Developmental Science 6:4 (2003), pp 413–429

Connectionist models of development

Yuko Munakata[1] and James L. McClelland[2]

1. *Department of Psychology, University of Colorado, Boulder, USA*
2. *Center for the Neural Basis of Cognition, Mellon Institute, Pittsburgh, USA*

Abstract

How have connectionist models informed the study of development? This paper considers three contributions from specific models. First, connectionist models have proven useful for exploring nonlinear dynamics and emergent properties, and their role in nonlinear developmental trajectories, critical periods and developmental disorders. Second, connectionist models have informed the study of the representations that lead to behavioral dissociations. Third, connectionist models have provided insight into neural mechanisms, and why different brain regions are specialized for different functions. Connectionist and dynamic systems approaches to development have differed, with connectionist approaches focused on learning processes and representations in cognitive tasks, and dynamic systems approaches focused on mathematical characterizations of physical elements of the system and their interactions with the environment. The two approaches also share much in common, such as their emphasis on continuous, nonlinear processes and their broad application to a range of behaviors.

Introduction

The study of development raises fundamental questions about how learning and change occur. For example, why are there sometimes stage-like transitions in development? Why are there critical periods for some types of learning? Why do children sometimes show U-shaped performance curves as they learn, initially getting worse at a skill before ultimately succeeding? What factors lead to developmental differences between individuals, and specifically to developmental disorders? Why do infants and children so frequently show apparent dissociations in their knowledge, seeming to know things when tested in one way but seeming completely unaware when tested in a different way? What neural developments contribute to each of these behaviors?

Connectionist models have been used to explore these and other puzzles of development. Such models support complex, nonlinear, emergent processes (as discussed by Thelen & Bates, this issue), of the sort that likely contribute to the developmental phenomena described above. Models allow exquisite control over and observation of these complex processes, to explore how they contribute to behavior. In this paper, we first provide a brief overview of the connectionist framework. We then discuss three types of contributions from specific connectionist models for informing our understanding of development:

1. In exploring the role of nonlinear dynamics and emergent properties, and their relevance for understanding nonlinear developmental trajectories, critical periods in development and developmental disorders.
2. In exploring the nature of the representations that lead to behavioral dissociations – simultaneous success and failure on different tasks meant to measure the same knowledge – particularly during development.
3. In providing insights into neural mechanisms, particularly why different brain regions become specialized for specific functions, in terms of computational tradeoffs that require the specialization of multiple brain regions for optimal performance.

We draw comparisons and contrasts with the dynamic systems framework at several points in our presentation, and we close with an overall discussion of the relation between the connectionist and dynamic systems frameworks.

Overview of connectionist framework

As elaborated below, processing in connectionist models occurs through the propagation of activation through networks of simple processing units (Figure 1). Knowledge that underlies processing is stored in connection weights between these processing units. Changes in

Address for correspondence: Yuko Munakata, Department of Psychology, 345 UCB, University of Colorado at Boulder, Boulder, CO 80309-0345, USA; e-mail: munakata@psych.colorado.edu or James L. McClelland, Center for the Neural Basis of Cognition, 115 Mellon Institute, 4400 Fifth Ave., Pittsburgh, PA 15213, USA; e-mail: mcclelland+@cmu.edu

© Blackwell Publishing Ltd. 2003, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

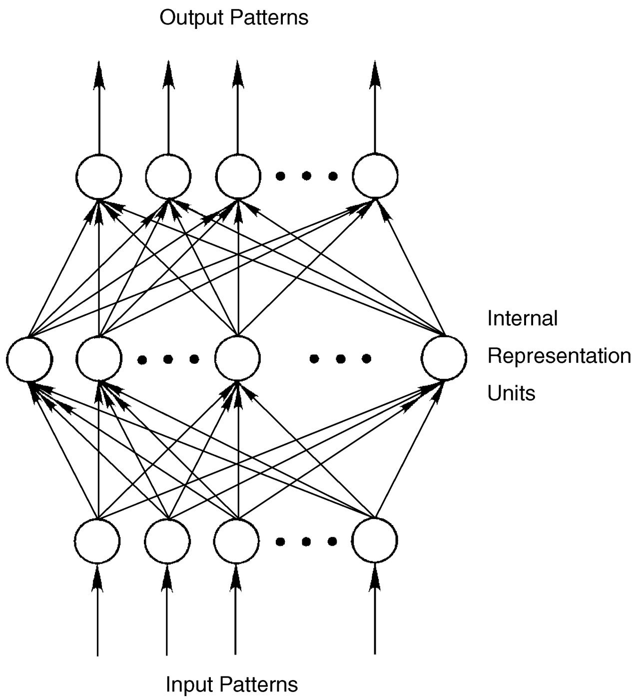

Figure 1 *A feed-forward network of the kind often used in many connectionist simulations. Note that the network can become a recurrent network by adding connections in the opposite direction, and/or by adding connections among the units within the same layer. From Rumelhart, Hinton & Williams (1986), Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland & PDP Research Group (Eds.),* Parallel distributed processing, *Volume 1, Figure 1, p. 320. Copyright 1986, MIT Press. Reprinted with permission.*

connections, driven by experience, provide a mechanism both for learning and for development. Connectionist models can be informed by neuroscience in a variety of ways, for example, in the activation functions governing the updating of processing units, the learning algorithms governing changes to connections, the structured organization of networks and the incorporation of neuromodulation. As we shall see, connectionist models can also inform neuroscience, for example, by providing computational insights into why different brain regions are specialized for different functions.

Activation dynamics

In the connectionist framework, the propagation of activation is a dynamic process operating in a continuous state space and evolving continuously over time, although this is not always stressed by all connectionists or in all models. In models that simulate this gradual activation process, it is typically formalized as a differential equation relating the rate of change of some variable (such as the activation of some unit) to the inputs it is currently receiving from other units via weighted connections. One simple such equation is found in the cascade model (McClelland, 1979):

$$\frac{da_i}{dt} = k\left(\sum_j a_j w_{ij} - a_i\right) \tag{1}$$

where a_i represents the activation of receiving unit i, a_j represents the activation of sending unit j, w_{ij} represents the connection weight from unit j to unit i, and k is a time constant that determines the rate of change of activation. This equation gives rise to a simple exponential approach to asymptote in a one-layer network or a more complex sigmoid-like approach to asymptote in a multilayer net.

In most contemporary models, the activation process is assumed to be nonlinear. A common implementation of a nonlinear activation function replaces Equation 1 with:

$$\frac{dnet_i}{dt} = k\left(\sum_j a_j w_{ij} - net_i\right) \tag{2}$$

$$a_i = f(net_i) \tag{3}$$

where net_i refers to the net input to unit i, and f is a simple nonlinear function like that illustrated in Figure 2.

In models that learn, the learning process often requires many presentations of an ensemble of training patterns. In this case, modelers often simplify the activation dynamics down to a single step computation, in

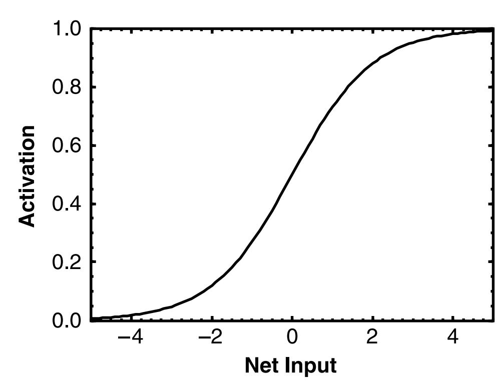

Figure 2 *Nonlinear activation function often used in neural network simulations. The actual function shown, called the logistic function,* $a_i = \frac{1}{(1 + e^{-net_i})}$, *is the function most frequently used.*

© Blackwell Publishing Ltd. 2003

which the value of k in Equation 2 is effectively set to 1. This can result in the loss of some interesting structure in the settling process, but has made it practical to study the dynamics of learning.

Learning dynamics

As with activation processes, learning rules are generally given in the form of differential equations, indicating the dynamic nature of the learning process. Changes to connection weights can arise due to self-organizing processes (e.g. based on a Hebbian type of mechanism whereby 'units that fire together wire together'), or due to error-driven learning, whereby feedback signals lead to weight adjustments that minimize the system's error. These ideas are formalized in differential equations such as:

$$\frac{dw_{ij}}{dt} = \varepsilon a_i a_j \tag{4}$$

$$\frac{dw_{ij}}{dt} = \varepsilon \delta_i a_j \tag{5}$$

where ε represents the 'learning rate' and δ_i represents unit i's contribution to the overall error, specifically, the degree to which a change in the input to unit i will produce a change in the overall error. The computation of the value of the δ variables is accomplished using a process much like the forward propagation of activation (Rumelhart, Hinton & Williams, 1986a). Equation 4 represents a self-organizing algorithm, whereby connection weights change according to the activations of sending and receiving units. Equation 5 represents an error-driven algorithm, whereby weights change to minimize the system's error.

General principles: representations, gradual learning and interactivity

The work reviewed here exemplifies several principles of connectionist models. First, there is an emphasis on the nature of the representations governing behavior. This is a crucial issue in the creation of models and in the understanding of how such systems work and develop. In the connectionist framework, representations can take the form of patterns of activity distributed across processing units. These patterns are graded in nature, with representations varying in strength in terms of the number of units contributing to them, the activation levels of those units, the amount of noise in the activation patterns, and so on. Exploring the nature of these underlying representations is critical for understanding how these systems behave in their 'mature' form – for example, to see the kinds of representations that can begin to solve the difficult computational problem of object recognition (Mozer, 1987), subserve reading (Plaut, McClelland, Seidenberg & Patterson, 1996) and support higher level cognition (O'Reilly & Munakata, 2000). Exploring underlying representations is perhaps even more important for understanding the *development* of these systems – for example, why behavioral dissociations are particularly salient during development (Munakata, 2001), and how systems can progress from knowing little about the category structure of the world to making fine-grained distinctions between categories and recognizing that different kinds of features are central to different categories (Rogers & McClelland, in press; McClelland & Rogers, 2003).

Second, in connectionist models, learning typically occurs gradually, as small changes are made to connection weights. Understanding how change occurs is a central goal of studying development, so connectionist models provide an important tool in allowing specific learning mechanisms to be implemented and their effects to be observed. Resulting models have demonstrated, for example, how self-organizing processes can lead systems to perseverate, repeating previous behaviors when they no longer make sense, in the same way that infants and children do (Munakata, Morton & Stedron, in press). Further, models using error-driven learning processes have helped to flesh out the Piagetian notions of assimilation, accommodation and equilibration, and their role in a variety of learning situations (McClelland, 1995).

Third, processing in connectionist models is typically highly interactive. That is, there is bi-directional communication between units within and between layers. Bi-directional connectivity makes networks much more complex dynamic systems, but does have advantages that are exploited in many models. Bi-directional connectivity allows activation signals to carry information that can be used to estimate the δ variables above (thereby providing a biologically plausible method of computing δ values (O'Reilly, 1996)). Also, as Hopfield (1982) first established, under many conditions networks with bi-directional connections will tend to settle to fixed-points or attractor states. The particular attractors in the network, the size of their basins of attraction and the resulting tendencies networks exhibit to settle into them, depend on the values of the connection weights, which in turn depend on the learning rule. Finally, interactivity has important consequences for how systems develop. For example, one part of a network may develop in a particular way in part because of how another part of the network is developing, due to the communication between these parts of the network. This point will be elaborated in the discussion of developmental disorders.

© Blackwell Publishing Ltd. 2003

Variants and additional mechanisms

It should be noted that the connectionist framework is by no means monolithic; many variants on the activation and learning mechanisms mentioned above have been proposed, and many additional mechanisms can be investigated within it. For example, 'generative' connectionist architectures have been investigated (Mareschal & Shultz, 1996), in which new units (along with connections to other units) are recruited into a network as a function of experience (see also Quartz & Sejnowski, 1997). Furthermore, it is possible to incorporate assumptions about changes that occur as a function of age that might interact with experience-dependent learning processes. All such approaches fall under the broad umbrella of connectionist models and should be considered in a larger review. In this brief review, we focus on work exploring the developmental implications of the simple formulation of connectionist mechanisms given above, seeking to explain developmental change as reflecting the operation of the learning process rather than age-dependent changes. Some specific points of contrast with other work in which age-dependent changes are invoked are considered at relevant points in our analysis.

Next, we explore how connectionist models have contributed to the study of nonlinear dynamics and emergent properties in development, the role of graded representations in behavioral dissociations and the specialization of different brain regions.

Nonlinear dynamics and emergent properties

In any complex system (human cognition, the weather, etc.), the whole is more than the sum of its parts. Complex processes like cognition cannot be reduced simply to the operation of individual neurons, the effects of neurotransmitters, etc. Instead, complex processes require an understanding of nonlinear interactions among a large number of components, and properties that emerge in systems as a result of such interactions. Models are essential for exploring this kind of complexity. The connectionist and dynamic systems frameworks have both emphasized the importance of this kind of complexity in human cognition (as discussed by Thelen & Bates, this issue). In what follows, we focus on simulations that explore the role of nonlinear dynamics and emergent properties as relevant for understanding nonlinear developmental trajectories, critical periods and developmental disorders.

Nonlinear developmental trajectories

Connectionist models have been used to explore a number of aspects of nonlinear dynamics and emergent properties that might subserve behaviors observed during development. For example, small, incremental changes in strengths of connections can lead to stage-like changes in the behavior of connectionist models (McClelland, 1995). Training in such networks is measured in epochs, which may consist of one presentation of each of the examples defining a training environment, or of a randomly chosen set of examples. Children often appear to show stage-like developmental changes, for example, in their understanding of balance scale problems (Inhelder & Piaget, 1958; Siegler, 1976). Figure 3 shows a network used to explore these changes, and a graph of the strengths of the connections in the network (McClelland, 1989). The graph illustrates that gradual incremental learning according to a simple (error-driven) connection-adjustment rule can explain why children's progress can appear highly stage-like in character. Specifically, the connection strengths (and as a result the overt responses that the network makes to inputs) exhibit periods of stability, followed by periods of rapid change, followed by further periods of stability once mastery up to a certain level has been achieved. This process is seen for both the weight dimension and the distance dimension, although it occurs more quickly for the weight dimension, because information about variation in weight is available more frequently than information about variation in distance (this accords with aspects of children's performance indicating that mastery of the distance dimension may in fact be very gradual in this task, Wilkening & Anderson, 1982). Concentrating on the fairly abrupt transition seen on the weight dimension, we can ask, what produces this stage-like progression in learning? In the network, the ability to represent a certain kind of information (say, the magnitude of weight on each side of the balance scale) is captured in the input-to-hidden connections, and the ability to use the representation is captured in the hidden-to-output connections. Both sets of connections initially have small random values (indicative of an initial state of ignorance of the relation between weight and the behavior of the balance scale). Each set of connections must become organized before changes in the other can be effective in allowing the network to produce correct outputs. This means that the δ terms introduced earlier are very small at first, so learning correspondingly progresses slowly. As the connections begin to become organized, the δ terms gradually become much larger; effectively the network becomes more sensitive to input about weight than it had been previously. The result is an acceleration of learning about the weight dimension, which only levels off again when the network has extracted all of the benefit available from it (in terms of eliminating the error in determining which

© Blackwell Publishing Ltd. 2003

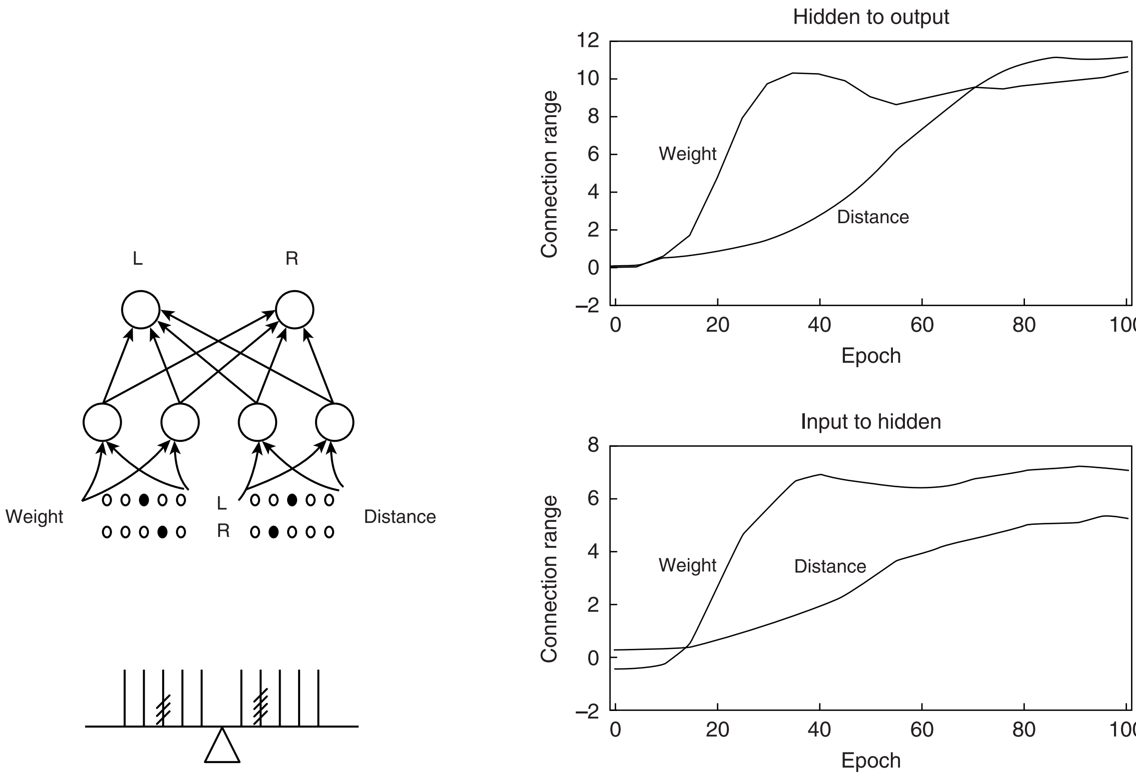

Figure 3 *(a) The simple three-layer network model of the balance scale problem of Inhelder and Piaget (1958) as later studied by Siegler (1976). (b) The state of the network's connection-based knowledge of the impact of weight and distance on the balance scale, as encoded in the connections from the input to the hidden layer and from the hidden to output layer as a function of ongoing learning through exposure to training examples. Connection range variable shown reflects the degree to which the connection strengths in the layer in question differentiate the different values of the weight and distance inputs. The x-axis represents amount of training in 'Epochs', each containing 100 randomly sampled training examples. From McClelland (1989), in R.G.M. Morris (Ed.),* Parallel distributed processing: Implications*, Figures 2.7, p. 25 and 2.12, p. 32. Copyright 1989, Oxford University Press. Reprinted by permission of Oxford University Press.*

side of the scale will go down with a given balance scale problem).

The McClelland (1989) model may not have captured all aspects of the transition from stage to stage that are seen in developmental data (Raijmakers, van Koten & Molenaar, 1996). This is an important issue that must be addressed in further research. One possible explanation for the shortcomings of the McClelland (1989) model is that it was completely deterministic. In subsequent work we have come to the view that this is an oversimplification; it appears to be a basic principle of human cognitive processes that they are intrinsically variable (McClelland, 1993; Siegler & Munakata, 1993). Intrinsic variability can be incorporated by assuming that noise is present in the net input to each unit. This tends to lead to variability in overt behavior, which will be greater around transitions than either before or after them. During transitions, the behavior of the network is fairly close to the margin between one or another mode of responding, so that random variability in processing would be more likely to lead to inconsistency and lability in the network's outputs. These are characteristics often exhibited at or near stage transitions (Siegler & Munakata, 1993). Other factors, such as attention, might be especially important during transitions, so that natural fluctuations in attention would make more difference and thereby contribute to increased variability in the transitions.

Further, connectionist models can show U-shaped performance curves as they learn, sometimes getting much worse on at least some items or subtasks in a domain as they progress from a very early stage where performance is uniformly poor toward overall mastery (e.g. Rumelhart & McClelland, 1986; Plunkett & Marchman, 1993; O'Reilly & Hoeffner, submitted). Sometimes (e.g. Rumelhart & McClelland, 1986), such U-shaped performance curves are produced in simulations in which the training corpus presented to the network shifts

© Blackwell Publishing Ltd. 2003

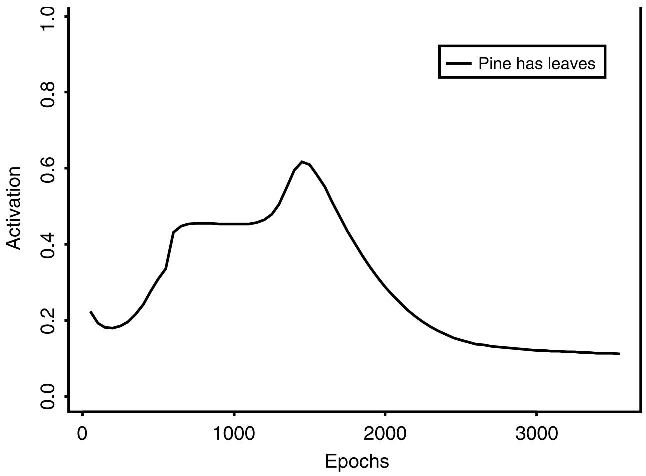

Figure 4 *Inverted U-shaped trajectory of the property 'has leaves' when given 'pine' as input, from a network learning various properties of a large number of different concepts. After an early period of low activation, the network transitions to a stage in which the activation of the 'has leaves' property first reflects the proportion of objects that have leaves (about .45), then moves in the direction of the proportion of plants that have leaves (much higher), before finally dropping downward to reflect the actual probability that a pine tree has leaves (which is 0 in the network's training environment). The figure is based on results of a simulation described in Rogers and McClelland (in press).*

over the course of development. Since some researchers have criticized the idea that the training environment might shift in this way (e.g. Pinker & Prince, 1988), it is important to note that U-shaped development can occur even in situations where the environment stays quite stable over time.

One case in point arises for a network's knowledge of the properties of atypical objects in a simulation model of conceptual development. Following the work of Rumelhart and Todd (1993), Rogers and McClelland (in press; McClelland & Rogers, 2003) considered a network learning about the properties of various specific objects chosen from the categories of birds, fish, trees and flowers. Among the attributes the network had to learn was whether or not a pine tree has leaves. The activation of the attribute 'has leaves' for the pine tree is shown in Figure 4. After an initial period of low activation of all properties, the network's activation of the 'has leaves' property comes to reflect the overall proportion of objects in the network's training environment that have leaves; this is about 0.45. At this point in training (covering about epochs 600–1300), the network has not differentiated the different kinds of objects, so the activation of 'has leaves' just reflects the probability of leaves given that the item is any object. However, because the plants the network is learning about all share many properties, and because the animals all share a different set of properties, the network gradually differentiates the plants from the animals. At this point, it still treats the pine like all other plants, most of which have leaves, and so it shows a sharp increase in the tendency to attribute leaves to the pine tree (up to a maximum activation of about .65). Thus, performance on whether the pine has leaves, previously indecisive, now becomes more clearly incorrect. With further exposure to the domain, the network learns to differentiate the pine from the other plants, thereby gradually overcoming this error, and producing an (inverted) U-shaped curve. The U-shaped trend over development reflects the interplay of progressive differentiation of concepts with differences in the probability of the property for concepts that are lumped together at different levels of granularity. As the representations of the various concepts differentiate progressively, its property attributions come to reflect the level of differentiation it has achieved, and therefore the probability of attributing leaves to pine trees undergoes a U-shaped progression. In this way, U-shaped learning can occur even in situations where the ensemble of experiences on which the network is trained remain stable over time.

It may be noted that U-shaped learning can also occur on a fine grain time scale, with many reversals in accuracy of performance with particular items (Plunkett & Marchman, 1993). Many times these micro-U-shaped trends can occur as a consequence of the effects of intrinsic variability in processing, coupled with effects of the random sequence of training examples; connection changes favoring a particular training example (e.g. live-lived) can lead to interference with performance on another (e.g. give-gave).

Critical periods

The analysis of critical periods in development has become an important area of application of connectionist approaches. The topic is fascinating to connectionists in part because of the general paradox of critical periods and in part because the dynamic properties of connectionist models suggest that they may offer some explanations of critical period effects.

Starting small

One widely discussed issue within the connectionist framework is the benefits of training networks on a simplified set of training examples before exposing them to the full complexity of a complex domain, such as sentential syntax (Elman, 1993). Similar ideas have been proposed

© Blackwell Publishing Ltd. 2003

TSXS	TXS	TSSSXS
TSSXXVV	TSSSXXVV	TXXVPXVV
TXXTTVV	TSXXTVV	TSSXXVPS
TSXXTVPS	TXXTVPS	TXXVPS
PVV	PTTVV	PTVPXVV
PVPXVV	PTVPXTVV	PVPXVPS
PTVPS	PVPXTVPS	PTTTVPS

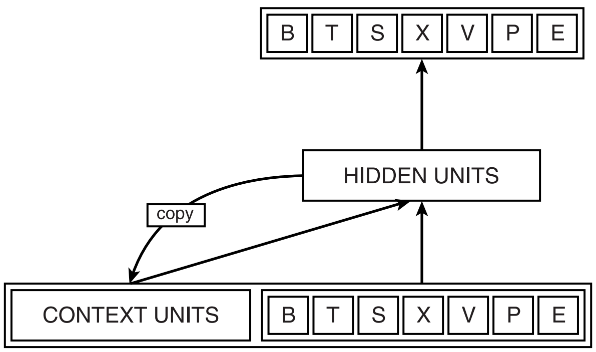

Figure 5 *(a) The fixed set of training sequences generated from a Reber Grammar (Reber, 1976) that were used to train the simple recurrent network shown in (b). Note that in the network there is an input and an output unit for each character appearing in the training sequences, as well as a set of hidden units. The task of the network is to predict the next element it is about to see, starting with B at the beginning of each sequence and ending with E at the end. Training sequences are chosen at random, and a large number of sequences are shown in each epoch of training. Input units are fully connected to hidden units, which are fully connected to output units. In addition, the network contains a set of context units, which serve to hold a copy of the activations of the hidden units from the previous time step; these units too are fully connected to the hidden units. All forward-going weights are subject to modification via back-propagation. From Servan-Schreiber, Cleeremans & McClelland (1991), Graded state machines: the representation of temporal contingencies in simple recurrent networks,* Machine Learning, *7, Figure 3, p. 164, Figure 12, p. 173, and Figure 4, p. 165. Copyright 1991, Kluwer Academic Publishers. Reprinted with permission.*

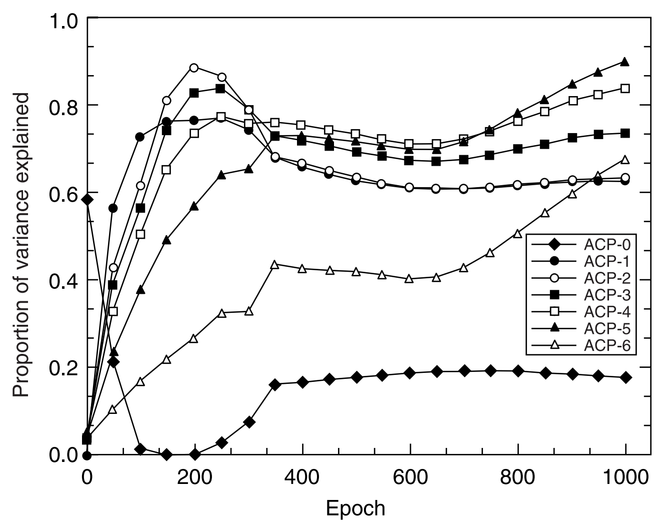

Figure 6 *The degree to which the predicted outputs of the network shown in Figure 5b correspond with the actual conditional probabilities (ACP) of occurrence of successive elements, conditional on taking different numbers of prior elements into account. The predictions based on taking no prior elements (ACP-0) into account are equivalent to the relative frequencies of occurrence of the elements without regard to preceding elements. From Servan-Schreiber, Cleeremans & McClelland (1991), Graded state machines: the representation of temporal contingencies in simple recurrent networks,* Machine Learning, *7, Figure 3, p. 164. Copyright 1991, Kluwer Academic Publishers. Reprinted with permission.*

based on behavioral investigations of critical periods in language learning (Newport, 1988, 1990). Essentially the notion is that a system that is only capable of processing a small amount of information at one time cannot be distracted by complex relationships, and will benefit from its simplicity by focusing on the simpler, lower-order regularities first. This idea has a great deal of appeal, and is likely to be highly relevant to our understanding of the ease with which young children acquire their first, native language. However, there are other perspectives on this issue that are worth serious consideration.

One observation is that connectionist networks tend to start small on their own accord, and no special interventions are necessary to force them to pay attention first to low-order statistical properties of inputs. An illustration of this property of simple recurrent networks (the type of network used by Elman) is presented in a simulation performed by Servan-Schreiber, Cleeremans and McClelland (1991). In this simulation, a simple recurrent network was trained with a specific set of examples generated from a stochastic finite-state transition network ('Reber Grammar', Reber, 1976, Figure 5). The task was simply to predict the next element in a stream of inputs representing the successive letters in the grammar. As Figure 6 illustrates, the network quickly picked up at the very beginning of training the 0-order structure of the sequence, that is, the base rate of occurrence of particular elements of the sequences (curve labeled ACP-0). Over the first 100 epochs or so, it came to exhibit sensitivity to first-order sequential structure; this is the conditional probability of an element, given only the immediately preceding element. This is illustrated by the first rising curve, ACP-1. Successively rising curves thereafter reflect increasing sensitivity to successively more and more preceding elements. When the simulation was terminated, the network was still progressing in its mastery of sensitivity to sixth-order

© Blackwell Publishing Ltd. 2003

structure, illustrated by the gradually rising ACP-6 curve. In this way, the network progressively came to take higher and higher order structure into account, based solely on the statistics in its training (many specific examples of U-shaped learning of the predictions of various elements in various contexts occur throughout this process). Thus, the network inherently tends to start small, exhibiting sensitivity first to low-order structure before picking up on more complex, higher-order sequential dependencies. A further relevant point of note is that work following up on Elman's (1993) study has not successfully replicated the finding that curtailing the complexity of the early training environment actually facilitated learning (Rohde & Plaut, 1999). Rohde and Plaut considered a wide range of variations of parameters of the network and the training regime, and found that the network typically performed best if trained on the full corpus from the beginning of training. One possible reason for this is consistent with an alternative account of the basis of critical period effects, which we now proceed to consider.

Entrenchment

A fairly straightforward way of accounting for reduction of plasticity as a function of prior learning is to assume that as a network learns, the knowledge that it has becomes somehow entrenched, so that it is more difficult to alter with subsequent learning. There are in fact several ways that entrenchment can occur in connectionist networks.

Commitment of units and connections. As units and connections become strengthened through training on a set of examples, it can become more difficult to get the network to learn additional examples added to the training set. This can occur for the following reason. As a network is trained, the connection weights become fairly large, as do the bias terms that determine the default activation states of units. Early in learning, the presentation of a new input tends to produce weak, intermediate levels of activation; but later, all inputs, whether old or new, tend to produce relatively extreme activations. Under these conditions, a fixed change to the strengths of the connections tends to produce rather little change in the output of the network. Because the changes made to connection weights in error correcting networks are determined by how strongly the change will reduce the error, the actual change that is made is a very small one. The result is a kind of double whammy, where the change is small and ineffectual, resulting in very little progress in a fully committed net. The effect becomes more and more pronounced as learning proceeds, resulting in stronger and stronger resistance to new learning. This form of entrenchment appears to underlie simulations of the 'age of acquisition' effect, in which words that are first presented early in the overall training of a network are more robustly learned than words that are first presented at a later point in training (Ellis & Lambon-Ralph, 2000; Zevin & Seidenberg, 2002).

Pruning of unused connections. A somewhat different although related effect is the fact that in trained networks, many connections that are not needed become very small. These small connections do not propagate activation; also, they prevent changes to connections below or above them in a multi-layer network from having any substantial effect. In general, the effect of pruning is to restrict the possibilities of reorganization in the network. This factor may also be at work in the simulation models of Ellis and Lambon-Ralph (2000) and Zevin and Seidenberg (2002).

Counter-productive Hebbian learning. A different kind of entrenchment effect arises in networks that have been trained with Hebbian, rather than error-correcting learning algorithms. Hebbian learning essentially amounts to a policy of strengthening whatever response the system makes to the inputs that it receives. Hebbian learning can be beneficial if the system is initially set up to make appropriate responses, since these will easily be strengthened. For example, suppose an infant comes with an auditory perceptual system capable of strengthening differential responses to clusters of speech sounds. Differential responses to sounds experienced in the environment will be strengthened, and a simple local interaction process among the units involved in representing different sounds can result in the gradual recruitment of all of the available representational units to represent one of the clusters of experienced sounds or another (see McClelland, Thomas, McCandliss & Fiez, 1999, for a model showing how this process can occur). Once this occurs, it becomes difficult to reorganize the system to respond to the set of clusters represented by a different language environment. For example, it may be difficult for a Japanese subject who has a single representation spanning the English /r/ and /l/. The presentation of either an /r/ or an /l/ will result in the activation of the existing, single representation. Hebbian learning at this point can actually be counter-productive, since it will tend to strengthen the now-established tendency of the given input to activate the established representation, thereby paradoxically strengthening the entrenched tendency to hear both the English /r/ and /l/ sounds as the same (again, see McClelland *et al.*, 1999, for further discussion of simulations and preliminary data testing this idea).

Overall, connectionist models provide a highly fertile environment for the consideration of alternative

© Blackwell Publishing Ltd. 2003

accounts of critical and sensitive period effects. The models can also be used to explore how the relative timing of biological events can interact with the learning process (Shrager & Johnson, 1996). In any such effort it should be borne in mind, however, that apparently biological events can be triggered by learning mechanisms. For example, a certain type of post-synaptic receptor molecule (the so-called NMDA receptor), which is crucial for triggering the biochemical processes that result in changes in synaptic efficacy, changes over the course of development in a way that may explain why changes in neuronal response properties are harder to induce in older than younger animals (Carmignoto & Vicini, 1992). Interestingly, it has recently been found that the change in the NMDA receptor is itself experience dependent, and indeed can be reversed if experience is withheld (Quinlan, Olstein & Bear, 1999). Overall it is apparent that degree of plasticity can change as a result of experience, not simply as a bi-product of some autonomous process driven simply by the passage of time.

Developmental disorders

Recently, connectionist models have proven useful for exploring developmental disorders in terms of nonlinear dynamics and emergent properties (Harm & Seidenberg, 1999; Hoeffner & McClelland, 1993; Karmiloff-Smith, Scerif & Thomas, 2002; Oliver, Johnson, Karmiloff-Smith & Pennington, 2000; Thomas & Karmiloff-Smith, in press). As elaborated below, these simulations have demonstrated the importance of considering processes of development in understanding developmental disorders (Karmiloff-Smith, 1998). This approach contrasts with one motivated by a static view of brain function, in which neural systems or modules are innately specified for particular functions; in this view, developmental disorders arise due to genetic alterations that target particular associated cognitive functions. This view has led some researchers to argue that impairments in adult (acquired) and developmental disorders have a similar underlying cause. That is, if a neural system is innately specified for a particular function, damage to that system late in life should lead to similar effects as a developmental disorder that affects the genes coding for that particular function (see discussion in Karmiloff-Smith *et al.*, 2002). However, this approach ignores the role that processes of development can play in developmental disorders. In contrast, within the connectionist framework, it is more natural to consider particular functions emerging in particular brain areas through a highly interactive process of development, with different regions developing as they do in part because of how other regions are developing (rather than simply having their functions prespecified). Connectionist models have highlighted how: (1) small, quantitative differences in the starting state of systems can lead through a process of development to qualitative differences in outcome, and (2) damage to a system early in development can lead to very different behaviors than damage to the same system late in development.

First, disorders can emerge through small changes in low-level properties of a system's start state, which interact with processes of development (Oliver *et al.*, 2000). For example, with small changes to the firing thresholds of units, networks showed large changes in the topography of their representations. When units fired too readily, or when firing was too difficult, networks failed to develop typical topographic representations of neighboring units that responded to similar stimuli. The nature of these representations was not prespecified in the networks, but emerged as the networks learned based on the activations of their units, which were affected by their firing thresholds. Such deviations in a system's ability to develop representations, emerging through a low-level change interacting with the developmental process could yield further emergent disordered processes (e.g. in the ability to learn language using atypical representations), though the exact computational consequences of non-topographic representations are uncertain. Similar ideas regarding the large effects of small differences in start state have been explored in the typical development of specialized systems. For example, small differences between units in their rate of activation updating can lead one set of units to become specialized in tracking the *location* of objects, while another otherwise identical set of units comes to specialize in the *identity* of objects (O'Reilly & Johnson, 1994; O'Reilly & McClelland, 1992).

Second, damage to a developing system can lead to quite different patterns of impairment than the same damage to a mature system, such that behaviors in cases of developmental disorders are likely to differ from cases of adult brain damage (Karmiloff-Smith, 1998). Damage to the mature system affects the specializations that have already emerged, whereas damage to the younger system more strongly affects *how* specializations in the damaged region and in other regions subsequently develop. Models from adult brain damage may thus not be extendable to developmental disorders. Patterns of impairment in models of word reading and past tense formation differed dramatically depending on whether they were damaged prior to or following training, despite identical levels, locations and types of damage (e.g. removal of connections, addition of noise to activation processing) (Thomas & Karmiloff-Smith, in press). In some cases, the developing networks required much

© Blackwell Publishing Ltd. 2003

more damage than the mature networks to show the same levels of impairment. In other cases, following the same damage, the developing and mature networks differed in their patterns of performance (e.g. with a developing network showing better performance on novel words than on familiar exception words, whereas mature networks showed the opposite pattern). In this way, in interactive systems where functions emerge with experience, the process of development can play a key role in shaping the effects of an atypical start state. Connectionist simulations hold the promise of illuminating how specific alterations identified at the neural level may develop in an interactive, emergent manner into the particular behavioral outcomes observed in developmental disorders.

We believe this work on developmental disorders in the neural network framework is quite consistent with the core principles of the dynamic systems approach, but we do not know of any dynamic systems simulations exploring this domain.

Graded representations in behavioral dissociations

People often show dissociations in their behavior, seeming to know things when they are tested in one way, while seeming unaware of this information when they are tested in another way. Such dissociations occur across perception, attention, memory, executive functioning and language, and are particularly salient during development and following brain damage. Connectionist models have demonstrated how graded representations can lead to such dissociations across domains and populations (reviewed in Munakata, 2001).

Here, we focus on models of dissociations observed in infants' and children's perseveration (Morton & Munakata, 2002a; Munakata, 1998; Stedron, Munakata & Sahni, 2002). Across the first several years of life, children show dramatic repetitions of prior behavior when those behaviors are no longer appropriate. For example, after infants search for a toy repeatedly in one hiding location, they perseverate in searching in that location even after watching the toy being hidden in a different location – a behavior known as the A-not-B error (Diamond, 1985; Piaget, 1954). Infants will even perseverate when they do not need to remember anything to succeed at a task. For example, when faced with two towels to pull – one with a distant toy on it and one with a toy behind it – infants will choose the towel with the toy on it. However, if the towels are switched so that the towel that was to the infants' left (e.g. with the toy on it) is now to the infants' right, infants perseverate, continuing to pull the towel on the same side as before though it does not yield the toy (Aguiar & Baillargeon, 2000).[1] Similarly, after 3-year-olds sort cards by one dimension (e.g. according to color), they perseverate in sorting according to that dimension even after repeated instruction to switch to sorting by a new dimension (e.g. according to shape) (Zelazo, Frye & Rapus, 1996). Six-year-olds show the same pattern when asked to judge the emotion of a speaker; they can succeed at an initial task of judging emotion based on the content of the speaker's utterances, but they perseverate with content when asked to switch to judging emotion based on tone of voice (Morton, Trehub & Zelazo, in preparation).

Connectionist models have been used to explore dissociations in two aspects of perseveration. First, although infants and children seem to know things (e.g. what information is relevant for retrieving distant or hidden objects, or what information to use for sorting cards by color) when they are tested on these tasks at the outset, they seem unaware of this information when they are tested on these tasks after engaging in conflicting behaviors (e.g. retrieving objects from a different location, or sorting cards by shape). Connectionist models have accounted for these dissociations in terms of a competition between two distinct types of representations – latent and active (Munakata, 1998) – each of which is graded in nature. Latent representations take the form of changes in connection weights that build from prior experience based on self-organizing learning (e.g. reaching to the left, or sorting cards by color). These latent representations are thought to rely on posterior cortical areas. Active representations take the form of maintained activations of processing units, and can represent currently relevant information (e.g. that the toy has been moved to the right, or that the new sorting rule is shape). These active representations are thought to rely on prefrontal systems. In tasks where children perseverate, these two types of representations compete; the currently relevant information must override what was relevant based on prior experience. Connectionist models have demonstrated that when the ability to maintain active traces is relatively weak, systems fall back on their latent representations and perseverate – in reaching for hidden objects (Munakata, 1998), in reaching even when all of the relevant task information is visible (Stedron *et al.*, 2002) and in sorting cards and judging emotion (Morton

[1] Infants will also perseverate in 'no-toy' versions of the A-not-B task, in which an experimenter waves one of two identical visible lids and infants are allowed to reach (Smith, Thelen, Titzer & McLin, 1999; Munakata, 1997). Although this task variant might appear to require no memory because nothing is ever hidden, infants must remember which of the two lids was waved to succeed in this task.

© Blackwell Publishing Ltd. 2003

& Munakata, 2002a). Thus, children may show dissociations in their performance based on how strong the latent representations are that need to be overcome. At the outset (e.g. before any reaching for toys or sorting of cards), there may be minimal latent representations (e.g. for where to reach or how to sort), such that children's active representations of the relevant task information are sufficiently strong to govern correct performance. In contrast, after repeatedly engaging in a behavior, latent representations are strengthened such that active representations of new information (e.g. a new location, a new sorting rule) cannot overcome the latent representations, resulting in perseveration.

Second, children can show remarkable dissociations between their perseverative behaviors and their apparent awareness of what they should be doing. For example, even as infants reach perseveratively for a hidden toy, they occasionally gaze at the correct hiding location (Piaget, 1954; Diamond, 1985; Hofstadter & Reznick, 1996). Perhaps even more compelling, even as children sort cards or judge emotions perseveratively (according to a previous rule), they can correctly answer questions about the new rule they should be using, such as where trucks should go in the shape game (Zelazo *et al.*, 1996), or what aspect of a speaker's voice they should be listening to (Morton & Munakata, 2002b; Morton *et al.*, in preparation). Connectionist models have demonstrated how such apparent knowledge–action dissociations could arise due to graded representations, with networks using weak representations for success at certain tasks but not others. For example, a network with only weak representations of a hidden toy was able to gaze correctly to the toy's location while reaching perseveratively (Munakata, 1998). The network's gaze system updated more frequently than the reaching system. This manipulation captured the fact that in the A-not-B task, infants' reaching is much more restricted than their gazing. This difference between reaching and gazing allowed the network to use weak representations to counter perseverative tendencies in its gazing in a way that the reaching system could not. Similarly, networks can use a weak representation of a new card sorting rule to answer questions about the rule ('Where do trucks go in the shape game?'), because there is no competition from latent representations (e.g. to sort by color) in this task (Morton & Munakata, 2002a). In contrast, the inherent conflict in the sorting task (e.g. a card is both red and a truck) requires the active representation of the new rule to be strong enough to overcome the latent bias toward the old rule.

Again, we believe that these connectionist simulations of perseveration and associated dissociations share much with the dynamic systems approach. One difference is that the connectionist models may better capture certain distinctions at the neural level, which may be lost in the abstractions of dynamic fields. For example, in the model of perseveration in the dynamic systems framework (Thelen, Schoner, Scheier & Smith, 2001), an infant's memory and stimuli that are presented to the infant are treated as qualitatively similar, as numerical abstractions that are summed as inputs to a dynamic field. However, environmental inputs and memory may be instantiated in qualitatively different ways in the neural system: certain forms of memory may be embodied in synaptic changes, whereas responses to particular environmental stimuli may be embodied in the firing of populations of neurons. As a result, memory and environmental inputs can have different consequences for behavior, and they can interact in nonlinear ways with other factors. This distinction between synaptic changes and neuronal firing is naturally captured in neural network models, in terms of the distinction between weights and activations.

Specialization of brain regions

The final contribution we discuss from the connectionist framework speaks to the relation between connectionist models and neuroscience. As discussed in the introduction, connectionist models can be informed by neuroscience in a variety of ways. But this relation is not a one-way street; models can inform neuroscience as well, rather than simply incorporating discoveries from neuroscience. In particular, models can help to illuminate why particular brain regions might become specialized for different functions (as discussed in the Developmental Disorders section), and why different regions might be required to subserve those specializations (rather than all of cortex, for example, subserving episodic *and* semantic *and* working memory). Here, we focus on how connectionist models have helped to explain aspects of the large-scale organization of the brain, in terms of computational tradeoffs. In a tradeoff, two objectives cannot be achieved simultaneously; a system must relinquish its ability to achieve one objective as it specializes on its ability to achieve the other objective. This approach to neural specialization provides an important complement to work demonstrating *what* brain areas contribute to different behaviors, by illuminating *how* and *why* those brain regions subserve their associated functions.

Connectionist simulations have demonstrated how computational tradeoffs may play a role in the specializations of the hippocampus and cortex in memory (McClelland, McNaughton & O'Reilly, 1995; O'Reilly

© Blackwell Publishing Ltd. 2003

& Rudy, 2001). These simulations have shown how a system that learns rapidly with non-overlapping representations is crucial for the recollection of particular episodes (such as meeting a particular person – e.g. remembering where you met her, what her name is, who you were with, and so on). The rapid learning with non-overlapping representations allows the system to quickly encode the memory and keep it distinct from memories for similar episodes (e.g. the meeting of other people). In contrast, simulations that learn slowly using overlapping representations tend to collapse across the differences of individual episodes; as a result, these systems instead specialize on representing the underlying structure of the environment (e.g. a schema for what typically happens in meeting people). Both types of representations and learning are useful, but there is a computational tradeoff between them; a single system cannot simultaneously specialize in both non-overlapping representations with fast learning and overlapping representations with slow learning. As a result, one neural system (the hippocampus) may specialize in the fast and non-overlapping functions, while another neural system (posterior cortex) may specialize in the slow and overlapping functions. This computational approach is consistent with (and may help to make sense of) findings from neuroscience regarding the anatomy and physiology of the hippocampus (Squire, Shimamura & Amaral, 1989). For example, areas of the hippocampus show very sparse levels of activity, which could contribute to relatively non-overlapping representations. Thus, a fundamental computational tradeoff in memory suggests the need for two specialized systems that the hippocampus and cortex appear to satisfy.

Connectionist simulations have also demonstrated how a computational tradeoff between interactive versus isolated representations may help to illuminate the specializations of posterior cortex versus prefrontal cortex, respectively (O'Reilly, Mozer, Munakata & Miyake, 1999; O'Reilly & Munakata, 2000). These simulations have shown how highly interactive representations, with numerous strong connections among units, activate related constructs from partial inputs to support schemas, inferences and semantic knowledge more generally (as emphasized in the original PDP volumes – McClelland, Rumelhart & PDP Research Group, 1986; Rumelhart, McClelland & PDP Research Group, 1986). In contrast, more isolated representations are required for systems to maintain representations over delays, in the absence of input, and in the face of noise (e.g. for working memory). Again, both types of representations are useful, but there is a computational tradeoff between them; a single system cannot simultaneously specialize on interconnected and isolated representations. As a result, one neural system (posterior cortex) may specialize on interconnected representations, while another system (prefrontal cortex) may specialize on isolated representations. This computational approach may help to make sense of findings from neuroscience regarding the anatomy (Levitt, Lewis, Yoshioka & Lund, 1993) and physiology (Rao, Williams & Goldman-Rakic, 1999) of prefrontal cortex, which may suggest more isolated representations in this region.

Although the work to date in this area has not been developmentally focused, it is developmentally relevant. Knowledge of the computational tradeoffs inherent in neural specializations, together with the time course of these specializations, could aid in understanding both typical and atypical development. For example, the computational account of hippocampal and cortical specializations may provide an explanation of infantile amnesia (McClelland *et al.*, 1995). In this account for the mature system, the hippocampus quickly encodes individual episodes based on cortical representations; episodic memories can later be retrieved when these distinct hippocampal representations activate relevant cortical representations. In the developing system, the hippocampus would quickly encode individual episodes based on early cortical representations, but retrieval of these memories would be hindered by dramatic changes to the cortical representations over the first years of life. That is, as the cortical representations reorganized, they would no longer serve to activate hippocampal representations, and hippocampal representations for particular episodes might activate cortical neurons that no longer represented those episodes. The result would be infantile amnesia. Note that many theories attribute infantile amnesia to representations changing, and becoming incompatible, so that prior experiences can no longer be accessed. However, these theories face a difficulty: if the representations change so dramatically that prior experiences are effectively no longer represented, one would expect dramatic general deficits to be observed as representations change. All previous knowledge would effectively be overwritten, so that children would not be able to recognize familiar objects, music, words, etc. But such dramatic general deficits are not observed. The computational tradeoff story can make sense of this. Cortical changes are slow, and possibly even slight, such that general deficits are not observed. But these slow and possibly slight cortical changes can lead to large changes in the hippocampal representations, leading to dramatic deficits, but only in episodic memory – hence infantile amnesia.

Further, understanding neural specializations in terms of computational tradeoffs could help to make sense of developmental changes in patterns of brain activity and

© Blackwell Publishing Ltd. 2003

effects of brain damage. For example, an area might subserve a computational function crucial for learning a skill, but unimportant for executing the skill once it is learned. An understanding of this kind of dependency at the computational level could help to explain why certain kinds of brain damage (or activity) are linked to the learning of skills such as reading but not to the expert execution of those skills (Stiles, Bates, Thal, Trauner & Reilly, 2002).

We believe that this type of connectionist work, on the specialization of brain regions, does not directly contradict the tenets of the dynamic systems framework. Nonetheless, currently such explorations may be more readily conducted within the connectionist framework, because aspects of these models may be more transparently mapped onto corresponding neural components.

Relation between connectionism and dynamic systems

Here we offer our thoughts about the similarities and differences between dynamic systems and connectionist models. In our view, connectionist models and dynamic systems are closely related theoretical perspectives. Both approaches place a strong emphasis on continuous, nonlinear processes. One could view dynamic systems theory in general as an overall mathematical theory that includes dynamic models of all kinds of things, including connectionist models, as special cases. However, in practice, the connectionist approach to modeling development is not a special case of the dynamic systems approach to modeling development. This latter approach, at least as practised by the dynamic systems researchers participating in the present special issue (Smith & Thelen, 1993; Spencer & Schöner, this issue) comes with some commitments that make it more specific than the whole class of possible dynamic systems models and that serve to differentiate the approach in some ways from connectionist approaches. It should be noted that there is a research tradition among a group of European developmentalists (including Moser, Raijmakers and others) that also identifies itself with a dynamic systems approach. For present purposes we confine ourselves to a comparison of connectionist models with the dynamic systems approach as characterized by Spencer and Schoner (this issue) and Smith and Thelen (1993). Comparison of connectionist models with the other work would be worthwhile, but space prevents such a consideration here.

While the connectionist work has tended to focus on topics in cognitive and linguistic development, most of the dynamic systems work has tended to focus on sensorimotor phenomena and real-time behavior in time and space. Relatedly, we think the dynamic systems work has tended to focus on direct mathematical characterization of the physical elements of the behaving system and on relationships among these elements and between them and the environment, while connectionist models have tended to focus on the representations that may underlie performance in cognitive and linguistic tasks. For example, Thelen's work on why babies will step right after birth and then cease to do so for some months makes reference to the dynamic properties of the legs as real physical objects, while McClelland's work on Siegler's balance scale addresses the child's developing representation of the weight and distance information required to predict how a balance scale would behave. In an area where the approaches have been applied to the same behavioral phenomena (the A-not-B error; Munakata, 1998; Thelen *et al.*, 2001) this difference between the approaches is apparent; Thelen *et al.* (2001) (see also Smith *et al.*, 1999; Thelen & Smith, 1994) use the dynamic systems framework to formulate theoretical proposals about the effect of performing an action on the creation of an attractor for that action, whereas Munakata (1998) is specifically concerned with the possible differences between two different kinds of representations of response tendencies, one short-lived representation in the activations of units and one longer-lived representation in the strengths of connections.

An important difference between many dynamical systems models and our own efforts is that we have stressed the role of learning as the engine of change in development, and we rely on such learning to account for developmental differences in performance. In contrast, in much of the dynamical systems work, developmental differences are attributed to differences in a control variable whose change as a function of age is assumed but not explained. For example, in neural network models of the development of object knowledge in infants (Mareschal, Plunkett & Harris, 1999; Munakata, McClelland, Johnson & Siegler, 1997), learning over the course of experience leads to the strengthening of connection weights, allowing networks to develop the ability to maintain representations of previously visible objects after they are hidden. This same kind of learning process is thought to lead to the ability of older children to maintain active representations of hidden objects for longer periods of time in the A-not-B task (Munakata, 1998). For comparison, in the model of Thelen *et al.* (2001), there is an external control variable that changes as a function of development, thereby changing the tendency of traces of immediately preceding activations to persist. We do not mean to suggest that dynamical systems researchers disagree with the idea that learning

© Blackwell Publishing Ltd. 2003

underlies these developmental changes; instead we suggest that such issues have often been outside of their focus. Learning could be incorporated into dynamical systems models, and we would view this as a desirable step for the future development of the framework.

Another difference between connectionist models and dynamic systems approaches is that dynamic systems models often tend to be formulated in low-dimensional systems with fairly simple dynamics, while connectionist models have tended to make use of systems that have a large number of dynamic variables – each of the units and connections in the network. Further complexity is added to the connectionist models since changes in the connection weights produce changes in the activation dynamics, and these in turn produce changes in the dynamics of learning (some of the specific ways in which this can happen have been discussed in qualitative terms above). The greater simplicity of the dynamic systems models affords an enviable mathematical tractability that is lacking in the connectionist models. This is not to say that there has not been formal mathematical work attempting to analyse connectionist networks; on the contrary, there is a vast literature of this type (e.g. White, 1989; Saad, 1998). However, while such analyses have shed some light on such things as the likelihood of convergence of neural networks and their tendency to generalize appropriately, they have not thus far succeeded in further elucidating their developmental dynamics beyond the kinds of qualitative statements about these issues that have been made earlier in this article.

It should be noted, however, that one difference between the dynamic systems approach and the connectionist approach to development is that to this point there have been no mathematicians with sufficient expertise in the analysis of dynamic systems who have tuned their particular mathematical tools for use in analyzing the properties of connectionist networks. While these networks may be more complex, an insightful mathematical analysis of the dynamics of development as it unfolds in such networks may be possible. We invite those mathematicians working in the dynamic systems framework to see just how far their tools might take us in understanding the dynamic properties of connectionist systems.

Conclusions

In closing we wish to emphasize the common ground that is shared between connectionist and dynamic systems approaches. Both approaches represent modern alternatives to frameworks that rely on discrete symbol-like constructs as the objects that are manipulated in real-time behavior and constructed to provide the basis for development. In this respect the models also share some commonalities with developments over the last 20 years or so in production system models (which now rely heavily on continuous and/or noisy activation values, e.g. Just & Carpenter, 1992; Anderson & Lebiere, 1998). There is also common ground in this respect with the growing use of Bayesian approaches to capture knowledge and inference (Anderson, 1990; Oaksford & Chater, 1994). A second commonality is that both approaches can be broadly applied to a wide range of different phenomena, and both allow detailed contact between theory and experiment. While there are differences between the details of the approaches and the distribution of phenomena addressed, we believe that useful insights have come from both and that further cross-communication between the protagonists of the two approaches is likely to contribute to the emergence of further insights into the dynamics of human development.

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.

Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J.R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Carmignoto, G., & Vicini, S. (1992). Activity-dependent decrease in NMDA receptor responses during development of the visual-cortex. *Science*, **258**, 1007–1011.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on $A\bar{B}$. *Child Development*, **56**, 868–883.

Ellis, A.W., & Lambon-Ralph, M. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**, 1103–1123.

Elman, J.L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, **48** (1), 71–99.

Harm, M.W., & Seidenberg, M.S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological Review*, **106** (3), 491–528.

Hoeffner, J.H., & McClelland, J.L. (1993). Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In E.V. Clark (Ed.), *Proceedings of the 25th Annual Child Language Research Forum* (pp. 38–49). Stanford, CA: Center for the Study of Language and Information.

© Blackwell Publishing Ltd. 2003

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

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, USA*, **79**, 2554–2558.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, **99**, 122–149.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, **2**, 389–398.

Karmiloff-Smith, A., Scerif, G., & Thomas, M. (2002). Different approaches to relating genotype to phenotype in developmental disorders. *Developmental Psychobiology*, **40**, 311–322.

Levitt, J.B., Lewis, D.A., Yoshioka, T., & Lund, J.S. (1993). Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 & 46). *Journal of Comparative Neurology*, **338**, 360–376.

McClelland, J.L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review*, **86**, 287–330.

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* (Chap. 2, pp. 8–45). Oxford: Oxford University Press.

McClelland, J.L. (1993). The GRAIN model: a framework for modeling the dynamics of information processing. In D.E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artifical intelligence, and cognitive neuroscience* (pp. 655–688). Hillsdale, NJ: Lawrence Erlbaum Associates.

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: Erlbaum.

McClelland, J.L., McNaughton, B.L., & O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**, 419–457.

McClelland, J.L., & Rogers, T.T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, **4**, 310–322.

McClelland, J.L., Rumelhart, D.E., & PDP Research Group (Eds.) (1986). *Parallel distributed processing. Volume 2: Psychological and biological models*. Cambridge, MA: MIT Press.

McClelland, J., Thomas, A., McCandliss, B., & Fiez, J. (1999). Understanding failures of learning: Hebbian learning, competition for representational space and some preliminary experimental data. In J. Reggia, E. Ruppin & D. Glanzman (Eds.), *Progress in brain research, Vol. 121, Disorders of brain, behavior and cognition: The neurocomputational perspective* (pp. 75–80). Amsterdam: Elsevier.

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

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

Morton, J.B., & Munakata, Y. (2002a). Active versus latent representations: a neural network model of perseveration and dissociation in early childhood. *Developmental Psychobiology*, **40**, 255–265.

Morton, J.B., & Munakata, Y. (2002b). Are you listening? Exploring a knowledge–action dissociation in a speech interpretation task. *Developmental Science*, **5**, 435–440.

Morton, J.B., Trehub, S.E., & Zelazo, P.D. (in preparation). Representational inflexibility in 6-year-olds' understanding of emotion in speech.

Mozer, M.C. (1987). Early parallel processing in reading: a connectionist approach. In M. Colthheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 83–104). Hillsdale, NJ: Lawrence Erlbaum Associates.

Munakata, Y. (1997). Perseverative reaching in infancy: the roles of hidden toys and motor history in the $A\bar{B}$ task. *Infant Behavior and Development*, **20** (3), 405–416.

Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: a PDP model of the $A\bar{B}$ task. *Developmental Science*, **1**, 161–184.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, **5** (7), 309–315.

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

Munakata, Y., Morton, J.B., & Stedron, J.M. (in press). The role of prefrontal cortex in perseveration: developmental and computational explorations. In P. Quinlan (Ed.), *Connectionist models of development*. East Sussex: Psychology Press.

Newport, E. (1988). Constraints on learning and their role in language acquisition: studies of the acquisition of American sign language. *Language Sciences*, **10**, 147–172.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science*, **14**, 11–28.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608–631.

Oliver, A., Johnson, M.H., Karmiloff-Smith, A., & Pennington, B. (2000). Deviations in the emergence of representations: a neuroconstructivist framework for analysing developmental disorders. *Developmental Science*, **3**, 1–23.

O'Reilly, R.C. (1996). Biologically plausible error-driven learning using local activation differences: the generalized recirculation algorithm. *Neural Computation*, **8** (5), 895–938.

O'Reilly, R.C., & Hoeffner, J.H. (submitted). Competition, priming, and the past tense U-shaped developmental curve.

O'Reilly, R.C., & Johnson, M.H. (1994). Object recognition and sensitive periods: a computational analysis of visual imprinting. *Neural Computation*, **6** (3), 357–389.

O'Reilly, R.C., & McClelland, J.L. (1992). *The self-organization of spatially invariant representations* (Parallel Distributed

© Blackwell Publishing Ltd. 2003

Processing and Cognitive Neuroscience PDP. CNS. 92. 5). Carnegie Mellon University, Department of Psychology.

O'Reilly, R.C., Mozer, M., Munakata, Y., & Miyake, A. (1999). Discrete representations in working memory: a hypothesis and computational investigations. *The Second International Conference on Cognitive Science* (pp. 183–188). Tokyo: Japanese Cognitive Science Society.

O'Reilly, R.C., & Munakata, Y. (2000). *Computational explorations in cogntive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

O'Reilly, R.C., & Rudy, J.W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychological Review*, **108**, 311–345.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Pinker, S., & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*, **28**, 73–193.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K.E. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, **103**, 56–115.

Plunkett, K., & Marchman, V.A. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, **48** (1), 21–69.

Quartz, S.R., & Sejnowski, T.J. (1997). The neural basis of cognitive development: a constructivist manifesto. *Behavioral and Brain Sciences*, **20**, 537–596.

Quinlan, E.M., Olstein, D.H., & Bear, M.F. (1999). Bidirectional, experience-dependent regulation of n-methyl-d-aspartate receptor subunit composition in the rat visual cortex during postnatal development. *Neurobiology*, **96**, 12876–12880.

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.

Rao, S.G., Williams, G.V., & Goldman-Rakic, P.S. (1999). Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. *Journal of Neurophysiology*, **81**, 1903.

Reber, A.S. (1976). Implicit learning of synthetic languages: the role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 88–94.

Rogers, T., & McClelland, J. (in press). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

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

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland & PDP Research Group (Eds.), *Parallel distributed processing. Volume 1: Foundations* (Chap. 8, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D.E., & McClelland, J.L. (1986). On learning the past tenses of English verbs. In J.L. McClelland, D.E. Rumelhart & PDP Research Group (Eds.), *Parallel distributed processing. Volume 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Rumelhart, D.E., McClelland, J.L., & PDP Research Group (Eds.) (1986). *Parallel distributed processing. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Rumelhart, D.E., & Todd, P.M. (1993). Learning and connectionist representations. In D.E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artifical intelligence, and cognitive neuroscience* (pp. 3–30). Cambridge, MA: MIT Press.

Saad, D. (1998). *On-line learning in neural networks*. New York: Cambridge University Press.

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.

Shrager, J., & Johnson, M.H. (1996). Dynamic plasticity influences the emergence of function in a simple cortical array. *Neural Networks*, **9**, 1119.

Siegler, R. (1976). Three aspects of cognitive development. *Cognitive Psychology*, **8**, 481–520.

Siegler, R., & Munakata, Y. (1993). Beyond the immaculate transition: advances in the understanding of change. *SRCD Newsletter, Winter Issue*.

Smith, L.B., & Thelen, E. (Eds.) (1993). *A dynamic systems approach to development: Applications*. Cambridge, MA: MIT Press.

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.

Squire, L.R., Shimamura, A.P., & Amaral, D.G. (1989). Memory and the hippocampus. In J.H. Byrne & W.O. Berry (Eds.), *Neural models of plasticity: Experimental and theoretical approaches* (pp. 208–239). San Diego, CA: Academic Press.

Stedron, J., Munakata, Y., & Sahni, S.D. (2002). In plain sight: simulating the role of memory in perseveration with visible solutions. Manuscript in preparation.

Stiles, J., Bates, E.A., Thal, D., Trauner, D., & Reilly, J. (2002). Linguistic and spatial cognitive development in children with pre- and perinatal focal brain injury: a ten-year overview from the San Diego longitudinal project. In M.H. Johnson, R.O. Gilmore & Y. Munakata (Eds.), *Brain development and cognition: A reader* (pp. 272–291). Oxford: Blackwell, 2nd edition.

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: MIT Press.

Thomas, M., & Karmiloff-Smith, A. (in press). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioral and Brain Sciences*.

White, H. (1989). Learning in artificial neural networks: a statistical perspective. *Neural Computation*, **1**, 425–464.

© Blackwell Publishing Ltd. 2003

Wilkening, F., & Anderson, N.H. (1982). Comparison of two rule-assessment methodologies for studying cognitive development and knowledge structure. *Psychological Bulletin*, **92**, 215–237.

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

Zevin, J., & Seidenberg, M. (2002). Age of acquisition effects in reading and other tasks. *Journal of Memory & Language*, **47**, 1–29.

© Blackwell Publishing Ltd. 2003