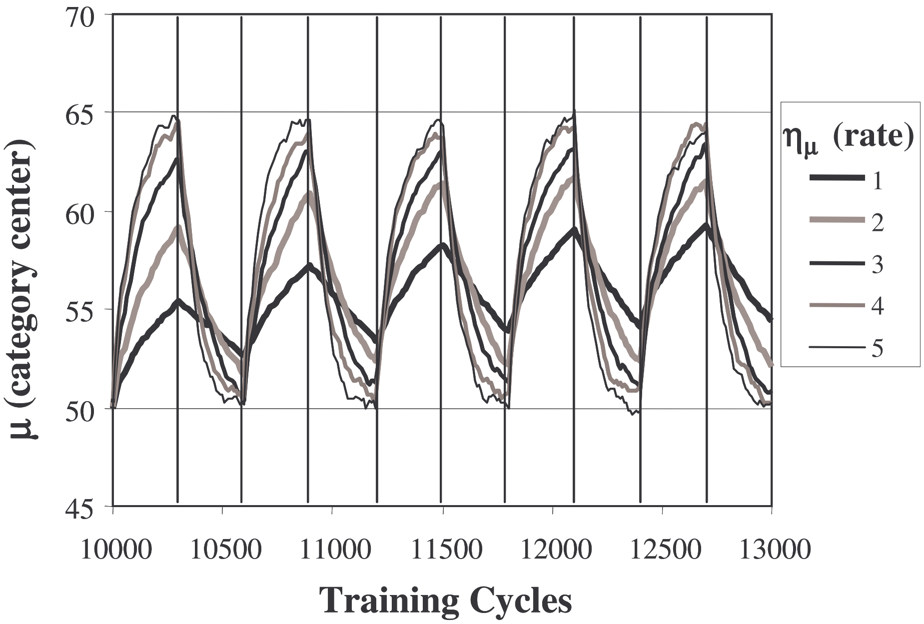


Figure 4: μ as a function of time and ημ for the voiceless category in plasticity simulations. After 10,000 cycles of training on categories centered at [0 and 50], the training set alternated every 300 epochs (vertical lines) between [0 50] and [10 65]. Note that only the voiceless category is shown for clarity of presentation.

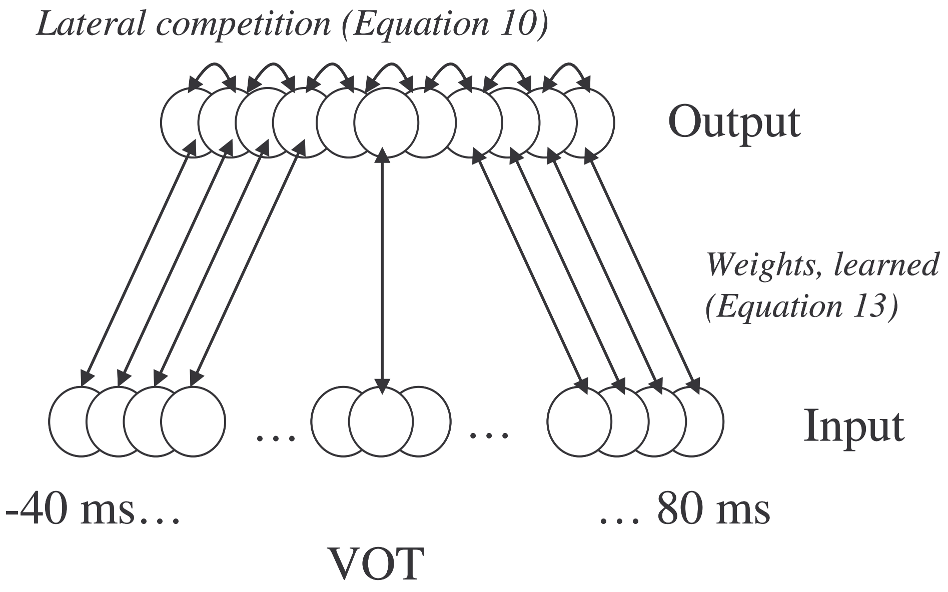


Figure 5: The Hebbian Normalized Recurrence network used for speech categorization.

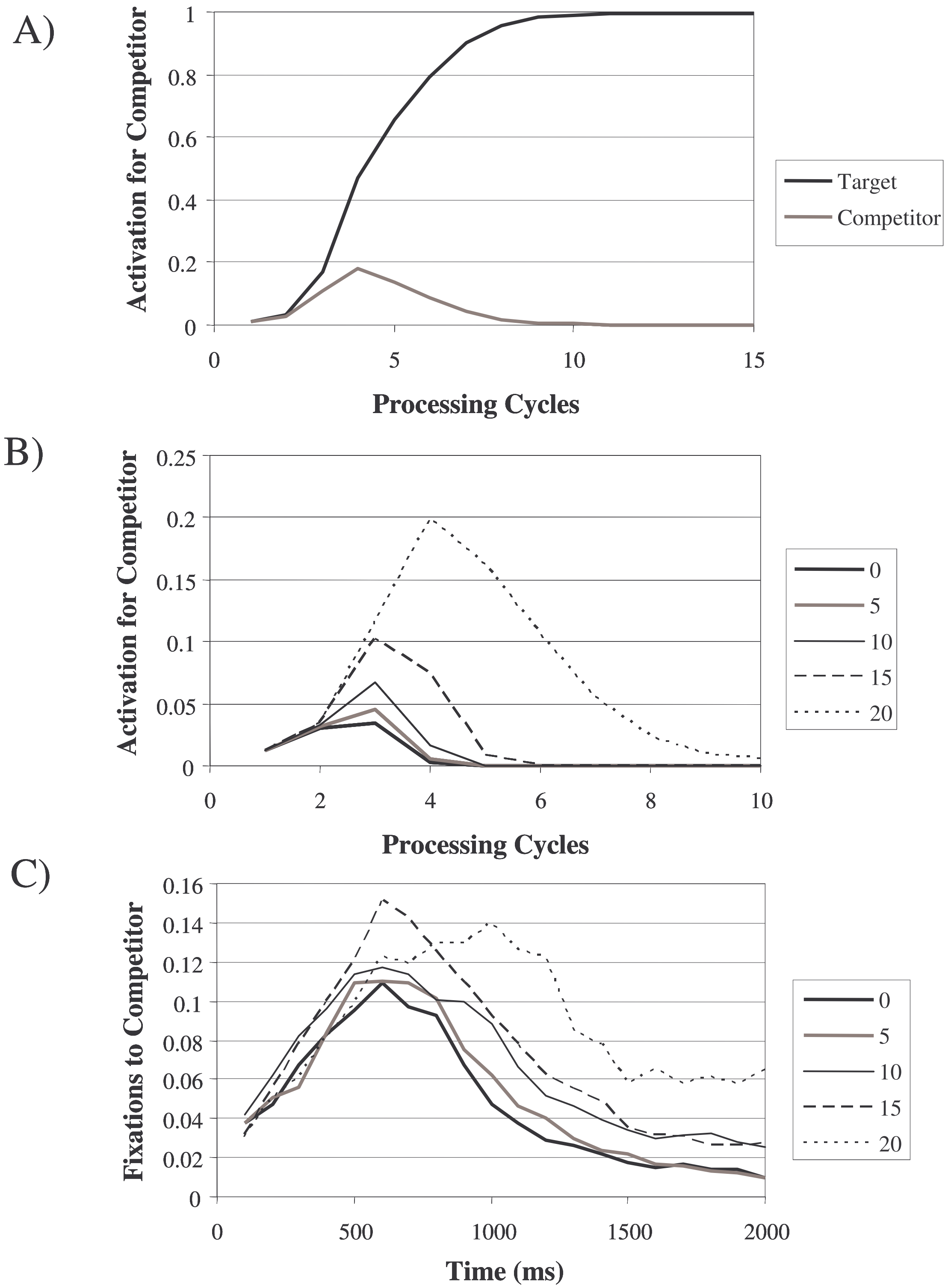


Figure 6: Temporal dynamics of processing in the Hebbian Normalized Recurrence network for speech categorization. A) Activation for targets and competitors for a single VOT as a function of time (processing cycles). B) Activation of voiceless competitor as a function of time and VOT. C) Probability of fixating the competitor a function of time and VOT for McMurray et al (2002).

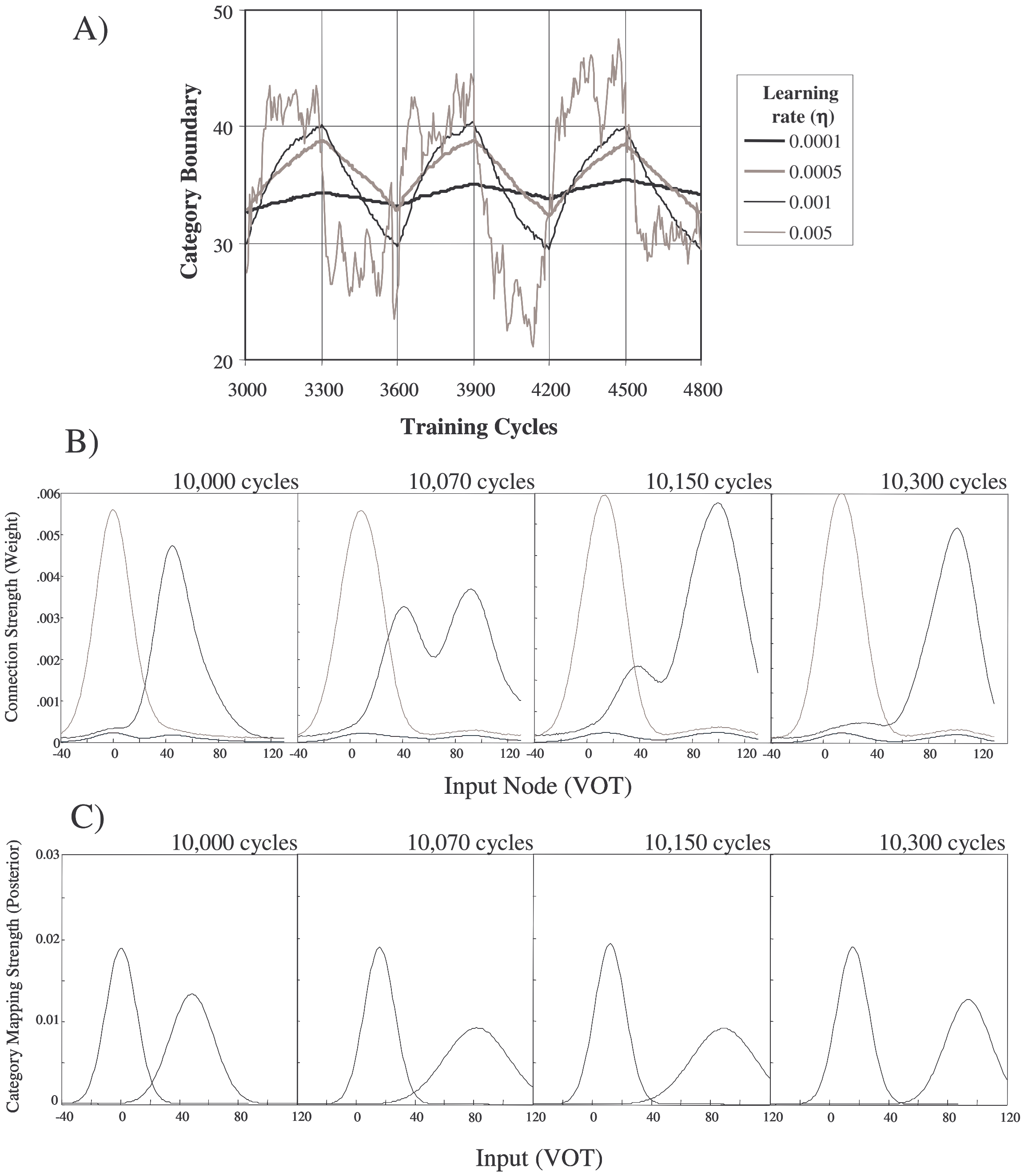


Figure 7: Plasticity simulations for Hebbian Normalized Recurrence. A) After 10,000 cycles of training on categories centered at [0 and 50], the training set alternated every 300 epochs (vertical lines) between [0 50] and [10 65]. Shown is category boundary as a function of time and η. B) HNR categories during a single category-shift from [0 50] to [15 95]. C) MOG categories during a single category-shift from [0 50] to [15 95].

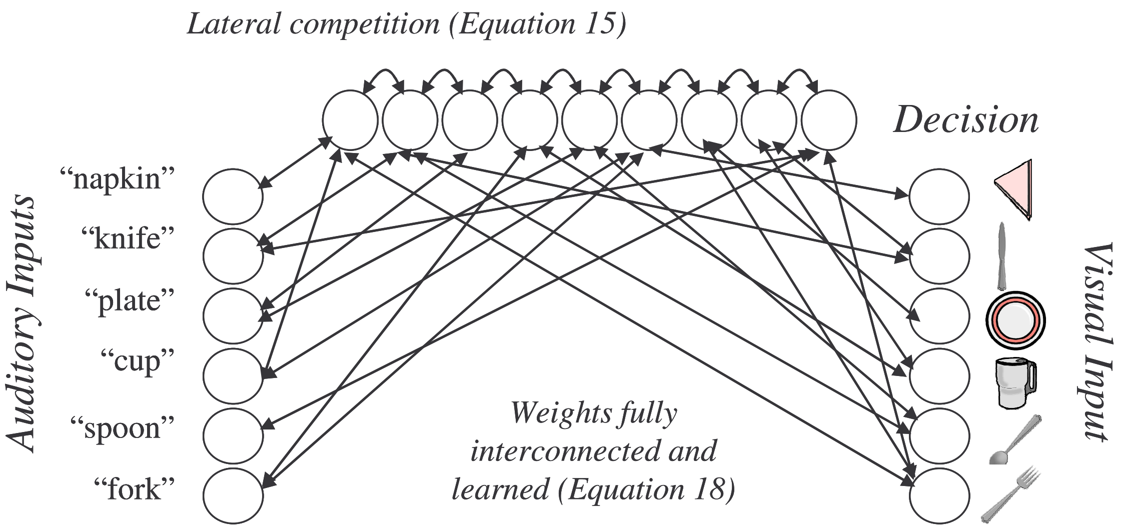
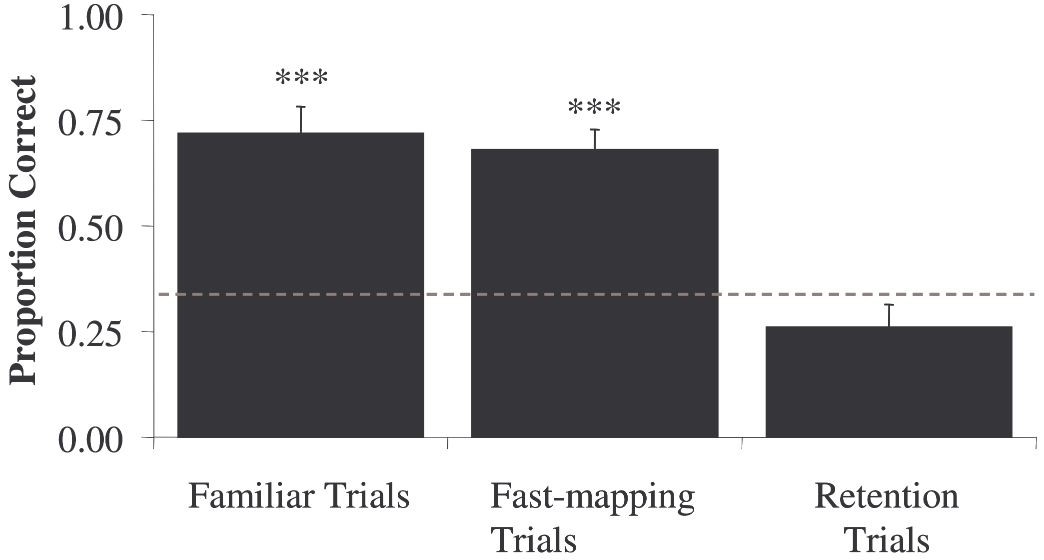


Figure 8: The Hebbian Normalized Recurrence network used for word learning.

A) Empirical Results



B) Simulation Results

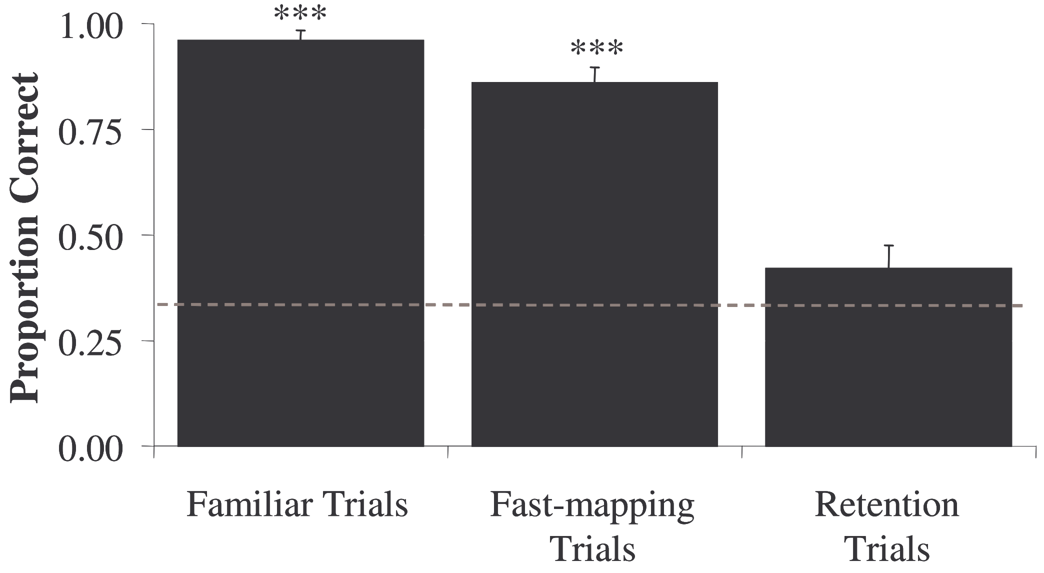
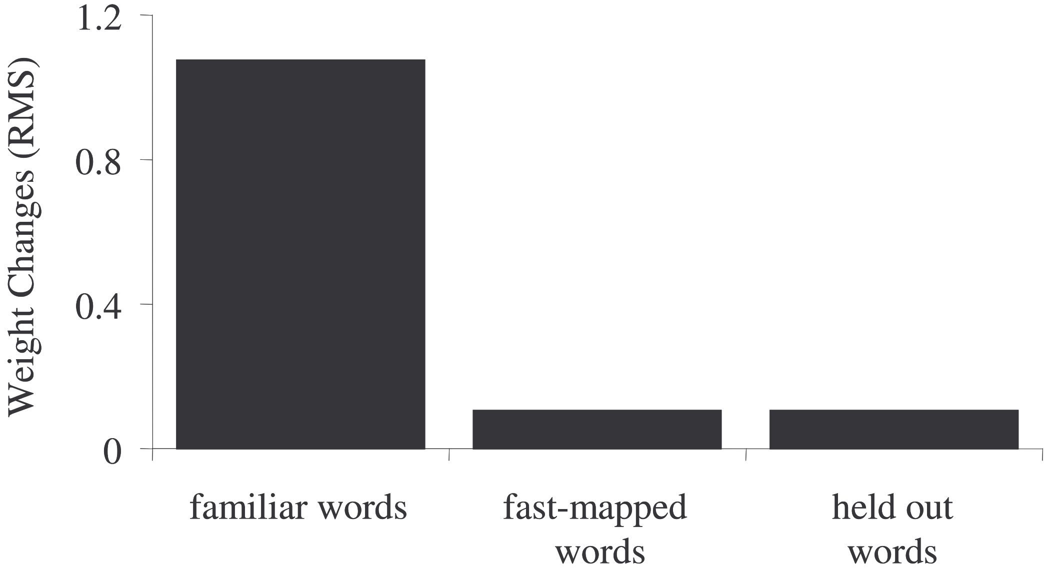


Figure 9: A comparison of empirical results from Horst & Samuelson (submitted) (A), and model performance (B). Error bars represent SEM, and *'s indicate significant deviation from chance (33%, the dashed line).

A) Learning during training phase



B) Learning during testing phase (fast-mapping)

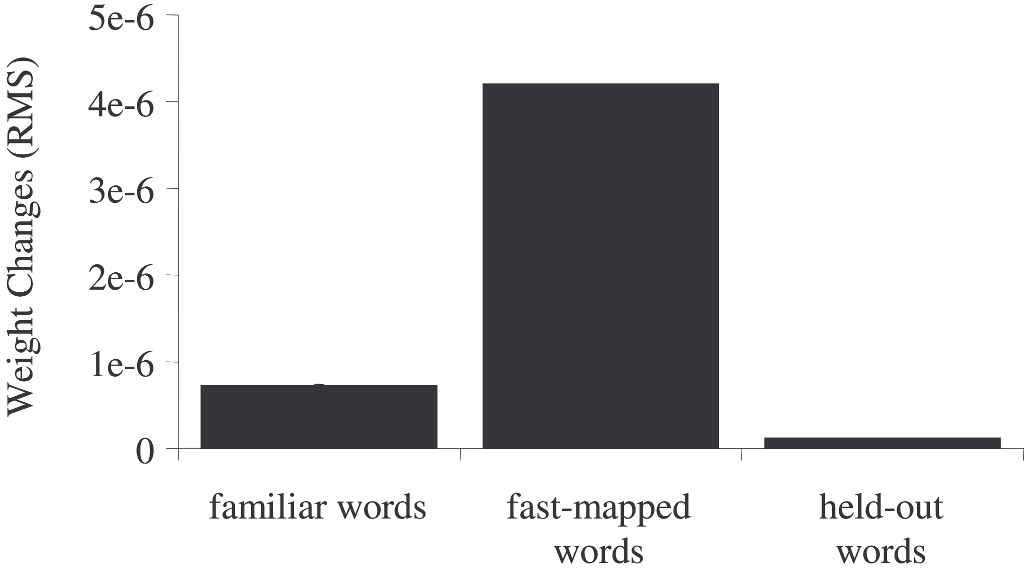
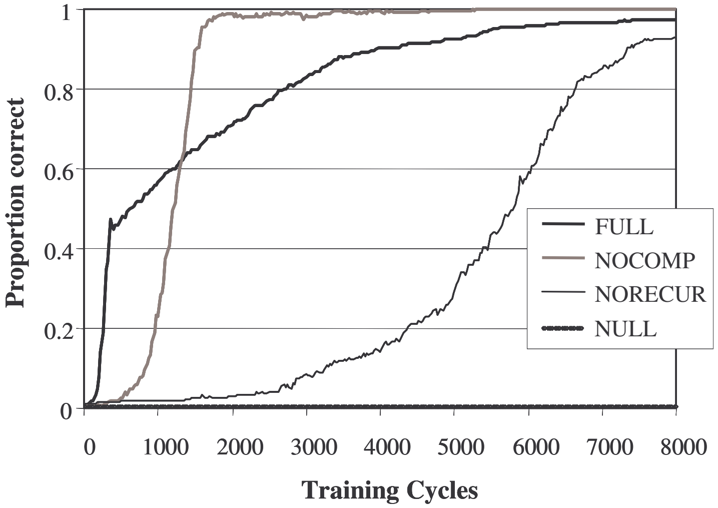
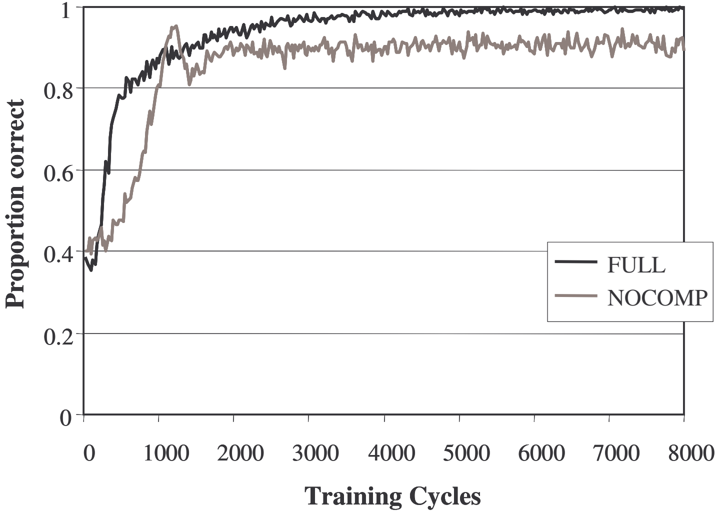


Figure 10: RMS weight change between the initialization of the model and the end of familiarization (Panel A), and between the end of familiarization and the end of the testing phase (Panel B). Note the change in Y-axis scale in Panel B.

A) Competence



B) Performance



C) Processing Cycles

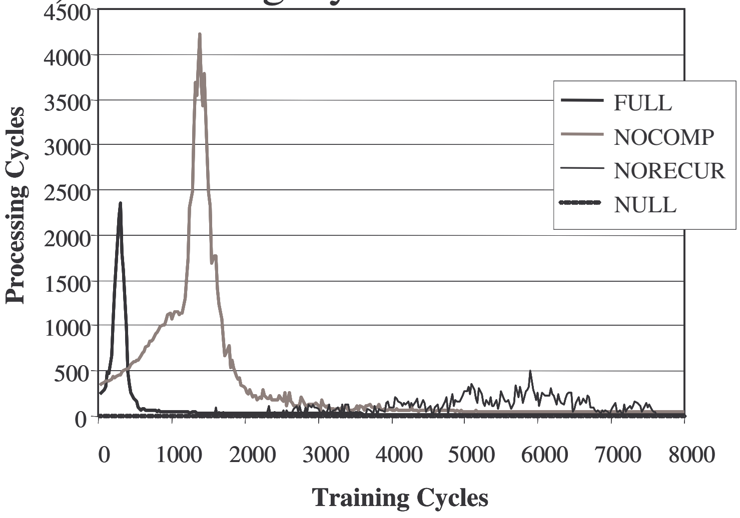


Figure 11: Performance of the word learning model as a function of training and model components. A) Performance measured by connections to the decision units. B) Performance measured in the visual array. C) Processing cycles (analogous to reaction time) as a function of learning.

Boxes

Box 1: Formulas for the Mixture of Gaussians model of speech categorization.

Note that while VOT is used in place of X, it is only used for the sake of concreteness of description—this model is intended as a general model for distributions along any single speech dimension.

$$G_i(vot) = \frac{\phi_i^{\ 2}}{\sum_{j=1}^{k}\phi_j^{\ 2}}\frac{1}{\sqrt{2\pi\sigma_i^{\ 2}}}\exp\left(\frac{(vot-u_i)^2}{2\sigma_i^{\ 2}}\right)$$

(4) Probability of a given VOT for model i. K represents the number of total models in the mixture, $\phi_i^2 / \Sigma\phi^2$represents the likelihood of model i, normalized against the likelihood of the other models. μ_i is the mean VOT (category center) of model i, σ_i is the width.

$$M(vot) = \sum_{i=1}^{k} G_i(vot)$$

(5) Probability of a given VOT for all models (i). The likelihood of any given VOT is the sum of the likelihoods of each model.

$$\Delta\phi_i = \eta\cdot\left(\frac{G_i(vot)}{M(vot)}\cdot\frac{2}{\phi_i} - \frac{2\phi_i}{\sum_{j=1}^{k}\phi_j^2}\right)$$

(6) Learning rule for adjusting ϕ_i (from which the prior likelihood of model i is computed) for a given VOT. Note that this rule only operates on a single model (the one with the highest likelihood) to create a winner-take-all form of competition. η represents a learning rate parameter.

$$\Delta\mu_i = \eta\cdot\left(\frac{G_i(vot)}{M(vot)}\right)\frac{(vot-u_i)}{\sigma_i^2}$$

(7) Learning rule for adjusting μ_i (the center of the category) given a single VOT. η represents a learning rate parameter.

$$\Delta\sigma_i = \eta\cdot\left(\frac{G_i(vot)}{M(vot)}\right)\left(\sigma_i^{-3}(vot-u_i)^2 - \sigma_i^{\ -1}\right)$$

(8) Learning rule for adjusting σ_i (the center of the category) given a single VOT. η represents a learning rate parameter.

Box 2: Formulas for Hebbian Normalized Recurrence.

$$o_x = o_x + \sum_{z \in i} w_{xz} \cdot i_z$$

(9)Activation in an output node (o_x) is the sum of activation from the input (i_y) nodes multiplied by the strength of the connection between them (w_{xy})

$$o_x = o_x^2 \Big/ \sum_{z \in o} o_z^2$$

(10) Output activation is squared and normalized to sum to 1 as a form of nonlinear competition. If this step is repeated many times (for any pattern of activation), this formula always yields a single output node active at 1, and all of the others at 0 (unless two nodes are perfectly tied). Thus, its limit behavior approximates Rumelhart and Zipser's (1986) competitive learning, while its moment-by-moment behavior is continuous. Also note that this is provably equivalent to lateral inhibition in which a given node inhibits all other nodes as a function of the proportion of total activation it holds (see McMurray, Horst & Samuelson, in preparation) and similar to Heeger's (1993) findings of normalization in population codes in the visual system.

$$i_y = i_y + i_y \cdot \sum_{z \in o} o_z w_{zy}$$

(11) Output activation spreads back to input layer (along same weights). Multiplying feedback by activation in the input layer prevents hallucination—feedback cannot affect an input node unless it is already active.

$$i_y = i_y \Big/ \sum_{z \in i} i_z$$

(12) Input layer is normalized to sum to 1.

$$\Delta w_{xy} = \eta \cdot \left(i_y o_x - w_{yx} \right)$$

(13) Weights are adjusted at each cycle through the model by a modified Hebbian rule. η represents the learning rate and is very low (~1e-6).

Box 3: Formulas for Hebbian Normalized Recurrence Word Learning Simulation.

$$d_x = d_x + \sum_{z\in a} w_{xz} \cdot a_z + \sum_{z\in v} u_{xz} \cdot v_z$$

(14) Decision node activation (d_x) (similar to output units) is the sum of the weighted (w_{xz})activation of the auditory input (a_z) nodes plus the weighted (u_{xz}) activation of the visual input nodes (v_z),

$$d_x = d_x^2 \Big/ \sum_{z\in d} d_z^2$$

(15) Decision activation is squared and normalized to sum to 1 as a form of nonlinear competition.

$$a_y = a_y + a_y \cdot \sum_{z\in d} d_z w_{zy}$$

$$v_y = v_y + v_y \cdot \sum_{z\in d} d_z u_{zy}$$

(16) Output activation spreads back to the two input layers (along same weights). Note that the only way that auditory and visual input arrays can affect each other is via this feedback mechanism (through the decision layer)—there are no direct connections.

$$a_y = a_y \Big/ \sum_{z\in a} a_z \qquad v_y = v_y \Big/ \sum_{z\in a} v_z$$

(17) Both input layers are independently normalized to sum to 1.

$$\Delta w_{xy} = \eta \cdot \left(a_y d_x - a_y w_{yx} - d_x w_{yx}\right)$$

$$\Delta u_{xy} = \eta \cdot \left(v_y d_x - v_y u_{yx} - d_x u_{yx}\right)$$

(18) Weights are adjusted at each cycle through the model by a modified Hebbian rule. η represents the learning rate. Note that the weight decay function (adapted from Adaptive Resonance Theory: Grossberg, ___) insures that weights for which either the decision or input (a or v) units are inactive (0) do not decay. This preserves plasticity so that these weights may be involved in learning other words.

Mechanisms for the Grand Theory

Heidi Kloos

University of Cincinnati

Guy C. Van Orden

Arizona State University & University of Cincinnati

Abstract

Cognition and development are exquisitely context sensitive, *context effects all the way down*. Conventional solutions to this fact perpetuate stalemates between models that address different context effects such as task effects. One path around the stalemates is to take context effects seriously. A new grand theory may originate in these facts. The pay-off is a perspective on modeling in which connectionist and dynamical systems models are about the same kind of thing—emergent control of human behavior.

Mechanisms for the Grand Theory

Modeling children's performance is essential to the discourse about cognitive development. Models are attempts to characterize the mechanisms of cognitive performance and cognitive development. In this chapter we discuss the kinds of mechanisms that models – or performance in general – could reveal. In particular, we distinguish between two views of mechanism that avail themselves to contemporary scientists. One view of mechanism sees cognition as reducible to a cognitive architecture of separate mental functions. The other view of mechanism represents cognition as emergent, highly dependent upon the fine details of contexts in which behavior occurs. Both views of mechanism provide causal explanations of cognition and development.

We distinguish between these views of mechanism to evaluate whether connectionist and dynamic-systems models may converge on a grand theory. We will argue that the answer depends on the kind of mechanism that they attempt to unravel. Models aimed at uncovering cognitive functions of behavior – whether connectionist, dynamic-systems, or any other model – are likely to end at loggerheads, opposing each other, with none able to make clear headway over the others. If on the other hand, the models aim to describe the array of constraints that produce specific behaviors, connectionist and dynamic-systems models can compliment each other towards a grand theory of development.

The chapter organization is as follows. In Part 1, we address why models aimed at reducing behavior to mental functions cannot make clear headway. The argument is that reduction of behavior requires human performance to be relatively context free. Yet, as we show on the example of balance task performance, human performance is highly context dependent, even in the sterile laboratory world of balance experiments. In Part 2, we elaborate on what such context dependence could mean. While not conclusive on its own, strong context dependence is consistent with the idea that cognition is softly assembled to suit the immediate task environment. Soft assembly offers a plausible alternative to hard assembled cognitive functions – functions that exist prior to and independently of the task context. We review more pointed evidence for soft assembly and discuss why models that take soft assembly seriously – connectionist or dynamic systems – anticipate the grand theory.

Part 1: Where Models Converge in Stalemate

At the center of the argument is strong context dependency in human performance. We develop this argument around well-studied examples from developmental psychology, in particular children's performance on balance tasks. However, our points pertain to cognitive modeling more generally.

Balance Performance and its Models

Picture a child in a balance beam experiment. The balance beam straddles a fulcrum and has pegs along its surface on which to set weights. The experimenter sets some weights and asks the child to predict the behavior of the beam. Will it stay balanced, will the right side tip, or will the left side tip? To perform correctly, in every case, the child must take into account the number of weights on each side of the bar, the distance of each weight from the fulcrum, and their product.

Different children give different kinds of answers to balance beam problems of this sort, but four kinds of answers, found across trials and experiments, are reliable and systematic (Siegler, 1981). (I) One class of answer appears to reflect an exclusive focus on the amount of weight on each side of the fulcrum while ignoring distance from the fulcrum. Children who answer this way consistently expect the side with the most weights to tip or that the balance scale should stay level when there are equal numbers of weights on each side. (II) A second class of answer appears to take into account distance from the fulcrum, but only when weights are equally distributed on both sides. Children who produce this kind of answer predict that the heavier side will tip. But when both sides have the same numbers of weights, they predict that the side with weights furthest from the fulcrum will tip. (III) A third class of answer takes into account both weight and distance from fulcrum, but does not suggest a systematic integration of the two dimensions. Children who produce this class of answers correctly predict the effect of weight when the distances are the same, and they correctly predict the effect of distance from fulcrum when the number of weights is equal. However, when both weight and distance from fulcrum differ, then the children simply guess. (IV) A final, fourth class of answers appears to take into account both relevant dimensions (number of weights and distance from the fulcrum) and integrate dimensions appropriately when the two dimensions are pitted against each other.

A variety of models have been proposed to explain the four kinds of answers and the transition from one type of answer to the next most sophisticated. Production-rule models suggest that children's performance derives from rule-like algorithms or strategies that can change rather suddenly as a result of experience (e.g., Klahr &

Siegler, 1978; Langley, 1987; Sage and Langley, 1983; Schmidt & Ling, 1996). Connectionist models suggest that children's performances are not a function of explicit rules but rather a gradual adaptation to statistical relations between balance appearance and response (e.g., McClelland, 1989; 1995; Shultz, Mareschal, & Schmidt, 1994; Shultz, Schmidt, Buckingham, & Mareschal, 1995). Dynamical systems models capture developmental changes as sudden jumps in the cusp catastrophe—a rule-like response looses stability and changes suddenly in a phase transition or bifurcation to a new rule (e.g., van Rijn, van Someren, & van der Maas, 2002).

A condensed overview of these models indicates some disagreement among them. Each proposes a different causal mechanism that could underlie balance performance. One solution to the disagreement might be a closer look at the fit between model output and human performance (van Rijn et al., 2002). Indeed, production-rule models have difficulty capturing the transition between each rule, and connectionist models do not do well with rule-like output unless it is a product of statistical relations in the training set. One could find a way to calculate overall fit across models in a single score and conclude that the model with the best score must be using the correct 'mechanism' – the mechanism that best mimics the underlying cognitive architecture.

We suggest that such a solution is premature because it does not adequately consider how context figures into the disagreement. The models that presently compete all emphasize the axiom that behavior can be the same across different laboratory contexts. Psychological constructs are only motivated by "behavioral consistency over varying contexts" (Embretson, 2006, p. 50; Cronbach & Meel, 1955).

Taken at face value, this axiom rules out context sensitive behavior as a topic of study, as it would not adequately motivate psychological constructs. Put another way, only context-free behavioral effects adequately pick out underlying mechanisms. Thus the models that would compete for best overall fit, at present, would all assume that behavior stands outside of context. But as we argue next task context effects also distinguish between the models.

Balance Performance is Context Sensitive

Adults, children, even toddlers move through uncountable varieties of contexts with fluid ease—from outdoor play on the slippery slide to indoor play on the paintball court, from being a peer to being a child or parent, and from quietly drawing to arguments about bed time. Even in the artificial and somewhat sterile environments of cognitive laboratories, context effects are paramount.

Take again the example of balance studies. The four types of balance-scale behaviors described above – Rules I, II, III, and IV – are well known and often cited. But it would be grossly misleading to suggest that this typology captures children’s full range of balance behaviors. In fact, it captures only a limited set of options allowed within the particular task context (Hardiman, Pollatsek, & Well, 1986). Limits come from a variety of methodological choices that are rationalized in terms of good experimental design. In this section, we discuss some of these choices in more detail.

One methodological choice is to hold back feedback during children’s test performance. A child is presented with a balance-scale scenario, in each measurement trial, and after predicting the behavior of the balance scale they progress immediately to the next trial. But feedback during testing changes the likelihood that a child’s performance will adhere to one of the four rules, especially when feedback is tailored to a child’s beliefs (Hardiman et al., 1986; Kliman, 1987). Siegler and Chen (1998) for example, carefully constrained the feedback provided to preschoolers to stop them from forming the rule *‘greater distance goes down’* instead of Rule I: *greater weight goes down* or Rule (II): *if weight is equal, greater distance goes down*.

Another choice is to provide children with only three answer options: ‘left side down’, ‘right side down’, or ‘balance’. If children are provided with more flexibile options, as when they are allowed to adjust the number of weights and their position on the beam, children’s performance can follow an entirely different rule called the *addition rule*. Under these circumstances, children integrate both the number of weights and their distance from the fulcrum according to an additive algorithm (Normandeau, Larivee, Roulin, & Longeot, 1989; Wilkening & Anderson, 1982). That is to say, they do not focus on one dimension only (Rule I – III) or integrate them according to the correct multiplication algorithm (Rule IV).

Yet another choice of method is the rather limited range of possible scenarios presented to participants. Only four pegs are fitted on each side of the fulcrum and only four weights can be placed on the pegs. This yields rather small discrepancies between the left and right arms of the balance beam. As it turns out, the magnitude of discrepancy between left and right side matters – an effect known as the *torque effect.* Children’s performance is better with larger differences in torque (Ferretti & Butterfield, 1986; Ferretti, Butterfield, Cah, & Kerkman, 1985; Jansen & van der Maas, 1997).

Another design choice is to balance out the number of trials of each type. Number-of-weights versus distance-from-fulcrum are pitted against each other in half of the trials while either one, or both, dimensions are held constant in the other half of the trials. Tasks that allow a greater and unbalanced variety of configurations of weights do not elicit performance of the four rules (Kliman, 1987). Children no longer form rules on the basis of differences in weights or distance from the fulcrum; they form rules about mathematical relations, specific to their narrow experience. For example, a child will abstract from examples like "*one* weight on the *ninth* peg balances *nine* weights on the *first* peg" to become "*one* weight on the n^{th} peg balance *n* weights on the *first* peg." And yet same child will generalize this belief no further, not even to accommodate two weights or more weights on the *n*th peg.

Finally, a fifth methodological choice limits proprioceptive information to information about weight (children are often encouraged to lift one or more weights) but no proprioceptive information is provided about how distance of weights affects balancing. The lack of proprioceptive information about how distance affects balance selectively increases the difficulty for children to correctly take into account distance (Hardiman et al., 1986; see also Gibbs, 2006). In particular, exclusive proprioception of weight could bias children's performance toward Rule I (*greater weight goes down*) and Rule II (*if weight is the same, greater distance goes down*), for instance, but against a rule based on distance alone. Indeed, proprioceptive information about distance leads children to perform as though the rule is *greater distance goes down* but completely ignore differences in weight (Karmiloff-Smith & Inhelder, 1975). And children will sustain this rule for some time despite contradictory evidence, suggesting that this "distance rule" is at least as stable as Rules I-IV.

The present list of context effects does not exhaust the list of task effects, which only seem to grow in number with each innovation in task context. Take for example children's performance in tasks in which children must integrate information about time, speed and velocity (Wilkening, 1981), task conditions no more complicated than balance tasks. In one task context, children estimated the distance an object traveled, given its velocity and travel time. In a second task context, children estimated the time an object traveled, given its velocity and travel distance. And in the third task context, children estimated an object's velocity given the time and distances traveled. Each of these three contexts is a permutation of the relations among distance, velocity and travel time. And each permutation asks the child for a prediction of one factor based on the other two factors as antecedents. The finding important for our purposes was that each task yielded a qualitatively different outcome. For example, 5-year-olds integrate speed and time multiplicatively to estimate the distance traveled, but they integrate speed and distance additively to estimate the time traveled, and they used distance exclusively to estimate speed—three qualitatively different outcomes.

Strong context dependence is of course not limited to balance-scale tasks or distance-time-speed tasks. Quite the opposite is the case. Context dependent performance is the rule not the exception in developmental psychology (e.g., Gigerenzer & Richter, 1990; Lawton, 1993). In fact, one could argue that the most salient outcome of American-style post-Piagetian research was to demonstrate extreme context sensitivity. Changes in task design yielded more and more sophisticated competence in younger children, making contextual support an important causal variable (cf., Keen, 2003). Nonetheless, despite omnipresent context effects, the effect of context is rarely studied in its own right.

Explaining Away Context Effects

One way to treat task effects is to assume that they are a simple add-on to other factors, notably to cognitive functions. For instance, if children simply performed better in one balance task than another then, all other things equal, task effects are superposed on other effects. In this case one could legitimately partition out task context as an independent source of variation in data. In a model, task context would then become a fixed parameter. But task and context effects rarely combine so straightforwardly, in balance performance or any other developmental phenomenon. As our review illustrates, children tested in a balance task do not simply perform better or worse. They exhibit different qualities of performance in different task contexts.

Another way to treat task effects is to account for children's performance in all of the different contexts. In fact the dynamical systems model of van Rijn et al. (2002) simulates Rules I-IV as well as the torque effect described by Ferretti et al. 1985. And McClelland's (1989) connectionist model can account for a difference in salience between weight and distance. Although these attempts do not capture the addition rules proposed by Wilkening and Anderson (1982) or idiosyncratic rules discussed by Kliman (1987), one could still imagine the possibility of a model that could take into account children's behavior across all contexts. However current attempts are far short of providing a comprehensive account.

Yet, another way to treat task-contexts effects is to equate different effects with different mental functions. This solution may appear clear-cut when tasks differ conceptually. For example, tasks that require a physical action such as balancing a rod can appear conceptually different from tasks that require a verbal judgment of balance (cf. Kirst, Fieberg, & Wilkening, 1993; Levin, Siegler, & Druyan, 1990). A physical action might shed light on implicit knowledge, while verbal action might shed light on explicit knowledge—two distinct forms of representation. The difficulty of this approach starts when tasks are conceptually alike. For example, it is not clear how a balance task with a small number of weights and pegs differs conceptually from balance task with a large number of weights and pegs. Both should entail the same balance knowledge, making guesses about different mental functions unpredictable and circular.

Finally, the most widely practiced response to task effects is to argue about which context is more transparent to mental functions, or equivalently which task context produces more “pure” data than another. In the example of the balance scale performance, Siegler (1981) defends the highly structured and methodologically precise task context of his assessment procedure, and Wilkening and Anderson (1982) justifies the legitimate expansions of task contexts to explore the task space. But in effect, there is simply no empirical basis on which to decide which task context is best. All task effects refer equally to changes in outcome measures of performance. Distinctions between task effects are supported only by intuitions about competence based on convention or esthetics, but not evidence. Consider however that task effects might not simply reveal superficial balance perspectives, but rather reveal that task and context are constitutive of children’s performance.

Arguments about the purity of data have not fared well in other domains. Conventional studies of adult cognition have virtually run to stalemate on the question of which task's data best reveal the architecture of cognition. Details of apparent stalemates have been described for perception (Uttal, 1990, 1997), reading (Van Orden & Kloos, 2005; Van Orden, Pennington & Stone, 2001), and memory processes (Watkins, 1990; Weldon, 1999). Similarly, functional neuroimaging of adult cognition shows signs of running to stalemate: subtle changes in task context cause cognitive functions to be in different parts of the brain (cf., Cabeza & Nyberg, 2000).

In sum, we have shown how context effects permeate our laboratory pictures of children's knowledge. Such task effects demonstrate that a child's performance is an interaction of the child's knowledge and the specific task constraints within which they act. The solutions discussed so far hold onto the idea that behavioral phenomena can be reduced to specific mental functions or processes. But these solutions have also lead to the stalemates about which task contexts separate context effects from cognitive functions most successfully. In the next section, we discuss a more radical solution to context sensitivity – one that conceives context effects as constitutive of a child's performance.

Part 2: Taking Context Effects Seriously

To this point we have shown why present solutions to context effects are unsatisfying or in a hopeless stalemate. In the remainder of the chapter, we will describe a way to circumvent stalemates about how to isolate cognitive functions from contexts effects. The argument rests on the distinction between hard-assembled and soft-assembled cognition (Turvey & Carello, 1981). So far we have discussed hard-assembled cognition – functions that exist off-line independently of a task context. Soft-assembled cognition on the other hand emerges out of collective action of the brain and body of the child in the immediate task context. Soft assembled cognition does not exist independently of the task context in which a child acts.

In Part 2, we will apply a lesson from physics to cognition. Laughlin (2005) describes examples in physics where theoretical debates ran to stalemate. Just as in the present case, those debates came to hinge on arguments about methodological purity that new data could not decide. Laughlin describes these debates as symptoms of applying an inappropriate conceptual framework to problems at hand. Physics experimenters treated soft assembled mechanisms of collective action as though they could be reduced to hard assembled mechanisms of component causes. Despite applying the wrong conceptual framework, experiments continued to yield interpretable outcomes. In the wake of each new experiment that “worked,” they continued to believe that the correct experiments could break through the falsity of other camps’ beliefs. But the other camps’ “successful experiments” trapped them in the same illusion. New data only perpetuated arguments indefinitely, hence stalemate.

The same may be true for cognition. Widespread context sensitivity in behavior could be a feature of soft-assembled collective action of the brain and body. If this is the case, then context effects become the basis for understanding behavior and need not to be explained away. In the ensuing sections, we develop this argument further. We illustrate how context effects permeate the brain and body, and we describe evidence for soft-assembly in adult cognition. Finally, we return to the issue of modeling and discuss whether connectionist and dynamical systems models compliment each other in a grand theory.

Cognition is ‘Context Effects All the Way Down’

The interaction between task context and cognition may permeate the brain and body below the level of a child’s judgments in a task. That is to say, context effects discussed in previous sections might just be the tip of an iceberg, and the iceberg itself might be context sensitivity all the way down. Below the iceberg’s tip, each muscle flex and each pattern of rhythmic cortical firing may create a context for every other muscle flex and neural firing (cf. Belen’kii, Gurfindel’ & Pal’tsev, 1967; Marsden, Merton & Morton, 1983).

Take for example the coordination of speech after an unexpected pull to a person’s jaw (Kelso, Tuller, Vatikiotis-Bateson & Fowler, 1984). Articulation compensates with movements in the upper and lower lips to preserve the flow of speech such that a listener could not distinguish between perturbed and unperturbed speech. The lip movements entail cortical interactions, the complex fluid matrix of neuromuscular interactions in the lips, modulation of the force of breath and the pace of respiration, and all else that makes up speech. Most important, the fluid compensation precisely expresses contextual constraints. Context in this case equals the unfolding of a spoken word as co-articulated speech. This word context surrounds the point in time at which the experimenter perturbs the jaw and precisely shapes compensation all the way down to insure intelligibility.

Related ideas have been advanced in the behavioral neuroscience of perception and action. In laboratory experiments, brain and body immediately reconfigure to accommodate changes in the task environment. For example, a slight change in the task (e.g., increasing the frequency of oscillation) leads to a new pattern in behavior in a virtually instantaneous phase transition across brain and body (Kelso, 1995). To accomplish this, global patterns of active constraint must connect all the excitable media of the body. Such patterns maintain identifiable but context-sensitive profiles to anticipate which actions are appropriate and possible, to anticipate how to act to satisfy the context of action. They do so to keep apace of a perpetually changing environment. In effect, brain and body combine constraints to poise each other, perpetually, ready for action.

Simultaneous change of brain and body may explain otherwise paradoxical ultra-fast cognition. For example, the time it takes to know that a picture was flashed is sufficient to know whether it was a bird or a car (Grill-Spector & Kanwisher, 2006). In other words, perceivers require no more time “for object categorization than for object detection” (p. 157). In this and other instances of ultra-fast cognition, the mind and body appear pre-prepared to act immediately in one “feedforward pass” (Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001, p. 6). To do so they must somehow capitalize on the arbitrary constraints of a task context. Cognition seems to use the fine points of contextual detail to pre-constrain possible actions. Preparative cognition stays in the loop, so to speak, by creating a perpetually constrained state of readiness for action.

The picture of context effects we have drawn expands on a famous description of development by connectionists—namely, that development is ‘*interactions all the way down*’ (Elman, Bates, Johnson, Karmiloff-Smith et al., 1996). Elman et al. use a newly hatched duck to illustrate how preparative constraints set up the potential for imprinting that is subsequently determined by the interaction with the environment. Beyond nature and nurture, a duck or a child is pre-prepared for developmental milestones and the milestones are realized in interactions with the environment. They reflect details of the environment as a consequence. Likewise cognition itself is preparative and only realized in interactions with the environment.

Hard Evidence for Soft Assembly

The preceding examples and arguments all push in the direction of soft-assembly, but they do not provide direct evidence. Direct support for soft assembly comes from evidence of fractal patterns in behavior – a pattern that repeats itself at every scale of observation. One way to see fractal structure in performance is to take repeated measurements and create a spectral plot that graphs amplitude of variation against its frequency. As it turns, the amplitude of variation gets larger with larger sample sizes. In other words, no measurement in a series of repeated measurements is independent of any other measurement – the context of measurement permeates the measurements themselves.

Fractal patterns appear to be universal in human behavior. They are documented widely in motor coordination studies and also in standard cognitive tasks (for reviews see Gilden, 2001; Riley & Turvey, 2002; Van Orden, Holden, & Turvey, 2003). In fact, as Machlup (1977) suggested decades ago, if you have not yet found a fractal pattern you have not taken enough data, “you have not waited long enough. You have not looked at low enough frequencies.” (p. 157). Universal fractal patterns imply that the processes of the brain and body are flexibly coupled. Such interactions among processes may be best thought of as interdependent, meaning that processes change each other’s dynamics as they interact. These interdependent processes allow context to permeate a system (for a fuller discussion see Van Orden et al., 2003).

Of course it is possible in theory to explain the fractal signature of behavior in terms of hard-assembled mechanisms. But the attempt leads quickly to absurd paradox. As it turns out, each measurement supplies its own fractal pattern, largely uncorrelated with fractal patterns generated by other measurements (Kello, Beltz, Holden, & Van Orden, 2006). For example, imagine taking two measurements per trial, such as the time until a response is made (e.g., stimulus to key press) and the duration of response (e.g., how long the key was held down). Both measurements from the same person in the same trial, trial after trial, and they both express the common grammar of fractal structure. However, the two fractal patters are uncorrelated. Thus, if one attempts to dissociate underlying mechanisms, each fractal pattern would require its own separate mechanism.

That might be all right if the complexity ended here, with two measurements, but it is trivial to up the ante on the number of measurements that will show independent fractal structure. Kello et al. (2006) illustrate this fact with repeated speech and the indefinite number of measurements one can take off the acoustic patterns of speech. Each measurement will yield its own fractal pattern, uncorrelated with the fractal patterns of the other measurements. It will appear as if each arbitrary measurement has an independent origin, and thus justifies its own hard assembled mental function. This absurd conclusion comes from thinking that one confronts hard assembled components in the first place.

The paradox disappears once we allow that cognition is soft assembled from component processes. Soft assembled behavior expresses contingent details of task demands and task contexts. We made that argument in previous sections. Contingent detail explains why we see so many mechanisms. The contingent details out of which task performances are constructed can include “sub-devices” for an indefinite number of subtasks—the ‘*respond to the signal*’ device versus the ‘*prepare the finger for the next trial*’ device, for instance. We escape circularity however because our analysis no longer targets this “functional level” of explanation.

Implications for Modeling

Our target for modeling is a level of emergent control, above the component details of task devices. A theory of emergent control makes progress so long as there actually are general principles of control to be discovered. In a way, this point about general principles simply takes a lesson from dynamical systems models. The cusp catastrophe for example is a very general account of control and qualitative change (Gilmore, 1981). And the search for empirical flags of the cusp catastrophe illustrates how one goes about establishing that cusp principles of control actually apply (van der Maas & Molenaar, 1992).

At the level of emergent control, task context effects are equated with task constraints in a model’s control parameters (Van Orden, Holden, Podgornik, & Aitchison, 1999). A control parameter is a ratio among the preparative constraints out of which behavior emerges. Values of the ratio will favor one or another outcome. For example, the ratio of weight to distance trials in a balance beam task could be conceived as a control parameter. The values of the ratio that favor distance trials yield different rules than the values of the ratio that favor weight trials.

Control parameters are most often associated with dynamical systems models but they are also discussed in the context of connectionist models (Farrar & Van Orden, 2001; Kello, 2003; Kello, Sibley, & Plaut, 2005; Rueckl, 2002). In a connectionist model, the ratio of weight versus distance trials is made explicit in the training regime and implicit in the weight matrix. In a dynamical systems model, this ratio could appear explicitly as a parameter in a system of equations. In both cases, the ratio controls competing outcomes that live on opposite sides of a critical value. The critical value is the point of equally distributed constraints, an imaginary point of no decision, a precisely balanced tug of war between equally compelling “rules” for instance.

This is a different view of modeling and control compared to previous schools of psychology. Models do not stand outside of time, in the sense of mental functions, except in the principles of their design. Most important, they do not capture phenomena outside of time. Models capture and make explicit the control of behavior emerging in time. Previous schools of psychology relegated control of behavior to relatively static loci in the environment (*Behaviorism*) or the organism (*Cognitivism*). The new grand theory will locate control in the perpetually changing interaction of child and environment.

Conclusions

Hard-assembled cognition makes context effects appear as a tribe of interlopers that must be conquered. From that perspective they are insidious, inserting themselves to undermine data that are otherwise equated with mental functions. Nonetheless cognitive performance is overly dependent on the particulars of context. A new grand theory of cognition and cognitive development must find its beginnings in this tribe.

Viable theory, methods, and models must embrace context sensitive phenomena. As for modeling, mimicking the control structure of human behavior captures the available causal basis of behavior. Therefore a grand theory of soft assembled cognition and development can embrace both dynamic systems and connectionist models.

We have argued that these attractive possibilities, perhaps inevitabilities, can be achieved by discarding the assumption of hard assembly. Fluid soft assembly of cognition is the essential human competence and performance is transparent to this competence. Competence as context sensitivity and performance as sensitivity to context are the two sides of the human coin.

References

Belen’kii, V., Gurfinkel, V., & Pal’tsev, Y. (1967). Elements of control of voluntary movements. *Biophysics*, *12*, 154-161.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*, 1-47.

Cronbach, L. J., & Meehl, Pl E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1999). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist, 61*, 50-55.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13*, 1-10.

Farrar, W. T., & Van Orden, G. C. (2001). Errors as multistable response options*.* *Nonlinear Dynamics, Psychology, and Life Sciences, 5,* 223-265.

Ferretti, R. P., Butterfield, E. C., Cah, A., & Kerkman D. (1985). The classification of children’s knowledge: Development on the balance-scale and inclined-plane tasks. *Journal of Experimental Child* Psychology, 39, 131-160

Ferretti, R. P., & Butterfield, D. C. 1986. Are children’s rule assessment classifications invariant across instances of problem types? *Child Development, 57,* 1419-1428.

Gibbs, R. W. (2006). *Embodiment and cognitive science.* New York: Cambridge University Press.

Gigerenzer, G., & Richter, H. R. (1990). Context effects and their interaction with development: Area judgments. *Cognitive Development, 5,* 235-264.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review, 108*, 33-56.

Gilmore, R. (1981). Catastrophe theory for scientists and engineers. New York: Dover Publications

Grill-Spector, K., & Kanwisher, N. (2006). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*, 152-160.

Hardiman, P. T., Pollatsek, A., & Well, A. D. (1986). Learning to understand the balance beam. *Cognition & Instruction, 3*, 63-86.

Jansen, B. R. J., & van der Maas H. L. J. (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review, 17,* 321-357

Karmiloff-Smith, A., & Inhelder, (1975). If you want to get ahead, get a theory. *Cognition, 3, 195*-212.

Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science, 12,* 79-83.

Kello, C. T. (2003). The emergence of a double dissociation in the modulation of a single control parameter in a single dynamic system. *Cortex, 39*, 132-134.

Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2006). *Fractal streams in human performance*. (submitted.)

Kello, C. T., Sibley, D. E., & Plaut, D. C. (2005). Dissociations in performance on novel versus irregular items: Single-route demonstrations with input gain in localist and distributed models. *Cognitive Science, 29*, 627-654.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 812-832.

Kirst, H., Fieberg, E. L., & Wilkening, F. (1993). Intuitive physics in action and judgment: The development of knowledge about projectile motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 952-966.

Klahr, D., & Siegler, R. (1978). The representation of children’s knowledge. In H. Reese & L. Lipsitt (Eds.), *Advances in child development and behavior* (pp. 61–116). New York, NY: Academic Press.

Kliman, M. (1987). Children’s learning about the balance scale. *Instructional Science, 15,* 307-340.

Langley, P. (1987). A general theory of discrimination learning. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development*. Cambridge, MA: MIT Press.

Laughlin, R. B. (2005). *A different universe: Reinventing physics from the bottom down*. New York: Basic Books.

Lawton, C. A. (1993). Contextual factors affecting errors in proportional reasoning. *Journal of Research in Mathematics Education, 24*,460-466.

Levin, I., Siegler, R. S., & Druyan, S (1990). Misconceptions about motion: Development and training effects. *Child Development, 61,* 1544-1557.

Machlup, S. (1977). Earthquakes, thunderstorms, and other 1/f noises. *Sixth International Symposium on Noise in Physical Systems* (pp. 157-160). National Bureau of Standards.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In M. G. M. Morris (Ed.), *Parallel distributed processing, implications for psychology and neurobiology* (pp. 8–45). Oxford: Clarendon Press.

McClelland, J. L. (1995). A connectionist perspective on knowledge and development. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 157–204). Hillsdale, NJ: Lawrence Erlbaum Associates.

Marsden, C. D., Merton, P. A., & Morton, H. B. (1983). Rapid postural reactions to mechanical displacement of the hand in man. *Advances in Neurology*, *39*, 645-659.

Normandeau, S., Larivee, S., Roulin, J. L., & Longeot, F. (1989). The balance-scale dilemma: Either the subject or the experimenter muddles through. *Journal of Genetic Psychology, 150*, 237-250

Raijmakers, M. E. J., van Koten, S., & Molenaar, P.C.M. (1996). On the validity of simulating stagewise development by means of PDP networks: Application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science, 20*, 101-136.

Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in elementary behaviors. *Journal of Motor Behavior, 34*, 99-125.

Rueckl, J. G. (2002). The dynamics of visual word recognition. *Ecological Psychology, 14*, 5-19.

Sage, S., & Langley, P. (1983). Modeling cognitive development on the balance scale task. In A. Bundy (Ed.), *Proceedings of the Eight International Joint Conference on Artificial Intelligence* (Vol. 1, pp. 94–96), Karlsruhe, Germany.

Schmidt, W. C., & Ling, C. X. (1996). A decision-tree model of balance scale development. *Machine Learning, 24,* 203-230

Siegler, R.S. (1981). Developmental sequences between and within concepts. *Monographs of the Society for Research in Child Development 46,* 1–74.

Sieger, R. S., & Chen, Z. (1998). Developmental differences in rule learning. *Cognitive Psychology, 33,* 273-310.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning, 16,* 57–86.

Shultz, T. R., Schmidt, W. C., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive development with a generative connectionist algorithm. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 205–261). Hillsdale, NJ: Lawrence Erlbaum Associates.

Turvey, M. T., & Carello, C. (1981). Cognition: The view from ecological realism. *Cognition*, *10*, 313-321.

Uttal, W. R. (1990). On some two-way barriers between models and mechanisms. *Perception & Psychophysics, 48*, 188-203.

Uttal, W. R. (1997). Do theoretical bridges exist between perceptual experience and neurophysiology? *Perspectives in Medicine and Biology, 40*, 28-302.

van der Maas & Molenaar, 1992

Van Orden, G. C., Holden, J. G., Podgornik, M. N., & Aitchison, C. S. (1999). What swimming says about reading: Coordination, context, and homophone errors. *Ecological Psychology, 11,* 45-79.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132,* 331-350.

Van Orden, G. C., & Kloos, H. (2005). The question of phonology and reading. M. J. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook*, 61-78, Oxford: Blackwell Publishing.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science, 25*, 111-172.

Van Rijn, H., van Someren, M., & van der Maas, H. L. J. (2002). Modeling developmental transitions on the balance scale task. Cognitive Science, 27, 227–257.

Watkins, M. J. (1990). Mediationism and the obfuscation of memory. *American Psychologist, 45,* 328-335.

Weldon, M. S. (1999). The memory chop shop: Issues in the search for memory systems. In J. K. Foster & M. Jelicic (Eds.) *Memory systems, processes, or functions*? New York: Oxford University Press.

Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental Study. *Cognitive Psycholgoy, 13,* 231-247.

Wilkening, F., & Anderson, N. H. (1982). Comparison of two rule-assessment methodologies for studying cognitive development and knowledge structure. *Psychological Bulletin, 92,* 215-237.

Are Dynamic Systems and Connectionist Approaches an Alternative to “Good Old Fashioned Cognitive Development?”

Lisa M. Oakes, Nora S. Newcombe, and Jodie M. Plumert

Abstract

The question of whether dynamic systems and connectionist approaches represent a new, unified theory of cognitive development leads naturally to an evaluation of what these approaches offer beyond traditional approaches to cognitive development. In our view, dynamic systems and connectionism bring a renewed focus on emergence in thinking about developmental change. Although this notion is also central to traditional theories of Good Old Fashioned Cognitive Development (GOFCD) such as those of Piaget, Gibson, and Vygotsky, the idea was lost in the heyday of nativism. In this chapter we trace the historical roots of concepts central to dynamic systems and connectionism, make connections between GOFCD and dynamic systems and connectionism, and illustrate through examples how work from more “traditional” perspectives has also revealed the process of emergence in developmental change. Finally, we consider connectionist and dynamic systems approaches to two classic areas of cognitive development, the A not B error and the balance scale problem, in the context of more traditional approaches to those problems. These comparisons allow us to examine how the new explanations refocus our attention on emergence and discuss the pitfalls that these approaches do and do not avoid.

Introduction

The goal of this volume, and the conference that motivated it, is to determine whether connectionism and dynamic systems are different theories of cognitive development or whether together they represent a paradigm shift in the field of cognitive development toward a new grand theory. The three authors of this chapter represent relative outsiders to this discussion. We each have studied cognitive development from a blend of “traditional” theoretical approaches, influenced by the theories of Piaget, Gibson, and Vygotsky, as well as by the broad range of theories that fall under the umbrella of information-processing approaches to cognitive development. We each represent a different blend of these theoretical frameworks, and we are all sympathetic to the general aims of and ideas behind connectionism and dynamic systems. We therefore have taken on the task of critically evaluating these “new” approaches from the perspective of Good Old Fashioned Cognitive Development (GOFCD), with an eye toward understanding what connectionism and dynamic systems bring to the field as well as understanding the extent to which they differ from more “traditional” approaches to cognitive development.

We have organized this chapter into four sections. In the first section we discuss what connectionism and dynamic systems bring to the study of cognitive development. Because many of the chapters in the volume deal directly with this issue, this section is relatively brief. In the second section, we examine how connectionism and dynamic systems relate to other GOFCD theories of developmental change. In some sense, this section traces the historical roots of the ideas in connectionism and dynamic systems. In addition, we show through examples of systematic programs of work examining developmental change in cognitive processes how the ideas inherent in connectionism and dynamic systems are not unique, although these “new” approaches may make some of these ideas more explicit and central than have others. Next, we evaluate the contribution of connectionism and dynamic systems in more depth by examining explanations of two historically important issues in cognitive development: infants’ behavior in the A not B task and children’s solutions to the balance scale problem. In this section we evaluate how these new theories compare to more traditional explanations of children’s developing behavior in these tasks. Finally we draw some conclusions about how connectionism and dynamic systems do or do not avoid the problems that have been the focus of criticisms of other theories of cognitive development.

Dynamic systems and connectionist approaches to the study of change

The field of cognitive development is broadly concerned with how children's thinking evolves from the preverbal representations of infancy to the high-level conceptual abilities of a 16-year-old learning calculus. The challenge, of course, is how to characterize and study such change and the causes of change. Dynamic systems and connectionist approaches to cognitive development are explicitly focused on understanding change over time. Moreover, understanding the *mechanisms* of change is critical for these viewpoints, and as a result they bring a focus on mechanism to the forefront. Each is concerned with demonstrating through mathematical models and careful empirical studies *how* change occurs, not simply documenting *that* change occurs. Proponents of the two viewpoints are concerned with how systems self-organize, with organization arising from a less organized (or sometimes unorganized) state through real-time processes and the dynamic activity of the system. The two approaches conceive of this self-organization differently. For dynamic systems, developmental change is an emergent product of interactions that occur on many different timescales (Smith & Thelen, 2003). Theories adopting this framework emphasize multi-causality and self-organization emerging out of the real-time dynamics of the child's own activity (Smith & Thelen, 2003). For connectionism, reorganization emerges out of nonlinearities in learning (Marchman, 1997; Thelen & Bates, 2003), and new structures only emerge from the interaction of the existing structure and environmental input (Bates & Elman, 2002; Elman, 2005). From an outsider's perspective, it is very difficult to distinguish between these two ways of thinking about change; for both, self-organization and emergent structure are a key feature of change, structural change emerges from change that occurs in real-time, and developing systems exhibit high levels of variability during the process of change. Central to both connectionism and dynamic systems, therefore, is the explicit idea that new structures and behaviors are emergent products of organism-environment interactions. Moreover, in both styles of theorizing, change at longer time scales necessarily emerges from change at shorter time scales because all behavior is linked together through time.

One consequence of these core ideas is that studies from a connectionist and dynamic systems perspective are often longitudinal. Because real-time change is intimately tied to change at longer time scales, development can be understood by observing change over many trials within an experimental session, over several sessions across several days or weeks (for examples see Joh & Adolph, 2006; Samuelson, 2002; Spencer, Vereijken, Diedrich, & Thelen, 2000; Thelen, Corbetta, & Spencer, 1996). Because the work from a GOFCD perspective often does not have as an explicit goal uncovering the mechanisms of change, longitudinal studies from these perspectives are relatively rare (for a notable exception, see Siegler, 1996). Instead, GOFCD theorists are generally interested in documenting changes in different states over relatively long time scales (e.g., months or years) using cross-sectional studies. It is difficult to reveal mechanisms of change over such time-scales. Thus, one contribution of the connectionism and dynamic system movements in cognitive development is to put the focus back on repeated observations over time. It must be pointed out that this approach is not new to cognitive development—Heinz Werner (1957) called it microgenesis, and Bob Siegler (1996) has strongly advocated and practiced this style of research over the past decade or more (see also Oakes & Plumert, 2002; Plumert & Nichols-Whitehead, 1996). It should also be noted that despite their interest in the connections between change on different time scales, most studies from a dynamic systems perspective examine change over relatively short time scales (such as trials or minutes). Thus, one weakness of many studies adopting a connectionist or dynamic systems framework is that they do not often examine changes over long time scales (Thelen's, work on reaching would be an exception, e.g., Spencer et al., 2000; Thelen et al., 1996).

Another key contribution of dynamic systems and connectionist approaches to cognitive development are the tools they provide for studying the emergence of new structures or behaviors from organism-environment interactions (Bates, Elman, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1998; Thelen, Schoner, Scheier, & Smith, 2001). As is evident from several of the chapters in this volume (see xxxxx, xxxxx, and xxxx), not only do connectionist and dynamic systems theorists value repeated observations over time, they also call for studies that seek to understand the processes that give rise to emergent behaviors. Such studies typically include perturbations or supports that change the organism-environment interaction. Usually, this entails manipulations of environmental structure (e.g., changing the salience of the B location and noting the effect on infants' reaching for the A location in the A not B task, Diedrich, Highlands, Spahr, Thelen, & Smith, 2001), but sometimes it entails manipulation of organism characteristics (e.g., teaching children shape-based categories in the lab, presumably changing how they approach the word-learning context in general, and noting the influence of this change on their rate of vocabulary acquisition outside the lab, Samuelson, 2000). Observing how manipulations of either the task or the organism changes the resulting behavior leads to a better understanding of the processes that give rise to behavior. Again, this type of approach is not unique to dynamic systems and connectionism. There are many examples from GOFCD and other approaches (e.g., see Lickliter, this volume) that explicitly engineer changes to either the organism or the environment to gain insight into developmental process (e.g., Oakes & Plumert, 2002; Plumert & Hund, 2001; Robinson, 2005).

As the chapters in this volume also make clear, these two theoretical paradigms have provided new mathematical or computational tools that make it easier to examine organism-environment interactions (see chapters xxxxxx, xxxxxx, xxxxx, this volume; and (Bates et al., 1998). For example, the dynamic field theory allows researchers to directly test how hypothesized inputs coming from the organism (e.g., memory or attention) and the environment (e.g., salience of perceptual information) interact to produce predictable patterns of behavior (e.g., Schutte, Spencer, & Schoner, 2003). Likewise, connectionist models of learning allow researchers to directly test how patterns of behavior emerge out of the interaction of simple processing units (e.g., Mareschal, Quinn, & French, 2002). Together, the conceptual and computational focus on how behavior emerges out of organism-environment interactions offers a significant step forward in our understanding of developmental process.

In summary, a major contribution of these two approaches to the field of cognitive development is a focus on emergence as a primary mechanism of change. Hence, any “grand new theory” from either a dynamic systems or connectionist perspective would seek to outline general principles that govern how new ways of thinking or behaving emerge from the interactions between the organism and the environment. Importantly, this “new” theory would make explicit this focus on the emergence of new behaviors through such interactions, and have as a central goal understanding those interactions rather than describing age changes in cognitive skills. In the next section we evaluate whether this would indeed be a new theory, and show how many GOFCD theories have also had this as a goal.

Is this a new way of thinking about developmental change?

GOFCD has long been interested in change, and the focus of much of the research and theorizing in the field of cognitive development is ultimately motivated by understanding developmental change. For example, information-processing theories, which has been criticized for focusing on *what* develops rather than on mechanisms of developmental change (Thelen & Smith, 1994), have described change in terms of increases in processing speed (Kail, 1986), numbers of relations a child can keep in mind (Halford, Wilson, & Phillips, 1998), the capacity and duration of memory stores (Case, 1985) and the availability of strategies for solving problems (Siegler, 1996). Although each of these examples is focused on describing what develops, they also represent an attempt to understand the mechanisms of change in broad cognitive abilities—that is, although these theories may not argue for qualitative shifts in cognitive structures or representational abilities, they attempt to account for the changes in behavior exhibited by children as they develop.

Why then does the emphasis on change in dynamic systems and connectionist approaches seem to be so unique and novel? In the 1980’s and 1990’s, nativist theories dominated the study of cognitive development. Such theories focused on identifying early emerging capabilities and not on how change occurs. This focus reflects, in part, the influence of Chomsky's (1968) notions that environmental events simply "trigger" pre-existing behaviors (i.e., those specified in biology). The highly publicized and influential work of people like Elizabeth Spelke (Spelke, Breinlinger, Macomber, & Jacobson, 1992; Spelke & Newport, 1998), Karen Wynn (1998) and Rochel Gelman (1978), all aimed at showing high-level cognitive competence at an early age rather than documenting the mechanisms that produce changes in cognitive abilities, led to the impression that the field of cognitive development had collectively lost an interest in understanding how change occurs.[1]

However, *emergence* as a mechanism of change has a long history in the study of cognitive development, and was central to the theories of Piaget, Gibson, and Vygotsky (a fact that is acknowledged by theorists from both connectionism and dynamic systems perspectives, see Bates & Elman, 2002; Thelen & Bates, 2003). Piaget, for example, proposed that new mental structures emerge through the dynamic interplay between the child's developing cognitive structures and input from the environment. Similarly, for Gibson, changes in the organism lead to increased sensitivity to environmental structure, which in turn leads to changes in the organism (at both neural and behavioral levels). Hence, change emerges out of cyclical organism-environment interactions over both shorter and longer time scales (Gibson, 1988; Gibson & Pick, 2000). For Vygotsky, new skills emerge at times when children are sensitive to social experiences that allow them to try out new ways of thinking and acting, sometimes referred to as the zone of proximal development (Wertsch, 1985). The notion that adult guidance must be developmentally appropriate necessarily implies that cognitive change emerges out of the interaction of the child and the social environment. Thus, the idea that behaviors emerge through interactions between the organism and the environment has been central to our understanding of development for quite some time, and these historical views on emergence have played an important role in the application of dynamic systems and connectionism frameworks to understanding development.

Importantly, however, the ideas about emergence in the theories of Piaget, Gibson, and Vygotsky had largely been lost or ignored over time, even by theorists who came from these traditions. Modern theorists whose work originated in a Piagetian tradition shifted focus from understanding how cognitive structures emerge out of the interaction of the child and the world to an almost exclusive focus on the cognitive structures or concepts themselves (Flavell, 1970). Gibsonian theorists have shifted away from viewing affordances as an emergent property of the interaction between the organism and environment to viewing affordances as objective properties of the environment - "information available about surfaces, places, obstacles, and things as well as about oneself" (Gibson, 2003, p. 293). Likewise, socio-cultural approaches to cognitive development have almost exclusively focused on the social environment and have had very little to say about how organism characteristics interact with social structure to produce changes in thinking. We believe these shifts in theoretical perspectives over time have occurred because the notions of *interaction* and *emergence* are very difficult concepts. It is difficult to think about behavior or thinking as being simultaneously determined by organism characteristics and environmental structure. It is much easier to assign causal priority to one or the other, rather than to both at the same time.

As a result, many studies of cognitive development from a GOFCD perspective may appear to be “merely” descriptive at first blush, rather than revealing processes of cognitive change. However, the rich descriptions of cognitive development amassed over the last several decades have provided understanding of the mechanisms of change, although the explicit goals of these investigations may not have always been to uncover such mechanisms. Thus, we have made progress in our understanding of mechanisms of change, and even in our understanding of how behavior and structure emerge from organism-environment interactions. This progress has not occurred despite a focus on description, but rather is intimately tied to the descriptive success. In part, the two enterprises are linked because the rich description of change strongly constrains the search for explanation. But even more than that, rich description focuses attention on plausible mechanisms and explanatory principles. These mechanisms can then be tested in studies that may themselves seem (on the surface) descriptive. That is, investigators often test explanatory ideas by evaluating predictions regarding “descriptive” questions, such as how manipulations in task characteristics or in environmental input will affect children’s behavior or cognitive abilities.

Consider an example of a cognitive change: the emergence of configural, as opposed to featural, processing of visual stimuli. We may simply ask: at what age do people process upright and inverted faces differently (an effect that has been interpreted as a hallmark of configural rather than merely featural processing)? This may seem a supremely descriptive question (although an improvement over a global question such as when children can process faces). Yet, the answer revealed by research asking this descriptive question turns out to be quite complex. On the basis of classic work by Carey and Diamond (1977) we would conclude that this change occurs at around 10 years of age. Carey and Diamond’s observation may lead one to conclude that what propels a change to configural processing are prolonged developmental changes—maturational and experiential factors that occur over many years. Evidence that configural processing is in place by 4 or 5 months (Bhatt, Bertin, Hayden, & Reed, 2005; Turati, Sangrigoli, Ruel, & de Schonen, 2004), in contrast, lead to a very different account of what factors influences a change from featural to configural processing. If configural face processing emerges early, then infants’ experience with faces during the period during which the visual system is organized becomes a potentially important factor contributing to that change. Indeed, early restriction of visual experience, as occurs when infants are born with congenital cataracts, has an enduring influence on face processing (Le Grand, Mondloch, Maurer, & Brent, 2004)—an observation that might appear to be merely a descriptive fact, but that is also a test of a mechanism.

Moreover, it turns out that the emergence of configural processing is extremely complex: there are three types of configural processing, and they mature at different rates (Mauer, Le Grand, & Mondloch, 2002). Although some kinds of configural face processing may be present in infancy, others may not reach adult levels until 14 years. In addition, effects of early visual deprivation can co-occur with surprising levels of adult plasticity (Maurer, Lewis, & Mondloch, 2005), as may arguably be seen in the advent of configural sensitivity to non-face stimuli (Gauthier & Tarr, 2002). Taken individually, all of this research could be characterized as descriptive; collectively the data (at a minimum) strongly constrain the family of possible accounts of developmental mechanisms. This body of work represents an important feature of rich description: Determining the timing of environmental effects and residual plasticity allows us to move beyond minimal explanations of developmental change (i.e., determining the timing and sequence of the acquisition of particular aspects of adult competence) to formulating sophisticated (and explanatory) accounts of the interaction of environmental inputs and neural substrates.

In our view, the (more or less explicit) goal of work conducted from a variety of GOFCD perspectives has been to understand exactly this kind of mechanism of developmental change. Examples from our own work will serve as illustrations of how one can conduct systematic studies of the conditions under which a particular behavior is and is not seen from a GOFCD approach. Such systematic investigations provide rich descriptions that tell us about the emergence of that behavior. Ours are not the only programs of research that exemplify this trend, but they illustrate how studies that do not have the same explicit goals of dynamic systems and connectionism can reveal a great deal about emergence and interactions. Importantly, many of these studies did have as a focus revealing the mechanisms of developmental change—or testing hypotheses about how such mechanisms constrain children's emerging behaviors. Moreover, these programs of research, although not explicitly adopting a dynamic systems or connectionist framework, are aimed at uncovering how child-environment interactions operate to allow new behaviors to emerge.

For example, in the area of infant categorization, Oakes and her colleagues have shown that infants' categories are highly flexible and context-dependent (see Oakes, Horst, Kovack-Lesh, & Perone, in press, for a review). At 10 months infants will respond to a category of *people* that excludes other land mammals (such as horses) in an object-examining task, but not in a less structured sequential-touching task (Oakes, Plumert, Lansink, & Merryman, 1996). By 13 months, infants will attend to this distinction even in the less structured task (Oakes et al., 1996). Similarly, Oakes and Ribar (2005) observed that 4-month-old infants would only respond to a narrow category of dogs *versus* cats in a visual familiarization task that minimized the memory and other cognitive demands, whereas 6-month-old infants would respond to that distinction even in a more cognitively demanding familiarization task. The point is that, in each of these investigations, infants' ability to respond to a particular category is a function of the interaction between their cognitive abilities (such as their ability to encode and recall individual items) and the structure of the task. When the task places demands on less sophisticated abilities, infants have difficulty recognizing subtle distinctions (such as that between dogs and cats). When the task places fewer demands on those less sophisticated abilities, or as those abilities develop, infants can more easily recognize subtle distinctions between categories. Infants' recognition of the category, therefore, is emergent: it resides in the interaction between their abilities and the task context. They may have knowledge and past experience relevant to recognizing the category in the moment, but their behavior in the experimental task reflects more than this existing knowledge and past experience. Similarly, infants clearly have the ability to form narrowly defined categories even at 4 months of age, but whether or not they actually form such categories at any given moment is complexly determined by the interaction of their existing knowledge, the cognitive abilities they use to form such categories, and the context in which they are forming those categories.

In a very different domain, Plumert and her colleagues have studied how categorical bias in memory for location emerges out of the interaction of the cognitive system and the task structure (see Plumert, Hund, & Recker, in press, for a review of this work). Categorical bias refers to the tendency to remember locations in the same spatial group (e.g., the same quadrant of a room) as closer together than they really are. In studies in which children and adults learn the locations of 20 miniature objects in a large, open box, memory for those location is determined not only by the age of the individual (i.e., 7-year-olds have less accurate memories than do adults), but also by contextual factors such as the cues available for organizing the locations into groups during learning (Hund & Plumert, 2003; Hund, Plumert, & Benney, 2002; Plumert & Hund, 2001), how long participants have to maintain memory for locations over time (i.e., how memory changes as the delays between learning and test are increased) (Hund & Plumert, 2002, 2005), and by the cues available at test (Plumert & Hund, 2001). Importantly, this work reveals that memory for locations in this context is not solely determined by task structure (e.g., presence or absence of boundaries during learning) or by developmental differences in the cognitive system (e.g., age), but rather performance in this task is complexly determined by the interaction of these two types of factors. For example, adults always show significant categorical bias when there is at least one cue that helps organize the locations into groups (i.e., boundaries that divide the space into four quadrants), but they do not show such bias when no organizing cues are available during learning. In addition, children and adults respond differently to the presence of cues (i.e., some cues lead to much greater categorical bias in the placements of adults than of children), suggesting they extract different things from their experience with these tasks even though the task structure is identical for all participants. These variations in how the same age group responds to different task structure and how different age groups respond to the same task structure support the idea that categorical bias emerges out of the interaction of the cognitive system and the task structure.

A final example from work by Newcombe and her colleagues also illustrates how spatial thinking emerges out of the interaction of the child and the task. In a task in which children are required to find a hidden object after being disoriented (i.e., they are turned around with closed eyes), Newcombe and her colleagues have found that children’s use of geometric properties (e.g., the relative length of walls defining enclosures) and featural information (e.g., colors or markings on surfaces) to reorient is not solely a function of the child’s underlying competence or developmental level. Rather, the exact mix of features children use to reorient depends upon a confluence of factors such as the reliability, variability, and usefulness of the sources of information and the certainty with which the sources of information have been encoded. This work is relevant to a debate regarding the origins of spatial knowledge. According to a nativist modular view, various sources of spatial information are processed independently in separable cognitive processing units (e.g., Wang & Spelke, 2002). Support for this position comes from the remarkable fact that a wide variety of animal species share a powerful sensitivity to geometric properties of enclosing spaces (e.g., the relative length of walls defining enclosures), using such information to reestablish spatial orientation after being disoriented (Cheng, 1986; Hermer & Spelke, 1996) for a review, see (Cheng & Newcombe, 2005) Based on findings that both rats and human children fail to use nongeometric (or featural) information (e.g., colors or markings on surfaces), even when use of featural information would disambiguate geometrically congruent locations, some have suggested that such geometric processing constitutes a specialized cognitive module that is impenetrable to nongeometric information, even when that information has been processed (Gallistel, 1990) and that this module is innately available (Hermer & Spelke, 1996).

Newcombe and Ratliff (in press) have proposed, in contrast, that the existing data on integration of featural and geometric information can be best explained by an adaptive combination approach in which the likelihood of using the two kinds of information varies depending on factors such as uncertainty, cue validity and cue salience (e.g., Ernst & Banks, 2002; Huttenlocher, Hedges, & Duncan, 1991). Newcombe and her colleagues have documented the conditions under which children use the colors of walls to find hidden objects, thereby showing that use of geometric vs. featural information is not an all-or-none phenomenon at a given age. For example, younger children are more likely to use featural information in larger rooms and under conditions where they can move about the space (Newcombe & Ratliff, in press). These differences in performance from children of the same age under different task conditions are difficult to explain from a modularity perspective. Hence, this program of research has provided the kind of descriptive data needed to differentiate between opposing theoretical views of the origins of fundamental spatial skills.

A closer look at explanations of change: How do dynamic systems and connectionist theorists explain change on the A not B and the balance scale tasks?

In this section we will further discuss the general approaches of dynamic systems and connectionism by examining specific examples of how they have been used to explain developmental change in two domains: the A not B error and the balance scale (or balance beam) task. These two domains provide particularly good contexts for understanding the contributions of these approaches because they are classic problems that have been studied from the perspectives of GOFCD, connectionism, and dynamic systems (although a formal dynamic systems model of the balance scale problem has not yet been proposed). Thus, these domains allow us to consider dynamic systems and connectionism explanations of well understood cognitive problems, as well as to consider how these approaches are (or are not) an advance over previous, more traditional cognitive development approaches.

The case of the A not B error. In his classic work on the developing understanding of object permanence during the sensorimotor period, Piaget noted that toward the end of the first year of life infants exhibit somewhat odd behavior toward hidden objects. Although infants can retrieve a hidden object by 8 or 9 months of age (when they are in Stage IV of Piaget's sensorimotor period), they make the "A not B" error. In the prototypical error, infants observe an attractive object hidden underneath a cloth, and they successfully retrieve the object. After several such hidings (repeatedly in the same location, location *A*), the experimenter hides the object (while the infant watches) under a different cloth (or location *B*). Infants in Piaget's stage IV of sensorimotor development (from approximately 8 to 12 months of age) search not where they have just observed the object hidden, but rather where they had successfully uncovered it on previous trials. In other words, after seeing the object hidden at location *B*, infants search for the object in location *A*.

This puzzling behavior has long been the focus of study in infant cognitive development. A large number of studies have been conducted to replicate the conditions under which infants make this error and to understand why infants fail to search in the location they have just seen the object hidden. Piaget's classic interpretation is that this behavior indicates an incomplete understanding of object permanence. According to Piaget, the infant "seems to reason as if the place where the object was found the first time remains where he will find it when he wants to do so" (Piaget, 1954, pp. 46-47) even though the infant directly observes the object hidden in the new location! The infant has developed awareness that the object can be retrieved even when it is out of sight, but the infant does not fully understand the relationship between the hiding place and where the object can be retrieved. Rather, according to Piaget, the infant conceives of the object *only* in the first place that he or she saw the object hidden and successfully retrieved it.

There have been many explanations for the A not B error, ranging from an appeal to innate or early emerging conceptual abilities (Baillargeon, 2004) to a discussion of the development of neuranatomical structures responsible for inhibiting prepotent responses (Diamond, 1991). The explanations provided by the dynamic systems and connectionist approaches grew out of these previous theories (and indeed rely on the large corpus of data that has been collected examining the conditions under which the A not B error is obtained). The dynamic systems and connectionism explanations differ from those that came before them in that they focus directly on the error as an emergent product of the development of the infants' underlying cognitive structures and the task demands. Previous explanations have focused on what the error reveals about the development of the type of underlying representations. Indeed, context effects in infants' performance were viewed as reflecting in the masking of underlying competence, and as a result the A not B error could be overcome by using tasks that more effectively tapped the underlying competence (Baillargeon & Graber, 1988).

In contrast, dynamic systems and connectionist theories of the A not B error provide explanations of how the error emerges and changes in real-time (i.e., in the context of a particular trial or experimental session), and how that emergence is a product of the task factors (such as the salience of a particular hiding event) interacting with developmental changes in memory representational abilities, motor abilities, and other factors. For example, the dynamic systems account is that the error is actually an error in reaching, determined by the competition between the transient memory of the hiding event and the longer-term motor memory for past reaches (Smith, this volume, Smith & Thelen, 2003; Thelen et al., 2001). In clever and systematic studies, Smith and Thelen and their colleagues have shown that the incidence of the error increases by increasing the strength of the motor memory for reaches at A (e.g., increasing the number of reaches at A) and decreasing the salience of the transient memory for the hiding event at B (e.g., making B more similar to A) and that the incidence of the error decreases by reducing the strength of the motor memory (e.g., changing the child's posture or arm weights before reaching on the B trials) or increasing the salience of the hiding event at B (e.g., making B very different from A) (Smith, this volume, Smith & Thelen, 2003). These types of manipulations are explicitly designed to perturb the components hypothesized to play a role in infants' behavior in the A not B task. In so doing, they reveal how organism-environment interactions give rise to the A not B error.

For Munakata and Morton's connectionist model of the A not B error, the strength of the underlying representation for the hiding event at A competes with the underlying representation for the hiding event at B. The system creates an active representation for the more recent hiding event (the working memory-like component) and a latent representation for previous hiding events (the long-term memory-like component) (Morton & Munakata, this volume, Munakata, 1998; Munakata, Morton, & Stedtron, 2003). Infants' representation of the location of the object is graded due to the competition between latent and active memory for the hiding event. As was true for the A not B model of Smith and Thelen, early in development the transient memory for the hiding event at B cannot compete with the stronger memory for the hiding event at A that was built up over time. With development, the active memory for the hiding event at B becomes more effective at competing, and as a result the error is reduced. Moreover, Munakata and Morton can evaluate how this process works in real time by examining how connection weights change as the child reaches several times to A and then sees the object hidden at B.

In both cases, on-line dynamics of the context is considered the cause of the behavior. Both theories have received significant support, with empirical tests confirming predictions made by each model. Clearly, therefore, the two approaches are similar in many ways. They both describe the behavior in terms of the competition between longer-term, more established memories and shorter-term, more transient memories. This raises the central question of this volume: Are these actually separate theories, or are they two variations of the same explanation of development? There are important differences in the details of the theories—a key component of Munakata and Morton’s model (Morton & Munakata, this volume, Munakata, 1998; Munakata et al., 2003) is the internal representation (in terms of connection weights and recurrence) of the hiding events, whereas the hiding event and object are not represented in any obvious way in the dynamic systems theory of Smith and Thelen (Smith, this volume, Smith & Thelen, 2003; Thelen et al., 2001). Similarly, the dynamic systems account is embodied, with the infant’s memory for his or her own action playing an important role in behavior. The connectionist model does not have an obvious component for memories for actions carried out by a real body in actual space.

Are these simply re-labeling of the same components? Perhaps the accounts of the A not B error are more similar than they are different, as well as having commonalities with more traditional ways of discussing the issues (Newcombe, 2001; Newcombe & Huttenlocher, 2000). Thus, this comparison may not be the best one for understanding the differences between the two approaches. But, what is clear is that the two explanations clearly derive from different theoretical stances, and as a result they have incorporated different components into their models and they have argued for the role of different aspects of development in this error. It may be that the “truth” is a blending of these two approaches, involving embodied cognition in which objects and locations are represented in long-term memory and carried out by physical bodies in real space. Indeed, as is clear from several chapters in this volume, in several domains theorists are blending of connectionism and dynamic systems to create new models of development (see xxxxx, xxxxx, and xxxxx). What is important for the present discussion is that both of these models of the A not B error have attempted to understand the error in terms of the real-time dynamics of the task.

The case of the balance scale task. In a very different series of observations, Inhelder and Piaget (1958) documented children's responses to the "balance scale" task. In this task, children are presented with a balance scale and they must predict whether scale will balance or tip when different configurations of weights are placed on each side of the fulcrum. To correctly solve this problem, children must consider not only the amount of the weight placed on each side, but the distance of the weights from the fulcrum. Based on children's performance on this task, Inhelder and Piaget argued for qualitative changes in children's thought. Concrete operational children were unsuccessful at solving this problem; they typically believe that weight is the only relevant factor. Children who had attained formal operations (and therefore were capable of propositional thought and hypothetical-deductive reasoning), in contrast, could successfully combine the two relevant variables in an accurate way, implicitly or explicitly computing torque.

As with the A not B task, there have been subsequent explorations of children's performance on the balance scale task from other perspectives than Piaget's. For example, a well-known series of studies by Siegler (1976, 1981), adopting an information-processing approach to cognitive development, used rule-assessment methodology to delineate a succession of four modes of reasoning on balance scale problems, beginning with children of 5 or 6 years. Importantly, Siegler's work clearly addressed questions about mechanisms of development, such as when, for example, he showed that it was possible to diagnose which children would respond to environmental input regarding how the balance scale worked.

Subsequent work on the balance scale problem has, for the most part, been within the GOFCD tradition. There has been discussion of issues such as whether rule-assessment methodology should be supplanted or supplemented by other techniques, for example, by latent class analysis (Jansen & van der Maas, 2002), and whether children younger than 5 years can also exhibit systematic (albeit one-dimensional) reasoning about these problems (Halford, Andrews, Dalton, Boag & Zielinski, 2002). From these studies, a rich portrait has been emerging of developmental change on this task, including an interesting suggestion that some aspects of change are abrupt and stage-like, while, at other periods, various rules or strategies co-exist as children experiment with solutions (Jansen & van der Maas, 2002).

There have also been efforts to model children's behavior. Some of them have used connectionist approaches (e.g., McClelland, 1989, 1995; Shultz, Mareschal, & Schmidt, 1994), while others have not. Dynamic systems theorists have not dealt with the balance scale, arguably because thinking about the problem is a more symbolic and logical task than A not B, and less intimately involved with motor experience. However, van der Maas and his colleagues have developed a cusp model of children's performance on the balance scale task, derived from catastrophe theory and with some over-arching points of contact with dynamic systems thinking (Jansen & Van der Maas, 2001). These models have been less successful in making novel predictions and stimulating new research than connectionist or dynamic systems models of A not B, and, as argued at length by van der Maas and Raijmakers (this volume), they do not even do a complete job of accounting for the known empirical phenomena. However, as was the case for models of the A not B error, such models represent a hybrid or mix of concepts from different traditions. Interestingly, van der Maas has reecntly offered an ACT-R model of children's performance on the balance scale task (see van der Maas & Raijmakers, this volume, van Rijn, van Someren, & van der Maas, 2003), which involves the postulation of explicit rules. This model seems to do the best job so far of capturing how children really behave. Although van der Maas and Raijmakers remain optimistic regarding the potential of neural networks with nonlinear properties, they suggest that more symbolic approaches seem to be working better in the short run.

Overall, the history of dynamic systems and connectionist approaches to the balance beam problem suggests caution about how easy it will be to apply the techniques used so far in these traditions to higher cognitive functioning. GOFCD has made good progress in delineating the progression of children's behavior in this realm, and in suggesting reasons for developmental change that may differ at different ages, and that include encoding of relevant information, working memory, experimentation with strategies, and direct instruction. It remains to be seen how successful nonlinear, dynamic models will be in explaining development in such domains, and how those models combine symbolic models with ideas from connectionism and dynamic systems.

Can dynamic systems and connectionism evade classic issues facing the study of cognitive development?

In this chapter we have examined the contributions of dynamic systems and connectionist viewpoints, and have argued that many of these contributions are not completely new ideas in the study of cognitive development. Rather, dynamic systems and connectionism bring ideas such as emergence front-and-center in our explanations of developmental change, and therefore play an important role in shaping how we talk and think about development. An important issue is whether these new frameworks avoid the problems of many previous theories of cognitive development. That is, the same criticisms have been raised about many different theories of cognitive development. If dynamic systems and connectionism are in a position to replace traditional GOFCD frameworks, they must be immune to those criticisms. We have identified at least four problems characteristic of many theories of cognitive development—some of these problems are solved by dynamic systems and connectionist approaches to cognitive development, others are problematic for these approaches, just as they are for GOFCD approaches.

First, all models of development must grapple with the difficult problem of characterizing the environment and the actual nature of the input and feedback that real children really receive.

Although the computational models developed from connectionist and dynamic systems must describe the input and feedback the models receive, as was discussed in the earlier section of the A not B error, the description of this input often appears arbitrary. When comparing the models of the A not B error, we made the argument that these two models may be simply using different labels for essentially the same input. Some connectionist models have attempted to understand how the statistics of the input influences what children learn. For example, Mareschal and his colleagues (Mareschal, French, & Quinn, 2000) found that variability in the head and face regions of dog and cat stimuli predicted both how connectionist models and 3- to 4-month-old infants learned the categorical distinction between dog and cat. French et al. (2002) found that by blurring the visual input in a way that mimicked the visual acuity of young infants, a connectionist model would to respond to the same kinds of categorical distinctions as do infants. A few theorists have attempted to characterize the kind of input children receive. Smith and her colleagues (Yoshida & Smith, 2001) have examined the structure of the kinds of nouns to which children from different language environment are exposed and how those different kinds of inputs help to create biases in children’s assumptions about the referents of new words. Clearly, characterizing the nature of the input is extremely difficult. However, as recently pointed out by Saffran (in press; Saffran, Reek, Niebuhr, & Wilson, 2005) we cannot really understand the learning mechanism without understanding *what* is being learned.

A second problem that faces all models of development is how to characterize the start states for development in the newborn, the end states in the adult (and indeed, determining whether or not there is an end state). Dynamic systems and connectionist models must be very explicit about the start and end states of their behavior (e.g., determining the weight or value of some parameter that refers to strength of memory trace). But, such determinations are quite arbitrary, and it is not often clear how they map onto actual developmental starting states in the child. In other words, where did those start states came from (i.e., what developmental events occurred before to create the starting state of the system)? In addition, it is not always clear whether and when development of this system ends. Because for dynamic systems and connectionism, change is always occurring in response to input, or as parts of the system interact, they have not taken seriously the task of characterizing the starting and ending states. Some might argue that this is not the goal of these models, but the initial state of the system (e.g., the limited motoric and perceptual abilities of the newborn) place important constraints on how the system operates and thus what kinds of organism-environment interactions are possible.

Third, theories that focus on emergence must attempt to explain the nature of organism-environment interactions. Clearly, one advantage to these approaches is that they have brought the discussion of change front and center—in some sense, connectionist and dynamic systems theories are defined by their approach to emerging structure and their focus on understanding that emergence. In this way, these frameworks appear to address the criticism of GOFCD theories of development that such theories are "merely" descriptive. Although we believe that many cognitive developmentalists have indeed been interested in the mechanisms of change (and have collected data relevant to understanding those mechanisms), we applaud the dynamic systems and connectionist frameworks for making this discussion explicit, and for adding tools to our arsenal of methods for understanding change.

Finally, development is characterized by variability, and theories of cognitive development must come to terms with the high levels of variability that emerge with developmental change and task variation. For many, such variability is a nuisance; the goal is to develop sensitive tasks that eliminate sources of extraneous variability and reveal children's true competence. Dynamic systems and connectionism embrace variability in behavior, and use variability over time or with task variations as a way of explaining developmental changes in the environment-organism interactions (see also Siegler, 1996). In fact, for the dynamic systems perspective, variability is a necessary precursor of change—systems must go through some sort of instability in order to change. Importantly, dynamic systems and connectionism provide an overall framework for thinking about the role of the environment or task, but understanding how specific tasks operate at specific points in development is a problem that can only be addressed by continued empirical work in the field.

Conclusions

Is this a new grand theory of development that represents a radical departure from "good old-fashioned cognitive development?" We don't think so. We think that, simply, but not at all trivially, this grand theory of development has helped to refine the explanations of development that have emerged in the field and to enhance our search for the principles of emergence. Thus, these new approaches have been and will continue to be important in shaping how we collectively think about developmental change. But, it is important to remember that the core notions of the mechanisms of developmental change have played an important role in our thinking about and studying of cognitive development for a long time, and that dynamic systems and connectionism are not immune to the criticisms that have been raised about GOFCD theories.

Author Notes

Preparation of this chapter was made possible by support from NIH grants HD49840 and HD49143 awarded to LMO, NSF grant BCS-0414302 awarded to NSN, and NSF grant BCS-0343034 awarded to JMP. We thank Bob Siegler for discussions about the state of work on the balance scale. Correspondences should be addressed to Lisa M. Oakes, Department of Psychology, University of Iowa, Iowa City, IA 52245 (JOHN: This will change after July 1, but I assume that for now we keep it Iowa)

Footnotes

1. It should be pointed out that although Rochel Gelman's views have historically been nativist in flavor, her recent chapter with Lucariello on learning in the third edition of *Stevens' Handbook of Experimental Psychology* (Gelman & Lucariello, 2002) is remarkably non-nativist. Thus, the reference here is to the work in the 1980's and 1990's that contributed to the nativist movement in the field of cognitive development, rather than the most recent writings by Gelman, or any of the cited authors.

References

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science, 13,* 89-94.

Baillargeon, R., & Graber, M. (1988). Evidence of location memory in 8-month-old infants in a nonsearch ab task. *Developmental Psychology, 24,* 502-511.

Bates, E. A., & Elman, J. L. (2002). Connectionism and the study of change. In M. H. Johnson (Ed.), *Brain development and cognition: A reader (2nd edition) (*pp. 623-642). Oxford: Blackwell.

Bates, E. A., Elman, J. L., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1998). Innateness and emergentism. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science (*pp. 590-601). Oxford: Basil Blackwood.

Bhatt, R. S., Bertin, E., Hayden, A., & Reed, A. (2005). Face processing in infancy: Developmental changes in the use of different kinds of relational information. *Child Development, 76,* 169-181.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195*, 312-314.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition, 23*, 149-178.

Cheng, K., & Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin & Review, 12,* 1-23.

Chomsky, N. (1968). *Language and mind*. New York: Brace & World.

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition*. Hillsdale, NJ: Erlbaum.

Diedrich, F. J., Highlands, T. M., Spahr, K. A., Thelen, E., & Smith, L. B. (2001). The role of target distinctiveness in infant perseverative reaching. *Journal of Experimental Child Psychology, 78,* 263-290.

Elman, J. L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences, 9*, 112-117.

Ernst, M., & Banks, M. (2002). Humans integrage visual and haptic information in a statistically optimal way. *Nature, 415,* 429-433.

Flavell, J. H. (1970). Concept development. In P. H. Mussen (Ed.), *Carmichael's manual of child pscyhology (Vol. 1*). New York: Wiley.

French, R. M., Mermillod, M., Quinn, P. C., Chauvin, A., & Mareschal, D. (2002). *The importance of starting blurry: Simulating improved basic-level category learning in infants due to weak visual acuity*. Paper presented at the Proceedings of the 24th Annual Conference of the Cognitive Science Society.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431-446.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gelman, R., & Lucariello, J. (2002). Learning in cognitive development. In C. R. Gallistel (Ed.), *Stevens' handbook of experimental psychology (3rd ed., Vol. 3,* pp. 395-443). New York: Wiley.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology (Vol. 39,* pp. 1-41): Annual Reviews.

Gibson, E. J. (2003). What psychology is about: Ruminations of an opinionated aged psychologists. *Ecological Psychology, 15,* 289-295.

Gibson, E. J., & Pick, A. D. (2000). An ecological approach to perceptual learning and development. *(2000), vi,* NY, US: Oxford University Press.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803-864.

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition, 61,* 195-232.

Hund, A. M., & Plumert, J. M. (2002). Delay-induced bias in children's memory for location. *Child Development, 73,* 829-840.

Hund, A. M., & Plumert, J. M. (2003). Does information about what things are influence children's memory for where things are? *Developmental Psychology, 39,* 939-948.

Hund, A. M., & Plumert, J. M. (2005). The stability and flexibility of spatial categories. *Cognitive Psychology, 50,* 1-44.

Hund, A. M., Plumert, J. M., & Benney, C. J. (2002). Experiencing nearby locations together in time: The role of spatiotemporal contiguity in children's memory for location. *Journal of Experimental Child Psychology, 82*, 200-225.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352-376.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolsecence*. New York: Basic Books.

Jansen, B. R., & Van der Maas, H. L. (2001). Evidence for the phase transition from rule i to rule ii on the balance scale task. *Developmental Review, 21,* 450-494.

Jansen, B. R., & van der Maas, H. L. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology, 81,* 383-416.

Joh, A. S., & Adolph, K. E. (2006). Learning from falling. *Child Development, 77,* 89-102.

Kail, R. (1986). Sources of age differences in speed of processing. *Child Development, 57,* 969-987.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15,* 762-768.

Marchman, V. A. (1997). Models of language development: An "emergentist" perspective. *Mental Retardation and Developmental Disabilities Research Reviews, 3,* 293-299.

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology, 36,* 635-645.

Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month olds' sequential category learning. *Cognitive Science, 26*, 377-389.

Mauer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255-260.

Maurer, D., Lewis, T. L., & Mondloch, C. J. (2005). Missing sights: Consequences for visual cognitive development. *Trends in Cognitive Sciences, 9*, 144-151.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In R. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology (*pp. 8-45). New York: Oxford University Press.

McClelland, J. L. (1995). A connectionist approach to knowledge and development. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competenece: New approaches to process modeling (*pp. 601-613). Mahwah, NJ: Erlbaum.

Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A pdp model of the ab task. *Developmental Science, 1*, 161-184.

Munakata, Y., Morton, J. B., & Stedtron, J. M. (2003). The role of prefrontal cortex in perseveration: Developmental and computational explorations. In P. Quinlan (Ed.), *Connectionist models of development (*pp. 83-141). East Sussex: Psychology Press.

Newcombe, N. S. (2001). A spatial coding analysis of the a-not-b error: What is "locatin at a"? *Behavioral & Brain Sciences, 24,* 57-58.

Newcombe, N. S., & Huttenlocher, J. (2000). *Making space: The development of spatial representation and reasoning*. Cambridge, MA: MIT Press.

Newcombe, N. S., & Ratliff, K. R. (in press). Explaining the development of spatial reorientation: Modularity-plus-langauge versus the emergence of adaptive combination. In J. M. Plumert & J. P. Spencer (Eds.), *Emerging landscapes of mind: Mapping the nature of change in spatial cognitive development*. New York: Oxford University Press.

Oakes, L. M., Horst, J. S., Kovack-Lesh, K. A., & Perone, S. (in press). How infants learn categories. In A. Woodward & A. Needham (Eds.), *Learning and the infant mind*. New York: Oxford University Press.

Oakes, L. M., & Plumert, J. M. (2002). Variability in thirteen-month-old infants' touching patterns in the sequential-touching task. *Infant Behavior & Development, 25,* 529-549.

Oakes, L. M., Plumert, J. M., Lansink, J. M., & Merryman, J. D. (1996). Evidence for task-dependent categorization in infancy. *Infant Behavior and Development, 19,* 425-440.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paried and successive presentation familairization tasks. *Infancy, 7*.

Piaget, J. (1954). *The construction of reality in the child*. Oxford, England: Basic Books.

Plumert, J. M., & Hund, A. M. (2001). The development of memory for location: What role do spatial prototypes play? *Child Development, 72,* 370-384.

Plumert, J. M., Hund, A. M., & Recker, K. M. (in press). Organism-environment interaction in spatial development: Explaining categorical bias in memory for location. In J. M. Plumert & J. P. Spencer (Eds.), *Emergiong landscapes of mind: Mapping the nature of change in spatial cognitive development*. New York: Oxford University Press.

Plumert, J. M., & Nichols-Whitehead, P. (1996). Parental scaffolding of young children's spatial communication. *Developmental Psychology, 32*, 523-532.

Robinson, S. R. (2005). Conjugate limb coordination after experience with an interlimb yoke: Evidence for motor learning in the rat fetus. *Developmental Psychobiology, 47,* 328-344.

Saffran, J. R. (in press). What can statistical learning tell us about infant learning? In A. Woodward & A. Needham (Eds.), *Learning and the infant mind*. New York: Oxford University Press.

Saffran, J. R., Reek, K., Niebuhr, A., & Wilson, D. (2005). Changing the tune: The structure of the input affects infants' use of absolute and relative pitch. *Developmental Science, 8,* 1-7.

Samuelson, L. K. (2000). *Statistical regularities in vocabulary guide language acquisition in 15- 20-month-olds and connectionist models*. Unpublished Doctoral Dissertation, Indiana University, Bloomington.

Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology, 38,* 1016-1037.

Schutte, A. R., Spencer, J. P., & Schoner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development, 74,* 1393-1417.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning, 16,* 57-86.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8,* 481-520.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46,* 84.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY, USA: Oxford University Press.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7,* 343-348.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605-632.

Spelke, E. S., & Newport, E. L. (1998). Nativism, empiricism, and the development of knowledge. *Damon, William (Ed); Lerner, Richard M (Ed), (1998). Handbook of child psychology: Volume 1: Theorectical models of human development (5th ed.). (pp. 275-340). xxii,* NJ, US: John Wiley & Sons, Inc.

Spencer, J. P., Vereijken, B., Diedrich, F. J., & Thelen, E. (2000). Posture and the emergence of manual skills. *Developmental Science, 3,* 216-233.

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: Are they really different? *Developmental Science, 6,* 378-391.

Thelen, E., Corbetta, D., & Spencer, J. P. (1996). Development of reaching during the first year: Role of movement speed. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1059-1076.

Thelen, E., Schoner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences, 24,* 1-86.

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

Turati, C., Sangrigoli, S., Ruel, J., & de Schonen, S. (2004). Evidence of the face inversion effect in 4-month-old infants. *Infancy, 6,* 275-297.

van Rijn, H., van Someren, M., & van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, *27*, 227-257.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences*, *6,* 376-382.

Werner, H. (1957). *Comparative psychology of mental development (revised ediition)*. New York: International Universities Press.

Wertsch, J. V. (1985). *Vygostky and the social formation of mind*. Cambridge, MA: Harvard Univeristy Press.

Wynn, K. (1998). Numerical competence in infants. In C. Donlan (Ed.), *The development of mathematical skills (*pp. 3-25). Hove, England: Psychology Press/Taylor & Francis.

Yoshida, H., & Smith, L. B. (2001). Early noun lexicons in english and japanese. *Cognition, 82*, B63-B74.

A Developmental Systems Theory Perspective on Psychological Change

Timothy D. Johnston[1] and Robert Lickliter[2]

[1]College of Arts and Sciences
University of North Carolina at Greensboro
Greensboro, NC 27402

[2]Department of Psychology
Florida International University
Miami, FL 33199

Authors Note: The writing of this chapter was supported in part by NICHD grant RO1 HD048423 awarded to RL.

A Developmental Systems Theory Perspective on Psychological Change

The two theories of psychological development that are the focus of this volume show how quantitative precision can be brought to bear on the problem of explaining psychological change. In any theoretical endeavor, precision often comes at the expense of generality and simplification – some features of the phenomena being explained must be ignored so as to create a more limited domain within which to achieve precision. Our aim in this chapter is to draw attention to some aspects of development that we believe are under-represented in both Dynamic Systems Theory (DST) and Connectionism. We intend this not as a criticism of either theoretical perspective but as a reminder that these aspects of development must be addressed in any eventual synthesis.

The approach we take is that of developmental systems theory, which shares with DST a focus on the whole organism in its environment and with Connectionism a focus on the biological support for psychological change. It adds to those perspectives a concern with more levels of biological analysis (including genetics), the incorporation of comparative data, and attention to evolutionary considerations in the analysis of psychological change. Although both DST and Connectionism have gone further than developmental systems theory in generating formal (mathematical) models of development and learning, developmental systems theory provides a broader context within which to pose questions about development and extends the analytic framework of the other two theories.

Nonetheless, there are broad compatibilities between developmental systems theory and the other theoretical positions under examination in this volume. In particular, all three perspectives reject the dichotomy between learned and innate features of the organism that has been such a persistent obstacle in developmental theory over the years and resist the tendency to ascribe features of behavior directly to genetic influences. We think the developmental systems perspective can enrich the accounts provided by DST and by Connectionism in the following ways: by explicitly including gene expression in the explanation of development; by broadening the domain of environmental influences on development to include social factors; and by connecting development with evolutionary change. The expanded and enhanced perspective on development provided by this account helps to situate the issues specifically addressed by DST and Connectionism in a broader theoretical context, and also illuminates a persistent problem for developmental theory, namely the relationships among the various time-scales over which change occurs.

1. The Developmental Systems Perspective

Developmental systems theory originated in opposition to accounts of development that assume a dichotomous relationship between genetic and environmental influences on development, such as the ethological theories of Konrad Lorenz (1937, 1965). Early architects of the theory include Lehrman (1953, 1970), Schneirla (1956), Bateson (1976), and Gottlieb (1976) and their writings have continued to exert an important influence on the field. Susan Oyama's *Ontogeny of Information* (1985) was a landmark synthesis of the motivating concerns of the theory, which has since been developed by a number of other authors (Oyama, Griffiths, & Gray, 2001; Robert 2004; West & King, 1987). Developmental systems theorists point out that all features of the organism – anatomical, physiological, behavioral, and psychological – emerge out of a set of interactions among a large number of factors. These include gene activity, hormones, social stimulation, food intake, and environmental temperature, among many others. Whereas some accounts of development focus on partitioning the organism's phenotypic characters among those that are genetically determined and those that are produced by the environment, developmental systems theory asserts that no such partitioning is possible, even in principle. Instead, the goal of developmental analysis should be to identify the ways in which various factors (including genetic and environmental ones) interact to produce all of the organism's phenotypic characters (e.g., Johnston & Edwards, 2002).

The specific factors that influence the development of any particular character of an organism (such as the mature song of a swamp sparrow or the perceptual abilities of a human infant) are determined by the ecological setting in which the organism normally develops. From a developmental systems perspective, understanding *what* influences are available in the ecological setting is critical to explaining *how* development occurs (Johnston, 1985; Lickliter, 2005). This ecological emphasis underlies another feature of developmental systems theory, namely its assertion of the continuity between developmental and evolutionary change. Because the course of development depends partly on the ecological factors to which the organism is exposed, changes in the phenotype of the mature organism may result from changes in those factors over long periods of time, as well as from alterations in the genetic makeup of the population resulting from natural selection, genetic drift, and other factors more usually considered in evolutionary analyses (Johnston & Gottlieb, 1990).

2. Resources for Development

Griffiths and Gray (1994) refer to the various factors that influence development as "developmental resources." These resources include all of the environmental factors shown to influence `development, as well as factors that exert their influence within the organism, such as hormones and the products of gene expression. This terminology was proposed by Griffiths and Gray to emphasize that none of the factors that affect development are privileged and to eliminate the implication that the genes, in particular, are uniquely capable of specifying or encoding aspects of the phenotype directly. As discussed in detail by many writers (Oyama, 1985; Johnston, 1987; Gottlieb, 1998; Lickliter & Honeycutt, 2003), the idea that the genes contain instructions specifying the construction of the phenotype (in the form of a plan or blueprint) has, in some form or another, been widespread in the literature on development for many years. Neither DST nor Connectionism espouses that nativist position and we will not reiterate the arguments against it here. Developmental systems theory also rejects the nativist position and is instead concerned with the ways and means by which the assembly of developmental resources is repeated within and across generations to produce behavioral phenotypes. Gene expression is, of course, a critical resource for behavioral development and is treated explicitly and in detail by developmental systems theory (Gottlieb, 1995; Johnston & Edwards, 2002; see further below).

The experience of the developing individual is another important resource for development. For example, children’s skills develop in large part on the basis of their experiences with the world, and both DST and Connectionism offer detailed accounts of the ways in which such experiences contribute to development. In many species, including humans in particular, the social group (including parents, siblings, and non-family members) is a very important source of experience for the developing individual. Parents provide experience for the young directly, for example by supplying food, warmth, and tactile stimulation, and also indirectly by modifying the young animal’s interactions with other features of its environment through mechanisms such as social facilitation, imitation, and scaffolding. The stabilizing and modifying roles of such factors both within and across generations are discussed further below.

Experience that constitutes a developmental resource for some behavioral or cognitive skill need not bear an obvious relationship to that skill. Much developmental theory focuses on cases in which skill development is based on direct experience with the skill (practice) or with the situation in which the skill is used (observation or imitation). For example, the entire field of learning theory assumes that a skill is learned as a function of experience with the task to which the skill is applied – Connectionism is explicitly based on matching functions between a skill such as word recognition and experience with speech sounds that include the words being learned. Developmental systems theorists have pointed out that not all relevant experience bears such a direct and obvious relationship to the skill or capacity under consideration. For example, Gottlieb (1997) has shown that the specific following response of a young duckling to the call emitted by its mother depends not on prior experience with the call, but on exposure to the vocalizations produced by the embryonic duckling itself in the days immediately preceding hatching. These embryonic calls are acoustically quite different from the maternal call and so this not an example of learning or familiarization. Exposure to the embryonic call is an essential, though non-obvious, resource for the normal development of the duckling's perceptual and behavioral specificity. Another example of such a non-obvious developmental resource is the importance of maternal licking and grooming of pre-weanling rat pups to their subsequent hormonal, physiological, and behavioral responsiveness to stress (Francis, Diorio, Liu, & Meaney, 1999). Pups receiving increased levels of maternal licking and grooming also show enhanced synaptogenesis in several brain regions (Liu, Diorio, Day, Francis, & Meaney, 2000) and a decrease in emotional reactivity to novelty in adulthood. These types of findings suggest that the developmental consequences of an organism's varied encounters with its environment are not always as obvious or straightforward as we might expect.

In the remainder of this chapter we will discuss some aspects of development that have not been much emphasized by either DST or Connectionism, but that are the focus of considerable attention within developmental systems theory: the role of gene expression, the contributions of social experience, and the relationship between developmental and evolutionary change. In a final section, we also address the challenge of how to deal with the fact that developmental change occurs on multiple time scales simultaneously.

3. Genes in developmental context

The idea that some behavior ("instinctive" or "innate") can be directly attributed to instructions provided by the genes is a long-standing feature of much developmental theory (Johnston, 1987, 2001). With the discovery of the genetic code by Watson and Crick in the middle of the last century, this idea seemed to gather greater weight and authority and the metaphor of the genetic blueprint, a set of precise and unequivocal instructions for the construction of the organism, gained wide currency. Numerous writers have criticized the blueprint metaphor and related ideas, and one important aim of developmental systems theory has been to replace this misleading but seductive metaphor with a more accurate account of genetic influences on development that incorporates our growing understanding of how genes make their essential contributions to the development of behavior.

Genes are segments of DNA that are transcribed in a complex process to produce RNA molecules, which move out of the cell nucleus to serve as templates for the construction (translation) of protein molecules. This process, from DNA to protein, is generally referred to as *gene expression*. It used to be thought that proteins are coded by contiguous stretches of DNA, arranged on the chromosome like beads on a string, but in fact, the picture has turned out to be far more complex. DNA consists of *exons* and *introns* that are initially transcribed sequentially into a primary RNA transcript. The intron sequences are then removed and discarded and the exons are spliced together into messenger RNA (mRNA) from which the protein molecule is translated. However, various kinds of alternative splicing can occur to produce different mRNA molecules from the same primary RNA transcript (Ast, 2005). Thus, the same stretch of DNA (the same gene, in conventional terms) can produce many different mRNA molecules, and hence many different proteins. The alternative splicing is carried out by a structure called the spiceosome and is regulated by specialized proteins produced within the cell in response to both intracellular and extracellular signals as development proceeds. This newly discovered one-to-many relationship between genes and proteins contradicts a long-standing dictum of molecular genetics ("one gene, one protein") and explains how the relatively small number of genes in the human genome (fewer than 25,000) can give rise to more than 90,000 different proteins.

This complex relation between genes and protein synthesis illustrates and underscores the complexity of the various roles that genes play in the development of behavior, and the relationships between the genes and other developmentally important factors. Figure 1 shows the place of the genes in the context of a developing organism, illustrating the influence of other factors, including experience (represented as sensory stimulation), the physical effects of the outer environment, and various internal elements such as individual cells and neural networks (Johnston & Edwards, 2002). As this depiction shows, genes do not affect behavior directly – genes are located within the nuclei of cells and only affect the development of behavior indirectly, through the regulated production of proteins and other gene products. Unlike many accounts of genetic contributions to development, this model shows that genes do not stand outside the developing system, acting on it as independent agents. Rather, they are embedded within the system, influencing its development through their effect on a complex cascade of interactions.

In the development of behavior, gene activity within the nervous system plays an especially important role. Genes, of course, exist within the nuclei of individual nerve cells, but nerve cells are organized into more or less complex neural circuits that control behavior. In explaining the neural control of behavior, it is the patterns of neural activity in those circuits that is most relevant. Figure 1 represents the nervous system in two different ways – as a collection of neural circuits that exhibit the patterns of activity controlling behavior and as a set of individual cells containing genes that help to regulate development. The dotted lines in the Figure show that the activity of individual cells is contained within the activity of the neural circuits that they make up. This allows us to depict two important features of behavioral development that are sometimes hard to keep in focus: although behavior is a global consequence of patterns of neural activity, the genetic contribution to change in those patterns (development) can only be effected within individual cells. We cannot explain the role of genes in development by postulating direct effects on complex neural circuitry, any more than on the behavior that those circuits control. Our models of gene action must recognize that genetic activity occurs within the nuclei of individual cells.

An important feature of the model depicted here is that all but the most transient effects of experience on behavior are mediated through molecular changes within individual nerve cells, including changes in gene activity. Notice that there is no arrow pointing directly from the box labeled “sensory stimulation” (the means by which experience is transduced by the nervous system) to that labeled “neural connectivity,” changes in which underlie the long-term, persistent changes in behavior involved in learning and development. In order for sensory experience to play a role in re-modeling neural circuitry (for example, by changing synaptic strengths), the activity in individual nerve cells that is produced by the experience must initiate biochemical events within the cell, including those that lead to changes in protein synthesis and gene activity.

Not all changes in protein synthesis involve modification in gene activity in the nucleus. A number of studies (see reviews by Steward & Schuman, 2001; Wang & Tiedge, 2004) have shown that translation of mRNA into proteins can occur in dendrites, close to synaptic sites. Both the transport of specific mRNA molecules from the nucleus and the synthesis of proteins in the dendrite have been shown to be regulated by neural activity, providing highly differentiated local control of the protein synthesis necessary for synaptic modification (Steward & Schuman, 2001, p. 314ff). Connectionism relies very heavily on changes in synaptic strength induced by experience to explain development. Those changes are produced by molecular events occurring both at the synapse and within the nucleus and developmental systems theory incorporates those events within an overall picture of developmental change.

As depicted in Figure 1, neural activity resulting from experience may also influence the transcription of genes in the cell nucleus. The recognition that experience produces long-term effects on behavior by modifying gene activity is contrary to any view that opposes experiential and genetic influences on behavioral change and reinforces the integrated perspective of developmental systems theory. A class of genes known as immediate-early genes (IEGs) has been shown to play a ubiquitous role in translating experience into changes in gene expression. Transcription of IEGs within specific neural structures typically follows sensory stimulation within a few minutes. For example, presentation of song results in transcription of the IEG *ZENK* (an acronym for four earlier names given to the same gene: *zif-*

268, egr-1, NGFI-A, and *Krox-24*) in parts of the songbird brain known to be involved in the perception and production of song (Ribeiro & Mello, 2000). *ZENK* has also been shown to be involved in variety of examples of neural and behavioral plasticity, such as long-term potentiation and associative learning (Knapska & Kaczmarek, 2004). The molecular pathways subsequent to *ZENK* transcription are still being worked out, but its product has been shown to regulate the transcription of synapsins (Thiel et al., 1994; Petersohn et al., 1995), which are part of the molecular structure of synapses. Presumably, such regulation plays a role in the modification of synaptic efficiency, which in turn produces the developmental changes in both neural activity and behavior explicated by connectionist models.

4. Social contributions to development

Developmental systems theory characterizes development as a situated, self-organizing, probabilistic process where pattern and order emerge and change as a result of complex interactions and relationships among developmental resources at multiple time-scales. From this perspective, what is inherited across generations is a structured developmental system that includes components internal (e.g., genes, cytoplasm, cells) and external (e.g., diet, light cycles, conspecifics) to the organism. Developmental systems theory recognizes that the recurrence from generation to generation of the specific resources and interactions that make up an organism's physical, biological, and social environments are as causally informative to the development and transmission of phenotypic traits as are genes contained within this system. A central concern for developmental systems theory is thus the identification of the developmental resources that allow for the repeated assembly of constructive interactions both within and across generations.

In this light, an ever-present feature of life for nearly all vertebrates is that the developing organism is situated in and interacts with some form of a social environment. Across a diverse array of species and ecologies, the typical circumstances of prenatal and postnatal development include conspecifics, especially parents and siblings. Like other aspects of the developmental system, these social companions are experiential resources for the developing individual and have been shown to play a significant role in the achievement of a wide range of species-typical behaviors, including food choice, predator avoidance, mate choice, and species identification (Galef & Laland, 2005; Gottlieb, 1997; Lickliter, 2005).

For example, research with several precocial bird species has consistently demonstrated that social experience with conspecifics both before and after hatching can have a significant effect on chicks' emerging perceptual preferences and early filial behavior (reviewed in Lickliter, Dyer, & McBride, 1993). In particular, the type, amount, and timing of sensory stimulation provided by the embryo or hatchling's social environment can significantly influence the young bird's arousal levels, perceptual processing, and learning and memory. For example, bobwhite quail chicks denied physical contact with broodmates during their prenatal development or reared with quail chicks of another species following hatching failed to prefer species-specific maternal auditory and visual cues at ages when normally socially reared chicks reliably demonstrate such filial preferences (Lickliter & Lewkowicz, 1995; McBride & Lickliter, 1993). Further, simply modifying the amount of tactile, auditory, or visual stimulation available from siblings in the period following hatching delays the emergence of perceptual responsiveness to maternal auditory and visual cues in quail chicks (Columbus and Lickliter, 1998).

Mallard ducklings denied physical contact and interaction with siblings following hatching likewise show altered patterns of auditory and visual responsiveness to their maternal hen. Gottlieb (1993) found that by manipulating the social stimulation available to ducklings immediately following hatching, he could affect both their arousal levels and their ability to learn non-conspecific maternal calls (in this case, a chicken maternal call). Specifically, he found that ducklings' malleability to learn to prefer the chicken maternal call over their own species-specific mallard maternal call was significantly affected by the presence or absence of tactile stimulation from siblings. Ducklings not permitted tactile contact with siblings slept less and were highly aroused before, during, and after exposure to recordings of the chicken call. In contrast, socially reared ducklings slept more and were less aroused than the isolates during exposure to the call. Ducklings that experienced normal levels of tactile contact with siblings appeared highly malleable, readily learning to prefer the non-conspecific chicken maternal call in the days following hatching, whereas those deprived of tactile contact with siblings did not. It is important to note that the presence of conspecifics is virtually guaranteed in the context of normal development for avian embryos and hatchlings, and early interaction with and stimulation from these social companions appears essential for the normal development of early perceptual and behavioral organization.

Additional evidence for the role of social factors as basic developmental resources is provided from research on the determinants of song learning in various species of songbirds. The presence of conspecifics that sing is a reliable and repeatable feature of the normal context of development for most songbirds and the opportunity to interact with social companions has been shown to influence the nature and course of song learning in a variety of oscine species. For example, West, King, and their colleagues have shown that juvenile and adult social experience with other cowbirds is a critical component of the development of male song patterns, female preference for male songs, and courtship and mating preferences in the brown-headed cowbird (e.g., West, King, & Arburg, 1988; West, King, & White, 2003; White, King, Cole, & West, 2002). Their work consistently shows that male and female cowbirds develop critical vocal and reproductive skills through recurring social interactions with their conspecifics. Further, male song patterns, female preference for songs, and courtship behaviors can be transmitted across generations (Freeberg, 1996). Young males that were captured from one geographic area and reared with adult males from another geographic area served as models for their own young, inducing them to behave like the males that had instructed their fathers. These types of changes in development that result in novel behavioral shifts that recur across generations can facilitate new organism-environment relationships. These new relationships can bring out latent possibilities for morphological, physiological, or further behavioral change. Changes in behavior driven by the physical or social environment can thus be the first step in creating new phenotypic variants on which natural selection can act (Gottlieb, 2002).

It is important to emphasize that social influences on development are not independent of the genetic influences discussed in the previous section. As already noted, exposure to song, a critical social factor in song development, also stimulates gene expression (e.g., Ribeiro & Mello, 2000) which is necessary for the changes in neural conductivity and connectivity that result in long-lasting behavioral change (i.e., learning of a species-typical song or song dialect). Indeed, we presume that any instance in which social experience produces a developmental change in behavior will involve changes in gene expression, as implied by Johnston and Edwards's (2002) model. Moffitt (2005) and Shanahan and Hofer (2005) provide overviews of recent studies that are beginning to identify the relationships between social experience and gene expression in human development and discuss various ways in which such studies could be extended.

5. The Relationship between Developmental and Evolutionary Change

In addition to including gene expression in explanations of development and broadening the domain of environmental influences on development to include social factors, developmental systems theory also addresses the relationship between developmental and evolutionary change. Developmental systems theory posits that development guides and constrains evolution in at least two major ways: First, it constrains phenotypic variation in such a way that the traits and characters presented to the filter of natural selection are not random. In other words, developmental processes and resources serve to limit the range of the possible in terms of phenotypic form. This is the *regulatory* function of development. Second, development provides variants in traits and characters that may lead to enduring alterations of the phenotype. This production of phenotypic novelties is the *generative* function of development and has significant implications for evolutionary change (Gottlieb, 2002; West-Eberhard, 2003). According to developmental systems theory, the regulatory and generative nature of development renders the natural selection of random mutations (the cornerstone of the neo-Darwinian view of evolutionary change) insufficient to account for evolution and its outcomes. Rather, the individual organism’s entire developmental system is the source of the stability *and* the variations in phenotype that researchers in evolutionary and developmental science seek to understand.

Given that all phenotypic traits arise during ontogeny as products of individual development, variations in the patterns and processes of development must be a primary basis for evolutionary change. Variations in morphologies, physiologies, and behaviors resulting from modification in developmental processes can place their possessors in different ecological relationships with their environments, and if these phenotypic variations provide slight advantages in survival and reproduction, then competitors without the novel phenotype will eventually decrease in frequency in the population. The novelty-generating aspects of the evolutionary process are thus the result of the developmental dynamics of living organisms, situated and competing in specific ecological contexts. From this view of evolution, genetic change is often a secondary or tertiary consequence of enduring transgenerational behavioral changes brought about by alterations in the organism's developmental system.

For example, a European passerine bird, the blackcap, has recently shown changes in its migratory behavior that appear to result in changes in mating behavior, size of egg clutches, and success at fledging young. At one time all European blackcaps migrated back and forth together, spending summers in northern Europe and the British Isles and winters in Portugal, Spain, and North Africa before gathering in mating grounds in southern Germany and Austria to breed. Over the last several decades, however, more and more blackcaps have remained north for the winter rather than heading south. Blackcaps were typically seen in the British Isles only during the summer months, but the number of them wintering in Britain and Ireland has soared over the last forty years. This change is likely due to the increased availability of winter provisioning provided by bird feeders, gardeners, landscapers, and other similar human activities. The resulting shift in migratory patterns has allowed northern wintering blackcaps to be exposed to the critical photoperiods that trigger migration and the onset of gonadal development some 10 days earlier than their southern wintering counterparts.

Even though all blackcaps continue to gather each year at the same mating sites in Germany and Austria, northern blackcaps typically arrive earlier at the breeding grounds and establish territories and mate with other earlier arriving birds; southern wintering blackcaps arrive at the same mating sites some two weeks later and are more likely to mate with each other, serving to reproductively isolate northern wintering birds from the later-arriving southern wintering population. Importantly, the shift in migratory pattern seems to confer an advantage to the northern blackcaps, who tend to lay about one more egg per season than do the late arrivals from the south (Bearhop et al, 2005).

The blackcap provides a compelling example of how a change in behavior (in this case, a change in migratory patterns brought on by changes in food availability) can lead to changes in the timing of breeding, which in turn can lead to the effective reproductive isolation of populations and ultimately, divergence and even sympatric speciation. Contrary to the assumptive base of the neo-Darwinian synthesis of the last century, the introduction of phenotypic variation upon which natural selection acts is not simply the result of random genetic mutations. Rather, variations in phenotypes and the resulting opportunities for evolutionary change are the result of a wide range of epigenetic processes occurring at different time scales and involving factors internal *and* external to the developing individual.

6. Time Scales of Change

The expanded perspective on development and evolution provided by developmental systems theory situates the issues addressed by Dynamic Systems Theory and Connectionism in a broader theoretical context and also highlights a persistent challenge for developmental theory, making sense of the relationship among the various time-scales over which change occurs. These time-scales include *real-time*, the immediate experiences and encounters of the individual with its physical, biological, and social environments, *developmental time*, the continuing influence of prior experiences and encounters on the individual's ongoing interaction with its varied environments, and *evolutionary time*, the transgenerational effects of the individual's experiences and activities during its ontogeny. Characterizing how real-time and developmental time combine to contribute to developmental change has been a primary focus of concern for both the DST and Connectionism approaches represented in this volume. Developmental systems theory highlights the additional complexity of evolutionary time, raising the important issue of how real-time and developmental time also combine with evolutionary time to contribute to evolutionary change.

As reviewed earlier, modifications in experience can initiate a range of physical and behavioral novelties, some of which can (given the availability of appropriate developmental conditions) persist across subsequent generations. Dramatic findings from developmental psychobiology indicate that modified early experience in one generation can predictably influence phenotypic outcomes in subsequent generations, even in the absence of the original experiential modification. For example, differences in physical (body weight, endocrine responses) and behavioral (fearfulness) measures have consistently been observed between groups of rats whose mothers (Denenberg & Whimby, 1963; Francis, Diorio, Liu, & Meaney, 1999; Whimbey & Denenberg, 1967) or grandmothers (Denenberg & Rosenberg, 1967) were handled or not handled as infants. Despite its obvious importance to both developmental and evolutionary concerns, these types of transgenerational effects on both physiological responsiveness (in particular, the development of the hypothalamic adrenocortical system) and behavioral responsiveness (including curiosity, novelty seeking, and emotional regulation) remain poorly understood. Once again, we emphasize that both experiential and genetic factors are involved in mediating transgenerational effects such as these. Maternal behavior has been shown to influence gene expression, as measured by increases in mRNA coding for proteins involved in behaviors known to be affected by differences in maternal care, such as stress response and spatial learning (reviewed by Meaney, 2001). These findings support the message of developmental systems theory that experiential and genetic influences are co-actors in developmental change, including that which extends across generations.

Developmental systems theory enriches the accounts of development provided by DST and Connectionism by emphasizing the importance of a transgenerational approach to the study of behavior, in which the development of one generation is studied in relation to the experiences and developmental resources of previous generations (Honeycutt, in press). Knowledge of the dynamics of developmental processes at multiple time-scales can lead to a deeper understanding of the ways and means of both developmental and evolutionary change.

7. Conclusion

Considering the number of different resources, factors, and interactions that must be taken into account in order to explain developmental change in behavior, and the variety of time scales over which that change occurs, it is not surprising that any theoretical explanation of development will lack either completeness or precision, and frequently both. Both Connectionism and Dynamic Systems Theory have gone further than most theories of development in providing quite precise explanations of certain developmental phenomena. In this chapter our aim has been to indicate some aspects of development that are not currently addressed by either theory but that will eventually need to be encompassed as our theories of development become more complete.

References

Ast, G. (2005). The alternative genome. *Scientific American, 292(4)*, 58-65.

Bateson, P. P. G. (1976). Specificity and the origins of behavior. *Advances in the Study of Behavior*, *6*, 1-20.

Bearhop S., Fiedler, W., Furness, R., Voter, S., Waldron, S., Newton, J., Bowen, G., Berthold,, P., & Farnsworth, K. (2005). Assortative mating as a mechanism for rapid evolution of a migratory divide. *Science, 310*, 502-504.

Columbus, R.F. & Lickliter, R. (1998). Modified sensory features of social stimulation alters the perceptual responsiveness of bobwhite quail chicks. *Journal of Comparative Psychology, 112*, 161-169.

Denenberg, V. H., & Rosenberg, K. M. (1967). Nongenetic transmission of information. *Nature, 216*, 549-550.

Denenberg, V.H., & Whimbey, A. E. (1963). Behavior of adult rats is modified by the experiences their mothers had as infants. *Science, 142*, 1192-1193.

Francis, D., Diorio, J., Liu, D., & Meaney, M. J. (1999). Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science, 286*, 1155-1158.

Freeberg, T. M. (1996). Assortative mating in captive cowbirds is predicted by social experience. *Animal Behaviour, 52*, 1129-1142.

Galef, B. G., & Laland, K. N. (2005). Social learning in animals: Empirical studies and theoretical models. *BioScience, 55*, 489-500.

Gottlieb, G. (1976). The roles of experience in the development of behavior and the nervous system. In Gottlieb, G. (Ed.), *Studies on the development of behavior and the nervous* *system, Volume 3: Neural and behavioral specificity* (pp. 25-54). New York: Academic Press.

Gottlieb, G. (1993). Social induction of malleability in ducklings: Sensory basis and psychological mechanism. *Animal Behaviour, 45*, 707-719.

Gottlieb, G. (1995). Some conceptual deficiencies in 'developmental' behavior genetics. *Human Development. 38*, 131-141.

Gottlieb, G. (1997). *Synthesizing nature-nurture: Prenatal roots of instinctive behavior*. Mahwah, NJ: Erlbaum Assoc.

Gottlieb, G. (1998). Normally occurring environmental and behavioral influences on gene activity: From central dogma to probabilistic epigenesis. *Psychological Review, 105,* 792-802.

Gottlieb, G. (2002). Developmental-behavioral initiation of evolutionary change. *Psychological Review, 109*, 211-218.

Griffiths, P. E., & Gray, R. D. (1994). Developmental systems and evolutionary explanation. *Journal of Philosophy, 91*, 277-305.

Honeycutt, H. (in press). Studying evolution in action: Foundations for a transgenerational comparative psychology. *International Journal of Comparative Psychology*.

Johnston, T. D. (1985). Environmental constraints and the natural context of behavior: Grounds for an ecological approach to the study of infant perception. In: G. Gottlieb, & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview,* (pp. 91-108). Norwood, NJ: Ablex.

Johnston, T. D. (1987). The persistence of dichotomies in the study of behavioral development. *Developmental Review, 7*, 149-182.

Johnston, T. D. (2001). Towards a systems view of development: An appraisal of Lehrman's critique of Lorenz. In: S. Oyama, P. E. Griffiths, & R. D. Gray (Eds.), *Cycles of contingency: Developmental systems and evolution* (pp. 15-23). Cambridge, MA: MIT Press.

Johnston, T. D. and Edwards, L. (2002). Genes, interactions, and the development of behavior. *Psychological Review*, *109*, 26-34.

Johnston, T. D. and Gottlieb, G. (1990). Neophenogenesis: A developmental theory of phenotypic evolution. *Journal of Theoretical Biology, 147*, 471-495.

Knapska, E. and Kazmarek, L. (2004). A gene for neuronal plasticity in the mammalian brain: Zif268/Egr-1/NGFI-A/Krox-24/TIS8/ZENK. *Progress in Neurobiology, 74,* 183-211.

Lehrman, D. S. (1953). A critique of Konrad Lorenz's theory of instinctive behavior. *Quarterly Review of Biology, 28,* 337-363.

Lehrman, D. S. (1970). Semantic and conceptual issues in the nature-nurture problem. In: L. R. Aronson, E. Tobach, D. S. Lehrman, & J. S. Rosenblatt (Eds.), *Development and evolution of behavior* (pp. 17-50). San Francisco: W. H. Freeman.

Lickliter, R. (2005). Prenatal sensory ecology and experience: Implications for perceptual and behavioral development in precocial birds. In: P. Slater, J. Rosenblatt, C. Snowden, T. Roper, H. J. Brockmann, & M. Naguib (Eds.), *Advances in the Study of Behavior,* Vol. 35. New York: Academic Press.

Lickliter, R., & Honeycutt, H. (2003). Developmental dynamics: Towards a biologically plausible evolutionary psychology. *Psychological Bulletin, 129,* 819-835.

Lickliter, R., & Lewkowicz, D. J. (1995). Intersensory experience and early perceptual development: Attenuated prenatal sensory stimulation affects postnatal auditory and visual responsiveness in bobwhite quail chicks. *Developmental Psychology, 31*, 609-618.

Lickliter, R., Dyer, A. B., & McBride, T. (1993). Perceptual consequences of early social experience in precocial birds. *Behavioral Processes, 30*, 185-200.

Liu, D., Diorio, J., Day, J.C., Francis, D. D., & Meaney, M. J. (2000). Maternal care, hippocampal synaptogenesis and cognitive development in rats. *Nature Neuroscience, 3*, 799-806.

Lorenz, K. Z. (1935). Der Kumpan in der Umwelt des Vogels. *Journal für Ornithologie, 83*, 137-213, 289-413.

Lorenz, K. Z. (1965). *Evolution and modification of behavior*. Chicago: University of Chicago Press.

McBride, T. & Lickliter, R. (1993). Social experience with siblings fosters species-specific responsiveness to maternal visual cues in bobwhite quail chicks. *Journal of Comparative Psychology, 107*, 320-327.

Meaney, M. J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience, 24*, 1161-1192

Moffitt, T. E. (2005). The new look of behavioral genetics in developmental psychopathology: Gene-environment interplay in antisocial behaviors. *Psychological Bulletin, 131*, 533-554. 2005

Oyama, S. (1985). *The ontogeny of information: Developmental systems and evolution.* Cambridge: Cambridge University Press.

Oyama, S., Griffiths, P. E., & Gray, R. D. (Eds.) (2001). *Cycles of contingency: Developmental* *systems and evolution*. Cambridge, MA: MIT Press.

Petersohn, D., Schoch, S., Brinkmann, D. R., & Thiel, G. (1995). The human synapsin II gene promotor: Possible role for the transcription factor zif268/egr-1, polyoma enhancer activator 3, and AP2. *Journal of Biological Chemistry, 270,* 24361-24369

Ribeiro, S., & Mello, C. V. (2000). Gene expression and synaptic plasticity in the auditory forebrain of songbirds. *Learning and Memory, 7,* 235-243.

Robert, J. S. (2004). *Embryology, epigenesis, and evolution*. Cambridge, UK: Cambridge University Press.

Schneirla, T. C. (1956). Interrelationships of the "innate" and the "acquired" in instinctive behavior. In: P. P. Grassé (Ed.), *L'Instinct dans le comportement des animaux et de l'homme,* (pp. 87-452). Paris: Masson.

Shanahan, M. J., & Hofer, S. M. (2005). Social context in gene-environment interactions: Retrospect and prospect. *Journals of Gerontology, 60B*, 65-76.

Steward, O., & Schuman, E. M. (2001). Protein synthesis at synaptic sites on dendrites. *Annual Review of Neuroscience, 24*, 299-325.

Thiel, G., Schoch, S., & Petersohn, D. (1994). Regulation of synapsin 1 gene expression by the zinc finger transcription factor zif268/egr-1. *Journal of Biological Chemistry, 269,* 15294-15301.

Wang, H. D., & Tiedge, H. (2004). Translational control at the synapse. *Neuroscientist, 10,* 456-466.

West, M. J., & King, A. P. (1987). Settling nature and nurture into an ontogenetic niche. *Developmental Psychobiology, 20,* 549-562.

West, M. J., King, A. P., & Arburg, A. A. (1988). An inheritance of niches: The role of ecological legacies in ontogeny. In: E. M. Blass (Ed.), *Developmental psychobiology and behavioral ecology*, (pp. 41-62). New York: Plenum.

West, M. J., King, A. P., & White, D. J. (2003). The case for developmental ecology. *Animal Behaviour, 66*, 617-622.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution.* Oxford, UK: Oxford University Press.

Whimbey, A. E., & Denenberg, V.H. (1967). Experimental programming of life histories: The factor structure underlying experimentally created individual differences. *Behaviour, 29,* 296-314.

White, D. J., King, A. P., Cole, A., & West, M. J. (2002). Opening the social gateway: early vocal and social sensitivities in brown-headed cowbirds. *Ethology, 108***,** 23-37.

Figure 1: Illustration of the various factors that interact in the course of behavioral development. From Johnston & Edwards (2002; Figure 3). Copyright 2002, American Psychological Association. Reproduced by permission.