Cognitive Development, 12, 563-586 (1997)
ISSN 0885-2014

ESSAY REVIEW

RETHINKING INNATENESS, LEARNING, AND CONSTRUCTIVISM: CONNECTIONIST PERSPECTIVES ON DEVELOPMENT

Thomas R. Shultz
McGill University

Denis Mareschal
Exeter University

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: MIT Press.

Plunkett, K., & Elman, J.L. (1997). *Exercises in rethinking innateness: A handbook for connectionist simulations.* Cambridge, MA: MIT Press.

One of the longest standing and most difficult issues in developmental psychology is that of transition mechanisms. How is it that the child develops from one mode or stage of functioning to another? Although the issue of transition is typically considered to be one of the two major issues in development, it has been estimated that only one percent of the literature addresses transition, the bulk of the literature focusing on the companion issue of what develops in children's cognition (Sternberg, 1984).

The book *Rethinking Innateness: A Connectionist Perspective on Development*, written by six of the major researchers in the area of connectionist approaches to development, is a major contribution to the issue of transition mechanisms in language and cognitive development. The work admittedly does not present a theory but rather a framework or set of conceptual tools for thinking about transitions in development. The title of our review deliberately modifies Elman et al.'s title,

Preparation of this paper was facilitated by a grant to the first author from the Natural Sciences and Engineering Research Council of Canada and by a grant to the second author from the Exeter University Research Fund. We are grateful to Paul Quinn and Yoshio Takane for comments on earlier drafts.

Direct all correspondence to: Thomas R. Shultz, Department of Psychology, McGill University, 1205 Penfield Avenue, Montreal, Quebec H3A 1B1 <shultz@psych.mcgill.ca>.

Manuscript received July 1, 1997; revision accepted August 3, 1997

adding the contrasting terms "learning" and "constructivism," and extending the discussion to "multiple connectionist perspectives." Their book covers learning and innateness in great detail, but as our review makes clear, it avoids serious consideration of constructivist ideas and alternative connectionist perspectives that are suggested by constructivism.

Our review begins with some background to situate the work in the historical and contemporary developmental literature. Then we focus on some of the principal contributions of the book, present a critique based primarily on what the book ignores, and attempt some degree of synthesis between competing connectionist approaches to development, integrating what the book covers and what it leaves out. Finally, we comment on the companion volume of computer exercises.

ISSUES OF TRANSITION AND INNATENESS IN PSYCHOLOGICAL DEVELOPMENT

Elman et al.'s topic of transition and innateness can be traced back to the beginnings and even to the pre-history of developmental psychology. Within the field of philosophy of mind, concern with the origin of knowledge polarized into rationalism vs. empiricism. Rationalism, the view that knowledge derives from the activity of the mind itself, can be traced as far back as Plato around 400 BC. Empiricism, the contrasting view that knowledge derives from sensory experience might be traceable to the Greek Stoics of around 300 BC, but it was certainly a central feature of the writings of Locke in the 17th century and Hume in the 18th century.

It is well known that both of these views on the origin of knowledge had their adherents throughout early scientific psychology and into the contemporary literature. Rationalism, over the years, became nativism, with its emphasis on innate determinants of development in language (Chomsky, 1971, 1975, 1980; Lenneberg, 1967; Pinker, 1994), cognition (Carey, 1991; Fodor, 1980; Spelke, 1994), and other aspects of development (Gesell, 1929). Often the thrusts of nativist arguments were indirect, noting the difficulty or impossibility of learning, with implications that knowledge representations must be innately determined. Detailed consideration of how behavior or psychological capacity actually emerges from the genotype was typically absent.

Empiricism became behaviorism, with its emphasis on learning from contingencies available in the environment. Behavioristic learning theory was formulated by Hull (1943) and has had many proponents in developmental psychology throughout the years (Bandura, 1977; Kendler & Kendler, 1959; Lipsitt, 1966; Spiker, 1963; White, 1963), but is perhaps less popular in contemporary work as an explicitly formulated theory. The difficulties that learning theories had explaining the basic facts of language acquisition (Chomsky, 1959; Pinker, 1994) probably account for much of its lost appeal. An aspect of classical learning theory that has had considerable staying power is Hebb's (1949) work on learning in cell-assemblies. The so-called Hebbian learning rule, which is still prominent in contemporary connectionism, specifies that links between neurons are strengthened in proportion to their simultaneous activity. In general, Hebb emphasized the importance of internal processes in distinction to Hull's focus on observable behavior. Connectionism derives more from the Hebbian tradition rather than the Hullian tradition.

Although the basic and long-standing debate between nativism and learning may have reached cliché status in psychology, the resolution proposed by Kant in philosophy is less well known to many psychologists. Kant's so-called constructivist resolution emphasized that both rationalists and empiricists were only partly correct in their accounts of knowledge acquisition. He argued that empiricists were correct in their claim that knowledge arrives through sensory impressions, and that rationalists were correct in their claim that knowledge must be processed by internal mental structures in order to be understood. As this processing occurs, initial mental structures become modified, and new structures are thereby constructed. The principal heir within psychology to Kant's resolution of the epistemological debate was Piaget (Piaget & Inhelder, 1969) who emphasized the construction of mental schemata in response to challenging experience, and the eventual assimilation of experience to these newly constructed schemata. Neo-Piagetians can be found in the contemporary literature on cognitive development (Case, 1985, 1992; Fischer, 1980; Pascual-Leone, 1970), but are less visible in the area of language development.

Elman et al. enter this venerable debate on the nature of transition mechanisms by doing some serious thinking about what it might mean for cognitive or linguistic abilities to be innate and how researchers might use neural network modeling to explore the ways these various forms of innateness might interact with learning. In the next section, we summarize and comment on some of the principal contributions that their book offers.

PRINCIPAL CONTRIBUTIONS

The contributions made by this book are many and varied. They offer a fresh perspective on developmental transitions. Here we stress demonstrations of the value of computational modeling; the importance of latent environmental structure; the eventual ossification of learning; what it means to be innate; arguments against representational innateness; the value of development; relations of psychological development to the connectionist, brain development, and dynamic systems literatures; the ubiquity of interactions in development; and suggestions for future research.

The Value of Simulation

Like other psychologists who have employed computational modeling, Elman et al. stress the value of such modeling and present many examples of its utility. In particular, modeling makes theorizing more rigorous and forces a level of precision that typically reveals inconsistency and incompleteness in theories. Building and testing computational models is indeed an excellent way to formulate and test psychological theory, as Elman et al. illustrate with reasonably detailed presentation or review of some 18 different models ranging from visual tracking of objects and imprinting to language acquisition and problem solving.

The particular type of modeling favored by Elman et al. is connectionist, also known as parallel distributed processing or neural network modeling. Such models typically operate below the level of symbols and rules favored by many other modelers (Anderson, 1993; Ling & Marinov, 1993; Newell, 1990; Klahr, 1989) and conform in a general way to known principles of brain functioning. These artificial neural networks are composed of simple, neuron-like units that change their activity levels (activations) moment to moment. The units are connected to each other by connection weights that are modifiable in size and sign during learning. As activation is passed among the units, governed by both the sending activations and the connection weights, such networks are considered to simulate cognition. Given an input (represented by a pattern of activation levels on a bank of input units) reflecting an instance of the type of problem on which a network has been trained, the network produces a response (represented by a pattern of activation levels on a bank of output units) that is interpreted as a judgment or a prediction.

The models featured by Elman and his colleagues are typically feedforward networks in which activation passes in a forward direction from one layer to the next. Occasionally, they discuss feedforward networks with a set of recurrent connections from an intermediate layer (known as hidden units) back to the input units. This recurrent network topology (architecture), popularized by Elman (1990), is useful for dealing with problems like sentence processing in which a network analyzes a sequence of items and keeps track of where each item is in the sequence. All of the network topologies covered in the book are of the static type, meaning that they are designed by hand and do not change as learning proceeds. This point becomes important when we discuss the implications of constructivism and alternative models that do change their topologies during learning.

Learning is typically accomplished in these networks with the so-called backpropagation algorithm. Backpropagation is probably the most commonly used connectionist algorithm for learning in multi-layered feedforward networks. In brief, it alters connection weights in order to minimize the discrepancy (error) between actual and target output activations, and propagates this error back through the layers of the network to adjust connection weights deep within the network. So, in these networks, activation flows forward, and error signals flow backward. Some alternative learning algorithms are discussed, but backpropagation is used in most of the featured models. Although the book does well at presenting a verbal description of backpropagation learning, the attempt at a more formal, mathematical presentation falls a bit short of what deeply curious readers may want to know. A good, next source for such readers is Hinton (1992). There are several other detailed presentations of backpropagation (Anderson, 1995; Hertz, Krogh, & Palmer, 1991; Rumelhart, Hinton, & Williams, 1986). The accompanying volume of exercises (Plunkett & Elman, 1997) also provides a more detailed presentation of the mathematics of backpropagation.

At first glance, artificial neural networks might seem to be a strange, even inappropriate choice for modeling innate characteristics. After all, these networks develop by learning, and much theoretical discussion of innateness focuses on symbolic rules, as opposed to connections (Pinker, 1994). The answer from Elman et al. is that neural networks can be used to implement innate features and thus explore the pervasive interactions between innateness and learning. Interesting examples of such interactions are presented in the book, although some of the best work on interactions between evolution and learning in networks (Nolfi, Elman, & Parisi, 1990; Parisi, Cecconi, & Nolfi, 1990) is curiously omitted. That work shows that it is possible to model evolutionary principles as neural networks reproduce over generations and that their evolutionary adaptiveness is greatly enhanced by learning. Nonetheless, Elman et al. do effectively illustrate the utility of computational modeling in general and connectionist modeling in particular.

In addition to the increased precision that comes with such modeling, it becomes possible to examine the insides of networks to determine what essentially they have learned and how that knowledge is represented. Although neural networks are notoriously more opaque in this respect than symbolic rule based models, such analysis is possible and fruitful as illustrated by several of Elman's models. Unfortunately, these network analysis techniques are still somewhat rare in connectionist modeling. They will likely prove to be essential in linking connectionist models to more conventional psychological theory. It is not enough to know that a network can learn and simulate psychological development; it is important to discover how it manages to perform as it does this and whether its emerging knowledge representations match those of children. Indeed, many of the models featured in this book could have benefited from deeper analysis of this kind.

Latent Environmental Structure

Notwithstanding the importance of innate constraints, a very important contribution of the connectionist movement has been to demonstrate that there is an abundance of latent structure in the environment that can be acquired by general purpose learning algorithms such as backpropagation. This point is well documented by Elman et al. in examples ranging from selectivity in visual neurons, to reading, to various aspects of language acquisition, including morphology, word meaning, and syntax.

Additionally, it is worth stressing that these complex latent structures in the environment could probably not be learned by traditional associative mechanisms such as classical conditioning or reinforcement. Single, direct connections between stimuli and responses, modifiable by reinforcement principles, would not possess sufficient computational power for such learning. Neural networks with their many units arranged on separate layers, non-linear activation functions, sophisticated learning rules, and informationally rich output targets are capable of learning complex functions that would elude less powerful learning mechanisms. Again, it is well to remember that connectionism has more in common with Hebbian than with Hullian or Skinnerian learning.

Ossification of Learning

One of enduring mysteries of development is that of ossification, sometimes seen as the opposite of plasticity. Why does learning (or further development) become more difficult as individuals mature? Such questions become important in consideration of critical or sensitive periods, during which learning is particularly effective. Maturational accounts have tended to explain ossification in terms of the genetically controlled reduction of learning capacity. In contrast, Elman et al. argue that ossification could be the result of learning itself. As a network learns, it gradually commits itself to particular solutions until it is no longer sufficiently flexible to undertake new learning. A well known result of ossification is that of limited recovery from brain damage.

As an example, Elman et al. cite the interesting simulations of Marchman (1993) who trained networks in grammatical development and then subjected them to lesions, by eliminating from 2% to 44% of the connection weights at different points in learning. She found that networks with small and/or early lesions were able to recover with further training. However, networks with large, later lesions failed to recover. The message seems to be that a network's learning capacity can be used up by learning itself.

Emergence of Functional Modularity

The presence of functional modules in the brain has often been presented as strong evidence for a nativist stance (Fodor, 1983). We believe that one of the strongest contributions of the book is to suggest that functional modules can emerge as part of the developmental process. There is a difference between starting with distinct functional modules and developing functional modules through interactions between the architectural or learning algorithm constraints and the environment. Jacobs, Jordan, and Barto (1991) showed how object identification would be naturally decomposed into "what" and "where" tasks in a system where multiple networks with different architectures competed to learn about objects. Differences in functionality emerged as an efficient way to split up a complex task among multiple networks. Similarly, Mareschal, Plunkett, and Harris (1995) described a model of infant visual tracking in which functional modules emerge as a result of different learning algorithms in different parts of the network. Some weight learning algorithms are better attuned to extracting temporal correlations from the environment whereas other weight learning algorithms are better attuned to extracting spatial correlations. As a result, this model develops one pathway to process spatial-temporal object information and a second pathway to process surface feature information.

Three Ways to be Innate

Arguably the book's most important contribution is its analysis of the different ways that some ability can be considered to be innate. Elman et al. argue that innateness can occur in representations, architectures, or timing.

Representations. The classic, default view is that innateness occurs at the level of representations. The basic idea of representational innateness is that children possess domain specific knowledge controlled by the genotype. Recent proposals have included innate knowledge of syntax (Pinker, 1994), number (Wynn, 1992), and physics (Spelke, 1994). The neural network analog to representational innateness would be to have many or all of a network's connection weights specified before learning starts.

Elman et al. come out very strongly against representational innateness, arguing that there is insufficient information in the human genotype for this to be feasible. Never mind psychological concepts, but the molecular parts of the body itself cannot even be fully specified in the genotype. They cite evidence that the human body contains 5×10^{28} bits of molecular information, but the human genotype contains only 10^5 bits of information. The implication, not well worked out in the book, is that the human genotype does not possess enough information to serve as a blueprint for possible innate aspects of language and cognitive development. Evidence for the initial equipotentiality of mammalian cortex, discussed later, is also used by Elman et al. to discredit representational innateness. If a part of the cortex does not initially know its eventual job, and can easily be recruited for other jobs, then how could its domain specific content be innately specified?

It is noteworthy that the network genotypes in the Nolfi et al. (1990) simulations on evolution and learning contained initially specified, informationally laden weights. Perhaps that is why Elman et al. chose to ignore that otherwise interesting work in the book—it effectively implemented the dreaded representational innateness.

Architectures. A second way for something to be innate is in terms of architectural constraints. Elman et al. further break architectural constraints down into unit, local, and global constraints. At the unit level would be features like firing thresholds, transmitter types, and learning rules. Connectionist analogs of unit constraints would be activation functions, learning rules, and the parameters of learning rules. At the local level, one could consider number of layers of neurons, connection density, and circuitry. Analogs in artificial neural networks would be network topologies and numbers of layers and units. And at the global level, there are connections between brain regions to consider. These could be implemented in connectionist systems by network modules that may have different, specialized jobs and that feed into other modules. Currently, such architectural decisions are made by the researchers and implemented by hand, except for the case of some generative algorithms, which we discuss later.

Timing. The third, and even more indirect, way for something to be innate is in terms of the timing of related events in development. An example is the evidence on spatio-temporal waves of cortical development. The locus of maximum neural plasticity begins in the primary sensory and motor areas, followed by the secondary association areas, and finally in the frontal areas (Thatcher, 1992). When developed and connected, these regions act as successively higher level filters of incoming information, from primary to secondary areas, and on to frontal areas.

Elman et al. discuss a simulation of these spatio-temporal cortical waves by Shrager and Johnson (in press). Their model consists of a 30 by 30 matrix of neuronal units, each unit having two-way connections with neighboring units and input from two external units. Random inputs of 0 and 1 from these two external units are provided and connection weights within the matrix are adjusted according to a Hebbian learning rule: connection weights increase between units that are simultaneously active. Under these conditions, some units become responsive to positive input from one of the external input units, others become responsive to input from both external inputs (conjunction), and still others become responsive to input from one or the other of the external inputs (disjunction). The main point though, of the simulations, was to study the effects of an external, "trophic" wave on knowledge development. This wave ensured that columns underneath it were more plastic and thus more able to learn than other columns. As the wave passed gradually from left to right, the network was observed to develop increasingly complex functions, including conjunction, disjunction, and exclusive-or.[1] Units on the left become fixed in learning simple functions (such as the presence or absence of a single input), but those units to their right are able to learn more complex functions by re-representing the output of the units to the left. The fact that the network can learn more complex functions, such as exclusive-or, under this trophic wave is important and suggests that greater cognitive complexity can be achieved by starting with simpler ideas.

Yet another example of timing effects is Elman's (1993) starting small effect with syntactic acquisition. His recurrent backpropagation network attempts to learn to predict the next word in English-like sentences with varying degrees of embedding. Although it is impossible to predict the exact next word under these circumstances, it might be possible to predict its syntactic category, by increasing activations for all known words of a particular category such as *noun*. Interestingly, Elman found that his network could not learn to predict syntactic categories unless it started with relatively simple sentences and gradually proceeded to more complex ones. He found that this could be implemented in two different ways—by staging the input from the environment (as in motherese) or by increasing the network's active memory capacity. Small active memory capacity ensures that only short, simple sentences can be processed; larger active memory capacity enables

[1]The exclusive-or function yields a value of *true* only in case one or the other of two input units is on, but not when both input units are on or when both are off.

the processing of longer, more complex sentences. Although staging the environment is an environmental constraint, increasing active memory could be considered as a maturational constraint. As in the Shrager and Johnson simulation, the implication is that it may help in learning complex material to start small, with simpler material. Complex learning can build on the simpler learning that precedes it.

On the cautionary side, it should be noted that Elman's networks do not learn a grammar in the sense that children do. His networks cannot, for example, generate grammatical sentences or make judgments about which sentences are grammatical and which are not. It only predicts, and weakly at that, the syntactic category of the next word in a sentence. Even so, it provides a good example of the importance of maturational timing constraints in development.

Why Have Development?

A very basic question that is rarely asked in developmental psychology is "Why should there be any development?" Why does our species not produce fully functioning, mature organisms at birth as some species do? Elman et al. focus on the psychological aspects of this question and identify a number of advantages for having a relatively long period of immaturity and development. In their view, development

1. Preserves maximal plasticity for dealing with internal damage or environmental change
2. Allows for phenotypic exploration of new solutions through learning
3. Allows time-dependent interactions to occur, and
4. Naturally decomposes difficult problems into simpler ones.

All of these points are illustrated to some degree in neural network simulations discussed in the book. This is a good example of how such simulations encourage both the asking and the answering of basic questions about psychological development.

Interesting Reviews

The Elman et al. book contains interesting reviews of several areas related to psychological development. These include literatures on neural networks, brain development, and dynamic systems.

Neural Networks. The review of neural networks provides a comprehensive and fairly painless introduction for novices. It covers units and connection weights; activation functions; some of the most common learning algorithms including the Hebbian rule, the delta rule,[2] and backpropagation; recurrence; modularity; and the

---

[2]The delta rule (called the perceptron convergence prodedure by Elman et al.) governs learning at one level of connection weights. Like backpropagation, it reduces the discrepancies between outputs and targets, but there is no need for propagation back to eawrlier levels.

distinction between supervised and unsupervised learning. This review provides a coherent and interesting introduction that could lead the reader to other, more technical sources (Anderson, 1995; Hertz et al., 1991; Rumelhart et al., 1986). Hertz et al. is an excellent source for a formal treatment of connectionism.

The three Rumelhart and McClelland volumes, which appeared in 1986 and came to be known at least around our laboratory as the PDP series, provide the basics in a less formal manner than Hertz et al. and ground the basics in various psychological issues. This is the classic three volume series that rekindled interest in connectionist modeling, and was no doubt an inspiration for the Elman et al. effort. Whether the Elman et al. series (also scheduled for three volumes) will be as successful as the original is doubtful if only because the focus here is on development whereas the focus of the original PDP series was much broader. Still, the Elman et al. series could do for developmental psychology what the PDP series did for cognitive psychology a decade ago, that is, to make neural network modeling accessible to a much wider range of practitioners.

Brain Development. The extensive review of brain development is extremely interesting, albeit highly selective and somewhat controversial. Considerable neuropsychological evidence is marshaled here against representational innateness. The idea that the genotype contains a detailed blueprint for cortical functioning is difficult to maintain against evidence that the cortex is initially relatively equipotential and plastic. Some of the most impressive evidence in this respect can be summarized under the maxims "When in Rome do as the Romans do" and "You are what you eat." Compelling demonstrations come from experiments with vertebrate animals that either transplant pieces of fetal cortex from one area to another, or redirect thalamic inputs from their usual targets somewhere else in the cortex. Under such abnormal conditions, the cortex takes on properties of the area that it is now in ("When in Rome do as the Romans do") or those of the input it receives ("You are what you eat"). It is as if auditory cortex becomes able to see and visual cortex becomes able to hear.

Data on how the human brain reorganizes itself in response to brain damage at various points in development and how language is organized in atypical populations such as people with Williams Syndrome and Downs Syndrome are reviewed to illustrate similar points. Basically, brain damage or deficits can be countered when healthy tissue (normally used for other functions) is able to take over the function of damaged tissue. Again, this would be difficult to understand under representational innateness. An important common conclusion that emerges from this work is that brain modularity is the outcome of experience, not of some genotypic blueprint, a point well made in an earlier book by Karmiloff-Smith (1992).

A useful summary table matches up linguistic and cognitive achievements with their suspected neural underpinnings over the period of childhood. The accompanying literature review emphasizes the twin processes of initial synaptogenesis and later synaptic pruning that characterize brain growth throughout childhood, and even later life.

Dynamic Systems. There is also an illuminating discussion of the dynamics of systems that evolve through time. The review focuses on systems in which the driving force is independent of time so that the development of the system is determined by the state of the system itself. One example is the vocabulary burst that occurs at around 2 years of age. Elman et al. suggest that the explosion in vocabulary is not driven by a maturational change or a change in the child's environment, but by the fact that the larger the child's vocabulary, the better he or she is at learning even more new words. Thus, gradual exposure to a constant linguistic environment leads to a non-linear, exponential explosion in vocabulary. The novel insight is that complex behaviors can emerge from one simple internal process. That process can be described by a simple equation that depends only on the current state of the system and does not assume any dramatic changes in the environment or the system. The language of first-order differential equations is suggested as appropriate for describing such processes.

Similar accounts have been proposed elsewhere in the literature (Thelen & Smith, 1994; van Geert, 1991; van der Maas & Molenaar, 1992). However, existing dynamic systems models of development often fall short by failing to specify the nature of the substrate they are discussing when talking about development. They focus on the shape of development, but fail to answer the questions of what develops and how development occurs. Connectionist modeling promises to fill that gap by providing a medium for discussing dynamic systems in terms of the processing involved in artificial neural networks. Unfortunately only a few pages directly address how differential equations could be related to particular models. Hence filling the gap between process and structure still remains a promise that needs to be fleshed out more fully.

Interactions Everywhere

The classical approach to dealing with issues of innateness and learning, particularly in ability testing, has been to attempt to quantify the proportions of variance in ability due to variation in genotype and environment. Unfortunately, this classic approach has not made much progress in establishing how abilities actually develop. Elman et al. argue that it is misleading to attempt such quantification because there are profound interactions between innate and environmental factors at every conceivable level from the molecular to the organism-environment level.

Interaction between brain systems is illustrated in a neurally inspired model of imprinting in chicks (O'Reilly & Johnson, 1994). Imprinting in chicks has been traced to two different brain systems, one of which predisposes attention to the head and neck regions of conspecifics and the other of which learns to follow the imprinted object. The two systems interact in that the first, attention system supplies the inputs to the second, learning system. The three layer model learned to map objects in particular locations onto preferences. Among the empirical findings with chicks that were simulated were the basic preference for the imprinted object over novel objects, subsequent reversal of preferences based on the relative length of the training sessions for the two objects, a critical period for reversibility depending on the time of exposure to the first object, a lingering preference for the first imprinted object over novel objects even after reversal, generalization to similar objects, and blended representations of two objects that are imprinted in rapidly successive presentations.

Interactions between brain and environment are illustrated in the two simulations on timing constraints that we reviewed earlier. They both show that complex knowledge can best be learned only after simpler relevant knowledge is already in place.

Suggestions for Future Research

Given that existing connectionist modeling only suggests some tentative solutions to developmental issues, the book closes with some useful suggestions for future research. These include the learning of multiple tasks in complex environments, the use of more active and goal-oriented models, models of social behavior and higher level cognition such as theory formation, and models that are more biologically realistic. We believe that most connectionist researchers and developmentalists would endorse these suggestions as being both worthwhile and challenging. We certainly do.

A CRITIQUE

As excellent as this book is, and it is the definitive treatment of connectionist approaches to development, it does have a number of shortcomings that surely will be addressed in future research. Chiefly, these shortcomings concern what the book lacks or ignores.

Real Constructivism

Our most serious reservation about this book is that it does not take constructivism seriously enough. This is somewhat surprising given the historical importance of constructivism in the debates on knowledge acquisition, the ample evidence reviewed even in this book for constructivist claims, the ready availability of connectionist learning algorithms based on constructivist principles, and the burgeoning evidence that models based on these principles have produced a number of successful simulations of development.

Elman et al. do seem to recognize the pivotal importance of constructivism in the debate on the origin of knowledge.

> Nativism and empiricism give the same answer to these questions: the structures are not new, they pre-exist, either in the organism or in the environment. Constructivism offered a very different view, considering development in terms of self-organizing emergent structures arising from the complex interactions between both organism and environment (p. 113).

Yet, curiously Elman et al. employ only static networks that never change their structure during learning; the only way these networks change is by adjusting their connection weights.[3] There is no question that these static networks develop new knowledge representations through connection weight adjustment; of course they do. But if that is all there is to constructivism, then Elman et al. are in the position of having the table turned, because that view could well be characterized as representational constructivism. Just as representational innateness is defined in terms of the sizes of network connection weights, so would representational constructivism. Connection weight changes in backpropagation learning tend to be quite small to guard against overshooting values that might minimize error. At some unspecified point, such small changes may induce new representations. But this seems far from the more dramatic development that Piaget had in mind when he hypothesized about major qualitative changes in processing mechanisms that enabled novel and more powerful types of representations that could not be achieved by younger children.

A key insight into what might be involved in such dramatic construction can be found in much of the brain development research reviewed by Elman et al., particularly the constructive events involving the birth and proliferation of new neurons and the establishment of new connections (synaptogenesis).Such constructive events could provide a network with additional computational power. Unlike the Elman et al. review, which emphasizes the one-time nature of neuron birth and synaptogenesis followed by extensive synaptic pruning over a longer period, a somewhat newer review indicates that synaptogenesis is driven by learning and continues throughout the life span (Quartz & Sejnowski, in press). Quartz and Sejnowski's reexamination of some of the same evidence reviewed by Elman et al. finds serious weaknesses in the evidence and identifies newer research suggesting that much more prominence should be given to synaptogenesis and dendritic arborization throughout the life span.

It has also been well known that neurogenesis (creation of new neurons) occurs throughout the life span, for example in the dentate gyrus of rodent hippocampus (Altman & Das, 1965; Cameron, Woolley, McEwen, & Gould, 1993; Kaplan & Hinds, 1977). A recent study showed that such neurogenesis is environmentally stimulated by living in enriched environments (Kempermann, Kuhn, & Gage, 1997). Once generated, the new neurons seem to facilitate new learning. Newly generated neurons, of course, indicate the presence of new neural circuitry.

Artificial neural networks that create new circuitry by growing new units and connections are available and fairly well known. One of these so-called generative

[3]Two exceptions are the syntax learning network of Elman that had its active memory capacity increased, and the cortical logic learning network of Shrager and Johnson that benefited from a trophic wave that enhanced the learnability of units as it washed over them. Of course, both of these techniques use factors that are external to the learning mechanism itself. As Elman et al. often point out, it is more elegant to use a single mechanism to explain transitions than an amalgam of different mechanisms.

learning algorithms has been used to model a range of developmental phenomena. This is the cascade-correlation algorithm (Fahlman & Lebiere, 1990), which Elman et al. mention but do not use. Cascade-correlation networks begin with just input and output units and attempt to reduce error at the outputs by adjusting their single layer of weights. The connection weight adjustment rule, known as quick-prop, is something like the delta rule discussed earlier, but quite a bit faster because it uses second, as well as first derivative information about the error surface (Fahlman, 1988). That is, it estimates not only the slope of the error curve, but also its curvature. This enables cascade-correlation to take larger and more decisive steps in making weight changes to reduce error.

When error can no longer be reduced in this fashion, the cascade-correlation algorithm recruits a new hidden unit, from a set of candidates, that is particularly good at tracking the network's error with its own activation. The connection weights going into these candidate hidden units are adjusted, again using quick-prop, but connection weights going into the output units are frozen during this period, the so-called input phase because of the focus on input-side weights. Once the best candidate at tracking error has been found, it is installed into the network, the remaining candidates are discarded, and training the output-side weights resumes with a return to the so-called output phase. At this point, the network has new computational power that it lacked before and thus can explore a new space of representations and hypotheses. With problems of much complexity, cascade-correlation networks can eventually recruit a number of hidden units, creating scalloped-shaped growth curves in which performance increases and then levels off over a number of cycles. Because each new hidden unit receives input, not only from the input units, but also from all earlier hidden units, the hidden layers become a set of knowledge filters at progressively higher levels in which earlier knowledge representations become re-represented and elaborated.

Generative networks are more than just a special case of static networks with some initial weights set to zero. The candidate units in waiting may exist somewhere in the complete system (the brain) or may be newly created, but they are not involved in any of the computations until they are brought into the network. They are not connected to the current network, are not taken account of in the weight learning algorithm, and cannot be brought on-line haphazardly as required by the task. Thus, generative networks differ fundamentally from static networks with certain connections set to zero. In such static networks, a zero connection weight is still accounted for in the learning algorithm and can be changed at any time in response to the same environmental pressures as any other weight. The difference between a generative network with k connections at time t, and a static network with n connections of which k ($k < n$) are non-zero at time t is that the generative network can only search a k-dimensional space whereas the static network can effectively search an n-dimensional space.

There are a number of these generative network algorithms available for learning in feedforward networks, but it is cascade-correlation that has been applied to a variety of developmental problem domains including the balance scale (Shultz, Mareschal, & Schmidt, 1994); seriation (Mareschal & Shultz, 1993); conservation (Shultz, 1996); causal prediction (Shultz, Schmidt, Buckingham, & Mareschal, 1995); integration of velocity, time, and distance cues for moving objects (Buckingham & Shultz, 1994); discrimination shift learning (Sirois & Shultz, 1997); and acquisition of personal pronouns (Oshima-Takane, Takane, & Shultz, 1997; Shultz, Buckingham, & Oshima-Takane, 1994). In some cases, static backpropagation networks were shown to be unable to capture the range of children's stages that were captured by cascade-correlation networks (Buckingham & Shultz, 1996; Shultz, Mareschal, & Schmidt, 1994). The capacity to grow in computational power appears to be necessary in these cases to capture the full range of stages found in children.

Another interesting feature of generative networks like cascade-correlation, in distinction to static networks, is that they are able to escape Fodor's paradox regarding the impossibility of experiential accounts of cognitive development. Fodor (1980) argued that experiential accounts were in principle incoherent because it was impossible to learn something that one did not already know, in the sense of being able to represent it. For example, one could not learn the concept of "green flower" unless one already possessed the concepts of "green," "flower," and "conjunction." Learning was thought by Fodor to be a matter of assembling concepts one already possessed to form hypotheses that could be tested against experiential feedback. We refer to this as Fodor's paradox because it seems to be a fundamentally sound argument against the apparently correct assumption that cognitive development is at least partly driven by experience.

A series of computational arguments have established that (a) a feedforward network, if it has enough hidden units, can approximate any computable function, (b) commitment to a particular network topology limits the range of functions that can be learned, and (c) generative networks can learn any computable function in polynomial time, i.e., in a reasonable amount of time (Mareschal & Shultz, 1996; Quartz, 1993). Because of point *b*, static networks cannot escape Fodor's paradox, but because of points *a* and *c*, generative networks can and do escape. The key to constructivist development is the ability to grow in computational power when needed, and generative networks such as cascade-correlation are able to grow by recruiting new hidden units when needed.

It is also worth noting that cascade-correlation can, in principle, achieve both the trophic wave of cortical development found in the Shrager and Johnson (in press) simulation and the starting small effect found in the Elman (1993) simulation. Cascade-correlation effectively implements a wave of development as it continues to create deeper networks with additional hidden layers. The difference between the two approaches is that this wave-like development is a natural feature of cascade-correlation learning rather than an externally imposed wave as in the Shrager and Johnson model. Cascade-correlation also implements starting small because it starts learning without any hidden units and adds new hidden units only as it needs them. This ensures that simple functions are learned first and that more complex ones are built on top of existing, simpler functions.

Elman et al.'s review of neurological growth provides abundant evidence that events somewhat analogous to cascade-correlation learning are common in mammalian brains. These brain events include experience mediated bursts in neural capacity (analogous to hidden unit recruitment in cascade-correlation), ossification of learning ability with time (analogous to the freezing of input side weights to hidden units in cascade-correlation), bursts in metabolic activity and performance occurring after bursts in neurogenesis and synaptogenesis (analogous to output phase learning in cascade-correlation), and synaptic pruning (analogous to the discarding of non-chosen candidate hidden units in cascade-correlation).

Thus, Elman et al.'s apparent representational constructivism can be contrasted with the constructivism found in cascade-correlation models that is based on architectural and timing considerations. Just as Elman et al. argue against representational innateness in favor of innateness based on architectural and timing constraints, we would argue against their representational constructivism (as a general account of development) in favor of a constructivism based on architectural and timing constraints. Although the details of the two arguments differ, they are quite analogous arguments. We agree that new representations can emerge from static networks, but we believe that this in itself is insufficient to account for much of the evidence coming from brain and cognitive development research, and from modeling studies.

Because static networks may be appropriate for modeling certain developmental domains and generative networks may be appropriate for modeling other domains, we devote a separate section to the issue of how one might determine which sort of model to use for any particular developmental problem.

Neglect of Other Important Models

Although it may seem odd to criticize a book that covers some 18 models for missing a few, we feel that the particular models chosen were not always the best available to illustrate intended points. As might be inferred from our earlier criticisms, we feel that more generative connectionist models could usefully have been discussed. Those cascade-correlation models that surpass the performance of static networks on the same problem might have been particularly interesting to discuss. For example, the cascade-correlation model of the balance scale captures all four stages seen with children—weight information, distance information when weights are equal on both sides, both weight and distance information but confusion when weight and distance information conflict with each other, and finally, mostly correct performance (Shultz, Mareschal, & Schmidt, 1994). In contrast, the presented balance scale model (McClelland & Jenkins, 1991) continues to oscillate between stages 3 and 4, never settling fully into the final stage 4.

Moreover, the McClelland and Jenkins model is presented in the book as an example of abrupt development from small, continuous changes in connection weights. However, it has recently been demonstrated that changes in this model's performance only seem abrupt due to the fact that the performance measure consists of diagnosis of qualitatively distinct rules (Raijmakers, van Koten, & Molenaar, 1996). When a continuous performance measure is used, this model shows only linear increases in performance. This was shown with a multiple regression technique in which performance was predicted by two predictors—continuous time and a binary variable indexing the periods before or after the major transition. Only the continuous time variable was a significant predictor of performance, indicating only linear increases in performance.

Unfortunately, no comparable longitudinal data from children are available on the balance scale. However, longitudinal data on conservation performance in children did show a sharp burst in continuous performance, indicated by a significant binary before-after predictor. Interestingly, a cascade-correlation model also reveals a burst in continuous conservation performance using the multiple regression technique (Shultz, 1997). Analysis of the knowledge representations in these networks showed that the performance bursts were due to the recruitment of key hidden units. So far, there is no realistic model of psychological development that we know of that shows sharp performance bursts due only to connection weight adjustments. Only toy, demonstration models of the sort presented in Elman et al. currently show such phenomena.

Another example of cascade-correlation networks surpassing the coverage of comparable static networks is in the area of the integration of velocity, time, and distance cues (Buckingham & Shultz, 1996). Cascade-correlation networks capture the three stages observed with children: identity (e.g., velocity is proportional to distance), additive (e.g., velocity is proportional to the difference between distance and time), and multiplicative (e.g., velocity is proportional to the ratio of distance to time) (Buckingham & Shultz, 1994). However, static backpropagation networks are either too weak to capture the final, multiplicative stage, or too powerful to capture the intermediate additive stage (Buckingham & Shultz, 1996).

No Consideration of the Opposition

Undoubtedly the most significant opposition to connectionist modeling of psychological development comes from symbolic computational approaches, particularly those that argue for or actually employ symbolic rules (Anderson, 1993; Klahr, 1989; Ling & Marinov, 1993; Newell, 1990). It has been argued by the symbolic camp that connectionist models would succeed only to the extent that they implement symbolic rules (Fodor & Pylyshyn, 1988). If connectionist models fail to implement symbolic rules, then it is argued they are doomed to failure because they will not show the systematicity, generativity, and universal generalization abilities of rules. Because we believe that connectionist models might be able to deal successfully with these issues, we were somewhat disappointed not to see them addressed in this important book.

WHEN TO USE STATIC OR GENERATIVE MODELS?

When might one want to use static networks to model cognitive development and when might it be more appropriate to use generative networks to model cognitive development? Even if generative networks are more powerful general learning systems than static networks, there may be domains in which it is more appropriate to use static networks than generative networks to model development. This issue is yet to be resolved and it would seem very likely that both types of mechanisms are present during development. Although we cannot currently offer a definitive solution to this puzzle we wish to suggest one way in which the two approaches may be reconciled. The roots to this answer lie in the arguments made about why development occurs and how learning constraints interact with environmental structure to produce task specific functional representations.

We suggest that static networks should be used to model domains that are constant across all individuals and for which evolutionary pressures have tightened up the dynamic constraints in order to accelerate the emergence of task appropriate representation. Examples of such domains might include basic spatial and temporal reasoning abilities, basic causal inference, memory, categorization, and possibly some aspects of language. These are all abilities that begin to develop very early in infancy and are found in all cultures. Environmental constraints are such that no matter where an infant is born he or she will need to develop this knowledge in a form that is consistent across all cultures.

Higher level cognitive skills that build on this initial learning tend to vary greatly across the world. The learning demanded of shepherd children in the Himalayas is very different from that demanded of Western children preparing for academic exams. One cannot realistically expect task specific networks to be preprogrammed in a genotype that would account for all the possible cultural variations in what a child might have to learn. Although there is some flexibility in the possible initial architectures that can be used to learn a task, in general, getting the architecture wrong can have a dramatic effect on whether a task can be learned or not (Mareschal & Shultz, 1996). Learning a wide range of tasks requires the ability to construct appropriate networks.

Learning about numbers (Sophian, 1996) is an excellent domain to illustrate the distinction between the role of static networks and generative networks in modeling cognitive development. Early quantification abilities (subitizing) and some cardinal and ordinal abilities have been demonstrated in infants and very young children. These properties are characteristic of all number systems. Other properties such as the use of numbers as symbols and arithmetic develop at a later age (Sophian, 1996) and show great cultural variability (Nickerson, 1988; Zhang & Norman, 1993). Learning to use these complex systems presupposes that the child has the prior ability to quantify. We suggest that learning basic universal cognitive skills such as quantification and some aspects of ordinality and cardinality that are acquired very early should be modeled using static networks. These abilities reflect regularities in all real world environments independently of cultural variation. In contrast, learning about the use of specific number systems should be modeled using generative networks that might build on the products of the prior early learning.

Many of the existing generative models have assumed prior quantification ability to solve a domain specific problem. Thus, whether it is a model of the balance-scale task (Shultz, Mareschal, & Schmidt, 1994), the potency and resistance task (Shultz et al., 1995), or estimating time, distance, and velocity relations (Buckingham & Shultz, 1994), the inputs to the network reflect magnitudes as evaluated by prior quantification. The fact that these networks can construct their own architecture gives them the flexibility to apply quantification knowledge to a broad range of possible task domains.

Static networks have been used to model a number of other basic infant abilities such as categorization (Mareschal & French, 1997; Quinn & Johnson, in press), memory interference effects (Mareschal & French, 1997), object permanence (Mareschal et al., 1995; Munakata, McClelland, Johnson, & Siegler, in press) and delayed response tasks such as the A-not-B task (Dehaene & Changeux, 1989). These are abilities found in every individual and can serve as building blocks for learning more complex idiosyncratic tasks. A domain with evidence of strong anatomical invariants across individuals may be another indicator of a task better suited to modelling with a static network.

THE EXERCISE BOOK

The Elman et al. book is accompanied by a volume of exercises entitled *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations* by Plunkett and Elman. In many ways, running networks is one of the most important things to learn about connectionism. Connectionist modeling is an applied skill. As with many other skills, one cannot completely understand it without practical experience. The hands-on experience makes many of the more abstract issues discussed in the Elman book sufficiently concrete. We have seen many people new to connectionist modeling find it surprisingly rewarding to design and build their first network and watch it learn a task.

The exercises in Plunkett and Elman can be broken down into two sections. The first section is the longer of the two and comprises the first eight chapters. This section is devoted to teaching about feedforward networks and backpropagation learning. Readers can work though a series of exercises specially designed to help them understand the inner workings of simple networks. It begins with a discussion of the functioning of individual units and of how to compute the flow of activity through a network. This is followed by a discussion of modeling methodology and how one might approach a particular task in order to model it. The next six chapters lead the reader through exercises that expand on such things as linear seperability in tasks, the role of hidden units and internal representations, autoassociation and categorization, conditions for generalization, receptive fields and spatial processing, simple recurrent networks, and temporal processing. The exercises involve reflection, computational modeling, and some hand calculations to ensure that the reader understands what the computer is doing during a simulation. At the end of each chapter are answers to the problems set in the exercises. These are pedagogical answers that aim to explain the results rather than just presenting the correct result.

The final four chapters differ from the first eight. Rather than presenting new material, they demonstrate how techniques can be applied to modeling cognitive development. There are chapters focusing on qualitative changes in learning, learning the English past tense, the balance-scale task, and the importance of starting small in development. Through these chapters, readers will be able to replicate and evaluate published simulations.

This companion volume is an exceptionally well put together pedagogical tool. The combination of lucid explanation, didactic questioning, hand calculations, and modeling would enable a fairly naive reader to get up to speed in modeling. The trick, however, is to work through all of the exercises. Cutting corners may seem quicker, but in the end will leave readers with many holes in their knowledge. After completing the entire book, readers would be able to reason confidently about feedforward backpropagation networks and be well placed to move on to more specific modeling techniques should they wish to.

The book assumes no more mathematical knowledge than that required to pass an introductory undergraduate statistics course. All the exercises are run with a specially designed windows-based program called *Tlearn*. The accompanying software can run on Macintoshes or PCs under Windows 95. There is inevitably a trade-off between the user friendliness and the power of a simulator. The Tlearn simulator is easy to come to grips with, but is somewhat limited in what it can do. It is primarily an educational tool, although models developed with this simulator have been published (Quinn & Johnson, in press). At the time of writing this essay the web page:

http://www.cs.cmu.edu/Groups/AI/html/faqs/ai/neural/faq.html

provided an overview of frequently asked questions about artificial neural networks. This included a descriptive list of other freely or commercially available neural network simulators.

CONCLUSIONS

The Elman et al. volume provides an excellent introduction to connectionism and how it might usefully be applied to basic issues of psychological development. The theoretical discussions of what it means for an ability to be innate are particularly illuminating, and the arguments for exploring developmental issues about transition with neural network models are compelling. While covering innateness and learning in great detail, the book curiously fails to take constructivism seriously enough, particularly in view of neurological, psychological, and modeling evidence for the importance of constructivist ideas. Still, the authors are to be commended for raising the debate about transition mechanisms in development to a new, higher, and more enlightening level. This book is an exemplary contribution to cognitive science, integrating material from a wide variety of cognitive disciplines, and is one of the most important books in cognitive development to emerge in recent years. Readers who wish to explore some existing connectionist models and learn to build connectionist models of their own will undoubtedly benefit from the well designed companion volume of computer exercises by Plunkett and Elman.

REFERENCES

Altman, I., & Das, G.D. (1965). Autoradiographic and histologic evidence of postnatal neurogenesis in rats. *Journal of Comparative Neurology, 124*, 319–335.

Anderson, J.A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

Anderson, J.R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Buckingham, D., & Shultz, T.R. (1994). A connectionist model of the development of velocity, time, and distance concepts. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 72–77). Hillsdale, NJ: Erlbaum.

Buckingham, D., & Shultz, T.R. (1996). Computational power and realistic cognitive development. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 507–511). Hillsdale, NJ: Erlbaum.

Cameron, H.A., Woolley, C.S., McEwen, B.S., & Gould, E. (1993). Differentiation of newly born neurons and glia in the dentate gyrus of the adult rat. *Neuroscience, 56*, 337–344.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 257–291). Hillsdale, NJ: Erlbaum.

Case, R. (1985). *Intellectual development*. New York: Academic Press.

Case, R. (1992). *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Erlbaum.

Chomsky, N. (1959). A review of B.F. Skinner's verbal behavior. *Language, 35*, 26–58.

Chomsky, N. (1971). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1975). *Reflections on language*. New York: Parthenon.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Dehaene, S., & Changeux, J. P. (1989). A simple model of the prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience, 1*, 255–261.

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.

Fahlman, S.E. (1988). Faster-learning variations on back-propagation: An empirical study. In D.S. Touretzky, G.E. Hinton, & T.J. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 38–51). Los Altos, CA: Morgan Kaufmann.

Fahlman, S.E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In D.S. Touretzky (Ed.), *Advances in neural information processing systems*, Vol. 2 (pp. 524–532). Los Altos, CA: Morgan Kaufmann.

Fischer, K.W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87*, 477–531.

Fodor, J. (1980). Fixation of belief and concept acquisition. In M. Piatelli-Palmarini (Ed.), *Language and learning: The debate between Chomsky and Piaget* (pp. 143–149). Cambridge, MA: Harvard Press.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT Press.

Fodor, J.A. & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3–71.

Gesell, A. (1929). *Infancy and human growth.* New York: Macmillan.

Hebb, D.O. (1949). *The organization of behavior: A neuropsychological theory.* New York: Wiley.

Hertz, J., Krogh, A., & Palmer, R.G. (1991). *Introduction to the theory of neural computation.* Reading, MA: Addison Wesley.

Hinton, G.E. (1992). How neural networks learn from experience. *Scientific American*, September, 145–151.

Hull, C.L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Jacobs, R.A., Jordan, M.I., & Barto, A.G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science, 15*, 219–250.

Karmiloff-Smith, A. (1992). *Beyond modularity.* Cambridge, MA: MIT Press.

Kaplan, M.S., & Hinds, J.W. (1977). Neurogenesis in the adult rat: Electron microscopic analysis of light radioautographs. *Science, 197*, 1092–1094.

Kempermann, G., Kuhn, H.G., & Gage, F.H. (1997). More hippocampal neurons in adult mice living in an enriched environment. *Nature, 386*, 493–495.

Kendler, T.S., & Kendler, H.H. (1959). Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology, 58*, 56–60.

Klahr, D. (1989). Information preocessing approaches. *Annals of Child Development, 6*, 135–185.

Lenneberg, E.H. (1967). *Biological foundations of language.* New York: Wiley.

Ling, C.X., & Marinov, M. (1993). Answering the connectionist challenge: A symbolic model of learning the past tenses of English verbs. *Cognition, 49*, 235–290.

Lipsitt, L.P. (1966). Learning processes of human newborns. *Merrill-Palmer Quarterly, 12*, 45–71.

Mareschal, D., & French, R.M. (1997). A connectionist account of interference effects in early infant memory and categorization. In M.G. Shafto & P. Langley (Eds.)., *Proceedings of the Ninteenth Annual Conference of the Cognitive Science Society* (pp. 484–489). Hillsdale, NJ: Erlbaum.

Mareschal, D., Plunkett, K., & Harris, P. (1995). Developing object permanence: A connectionist model. In J.D. Moore & J.E. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 170–175). Hillsdale, NJ: Erlbaum.

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. 676–681). Hillsdale, NJ: Erlbaum.

Mareschal, D., & Shultz, T.R. (1996). Generative connectionist networks and constructivist cognitive development. *Cognitive Development, 11*, 571–603.

Marchman, V. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*, 215–234.

McClelland, J.L., & Jenkins, E. (1991). Nature, nurture, and connections: Implications of connectionist models for cognitive development. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 41–73). Hillsdale, NJ: Erlbaum.

Munakata, Y., McClelland, J.L., Johnson, M.H., Siegler, R. (In press). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review.*

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Nickerson, R. (1988). Counting, computing and the representation of numbers. *Human Factors, 30*, 181–199.

Nolfi, S., Elman, J.L., & Parisi, D. (1990). *Learning and evolution in neural networks.* Techical Report No 9019, University of California, Center for Research in Language, San Diego.

O'Reilly, R.C., & Johnson, M. (1994). Object recognition and sensitive periods: A computational analysis of visual imprinting. *Neural Computation, 6*, 357–389.

Oshima-Takane, Y., Takane, Y., & Shultz, T.R. (1997). *The learning of first and second person pronouns in English: Network models and analysis.* Submitted for publication.

Pascual-Leone, J. (1970). A mathematical model for transition in Piaget's developmental stages. *Acta Psychologica, 32*, 301–345.

Parisi, D., Cecconi, F., & Nolfi, S. (1990). Econets: Neural networks that learn in an environment. *Network, 1*, 149–168.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child.* New York: Basic Books.

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York: William Morrow.

Plunkett, K., & Elman, J.L. (1997). *Simulating nature and nurture. A handbook of connectionist exercises.* Cambridge, MA: MIT press.

Quartz, S.R. (1993). Neural networks, nativism, and the plausibility of constructivism. *Cognition, 48*, 223–242.

Quartz, S.R., & Sejnowski, T.J. (In press). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences.*

Quinn, P.C., & Johnson, M.H. (In press). The emergenece of perceptual category representations in young infants. *Journal of Experimental Child Psychology.*

Raijmakers, M.E.J., van Koten, S., & Molenaar, P.C.M. (1996). On the validity of simulating stagewise cognitive development by means of PDP networks: Application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science, 20*, 101–136.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1 (pp. 318–362). Cambridge, MA: MIT Press.

Shrager, J., & Johnson, M.H. (In press). Modeling the development of cortical function. In B. Julesz & I. Kovacs (Eds.), *Maturational windows and cortical plasticity in human development: Is there a reason for an optimistic view*? Reading, MA: Addison Wesley.

Shultz, T. R. (1996). A generative neural network analysis of conservation. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 65–66). Hillsdale, NJ: Erlbaum.

Shultz, T.R. (in press). A computational analysis of conservation. *Developmental Sciences.*

Shultz, T.R., Buckingham, D., & Oshima-Takane, Y. (1994). A connectionist model of the learning of personal pronouns in English. In S.J. Hanson, T. Petsche, M. Kearns, & R.L. Rivest (Eds.), *Computational learning theory and natural learning systems*, Vol. 2: *Intersection between theory and experiment* (pp. 347–362). Cambridge, MA: MIT Press.

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: Erlbaum.

Sirois, S., & Shultz, T.R. (1997). *Neural network modeling of discrimination shifts: A developmental approach.* Submitted for publication.

Sophian, C. (1996). *Children's numbers.* Boulder, CO: Westview Press.

Spelke, E.S. (1994). Initial knowledge: Six suggestions. *Cognition, 50*, 431–445.

Spiker, C.C. (1963). Verbal factors in the discrimination learning of children. In J.C. Wright & J. Kagan (Eds.), *Basic cognitive processes in children*, Vol. 28: *Monographs of the Society for Research in Child Development* (pp. 53–69).

Sternberg, R.J. (Ed.). (1984). *Mechanisms of cognitive development.* New York: W. H. Freeman.

Thatcher, R.W. (1992). Cyclic cortical reorganization during early childhood. *Brain and Cognition, 20*, 24–50.

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

van der Maas, H.L.J., & Molenaar, P.C.M. (1992). A 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.

White, S.H. (1963). Learning. In H.W. Stevenson (Ed.), *Child psychology: The sixty-second yearbook of the National Society for the Study of Education*, Part 1 (pp. 196–235).

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749–750.

Zhang, J., & Norman, D.A. (1993). A cognitive taxonomy of numeration systems. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 1098–1103). Hillsdale, NJ: Erlbaum.