22 Neural Network Models of Cognitive Development

YUKO MUNAKATA, JENNIFER MERVA STEDRON, CHRISTOPHER H. CHATHAM, AND MARIA KHARITONOVA

This chapter covers neural network modeling (also known as connectionist or parallel-distributed processing modeling) as a tool for studying developmental cognitive neuroscience. Neural network models provide a powerful method for exploring the complex relation between brain development and cognitive development. This chapter reviews what neural network models consist of, why modeling is useful, and how models have helped to address fundamental questions about development. Important challenges for this methodology are also discussed, along with productive directions for future work within the neural network modeling framework.

Neural network models provide a powerful tool in the study of developmental cognitive neuroscience. Such models implement neural processes in computer simulations, in the form of mathematical equations that characterize neural activity and learning. Neural network simulations thus allow an exploration of the role of neural processes in behavior. The modeling methodology provides an important complement to other methods, by building upon findings from other studies and pointing the way toward new studies to advance our understanding of the relation between brain and behavior.

In this chapter, we cover neural network models of cognitive development from the perspective of answering three critical methodological questions: why, what, and how? More specifically, we explain *why* it is important to use neural network models in the study of developmental cognitive neuroscience and to explain more about *what* the nuts and bolts of neural network models entail. We then describe *how* neural network models have been used to address fundamental developmental questions about the origins of knowledge and how change occurs. We also discuss challenges relevant to each of these issues of the why, what, and how of neural network modeling. This chapter aims to confer an appreciation of the potential contributions of neural network models to the advancement of developmental cognitive neuroscience, as well as the ability to critically evaluate both the over- and underselling of this methodology.

Why

First, we describe some of the benefits of neural network modeling (adapted from O'Reilly and Munakata, 2000; see also Seidenberg, 1993; Rumelhart and McClelland, 1986; Elman et al., 1996). All these benefits are demonstrated by specific models covered later in the "How" section, and they support a productive interchange between modeling work and other methodologies. Some of these benefits are arguably conferred to some degree by purely verbal theories; however, implementing a working model of a theory is both more demanding and more powerful than simply stating the theory, and so provides greater benefits.

Models Allow Control

Models can be manipulated, lesioned, tested, and observed much more precisely than the thing being modeled (whether the thing is a single neuron, a small collection of neurons, a human infant, a monkey, and so on). Such control enables a clearer picture of the causal role of different factors. For example, in this chapter, we will see how such control allows an assessment of long-term effects of word frequencies in language learning.

Models Help Us to Understand Behavior

With such control, we can watch a model in action to get a sense of why behavior unfolds as it does. Seemingly unrelated or even contradictory behaviors can be related to one another in nonobvious ways through common neural network mechanisms. Further, neural network models can provide an important bridge between neural and cognitive aspects of behavior. Lesioned models can also provide insight into behavior following specific types of brain damage and, in turn, into normal functioning. In this chapter, we will see how models can help us to understand various potentially puzzling aspects of children's behavior, including nonlinear trajectories in their development.

Models Deal with Complexity

Complex, emergent phenomena (the brain is more than the sum of its parts) can be captured in models in principled, satisfying ways. Such

emergent phenomena arise from the complex interactions of multiple elements of a model, without being obviously present in the behavior of the individual elements. Without the models and the principles, such complexity might otherwise be lost in vague, verbal arguments. In this chapter, we will see how models have provided insight into the emergence of complex phenomena in domains as diverse as infant object processing and children's semantic development.

Models Are Explicit Creating an implemented model forces you to be explicit about your assumptions. For example, what do children encode about a particular task, and how? How do they subsequently process this information? What kinds of mechanisms support their learning in this task? Explicitness about such assumptions confers many potential benefits, including the generation of novel, empirically testable predictions and the deconstruction of black box constructs. In this chapter we will see the advantages of such explicitness in a model that deconstructs the notion of an object-permanence concept and leads to novel predictions subsequently tested in infants.

What

We will see all these benefits in action when we consider how models have been used to explore issues in cognitive development, in the next section. First, we consider the nuts and bolts of what neural network models are, which will provide the foundation for understanding their contributions. Here, we focus on five critical elements of neural network models: units, weights, net input and activation functions, and learning algorithms (for more extensive treatments of these and other nuts and bolts of neural network models, see Elman et al., 1996; O'Reilly and Munakata, 2000; Rumelhart and McClelland, 1986). Each of these elements maps onto neural constructs while capturing important aspects of psychological processing, allowing neural network models to provide an important step in understanding the relation between neural and cognitive development.

Units and Weights Neural network models consist of two basic elements: units and weights (figure 22.1a). In models most closely tied to the underlying biology, each unit corresponds to a neuron, the activity of each unit corresponds to the spiking of a neuron, and each weight corresponds to a synapse (the strength of the weight corresponds to the efficacy of the synapse). Models of psychological phenomena are much more scaled down; single units correspond to collections of neurons or even entire brain regions, the activity of each unit corresponds to the overall firing rates of these neurons, and the weights between units represent synapses between the groups of neurons.

Units communicate with one another by means of their weights. Each unit receives activity from other units by way of its weights, and if enough such input is received, the unit becomes active. The unit then sends this activity to other units by means of its weights to those units, an action that in turn influences the activity of those units.

In most network simulations of behavior, units are organized into layers. An input layer (or layers) receives information that reflects the external world, in the form of patterns of activity on the units in the layer. Networks are described as "perceiving" their environments when they receive this input information, with the particular type of perception (seeing versus hearing, and so on) depending on the modality that the input layer represents. In the simplistic example shown in figure 22.1b, the network sees the word "dog" when its input units for the letters *d*, *o*, and *g* are activated. An output layer (or layers) produces patterns of activity that are interpreted in terms of some response behavior. For example, the network in figure 22.1b can say either "cat" or "dog," by activating the corresponding output unit. (For much more realistic models of word reading, which incorporate semantic representations and more complex phonological and orthographic representations, see Harm and Seidenberg, 2004; Plaut et al., 1996; O'Reilly and Munakata, 2000.) Additionally, some number of hidden layers may sit between the input and the output layers, providing the network with the capability to transform input information in useful ways to support meaningful behavior.

Units can be connected by means of their weights in a variety of ways (figure 22.1a). Feedforward weights connect units in the input layer(s) to units in the hidden layer(s), and units in the hidden layer(s) to units in the output layer(s). Feedback weights may connect the units in the reverse direction (output to hidden to input). Lateral weights connect units to other units in the same layer. Recurrent weights connect units to themselves. In addition to these different directions of connectivity, weights may vary in whether they are excitatory (increasing the input to the receiving unit) or inhibitory (decreasing the input to the receiving unit). Recurrent weights that are excitatory allow units to maintain their activity by continuing to excite themselves. Lateral weights that are inhibitory lead units within a layer to compete with one another for activity, also helping active units to maintain their activity by inhibiting the activity of competing units.

As will be discussed further, "knowledge" in the neural network framework takes the form of patterns of activity across the processing units and patterns of connectivity in the weights. Knowledge is thus embodied in the processing machinery (in contrast with the traditional computer metaphor, in which knowledge structures [RAM] are separable from processing [CPU]). This embodied character of knowledge in the neural network framework makes it a

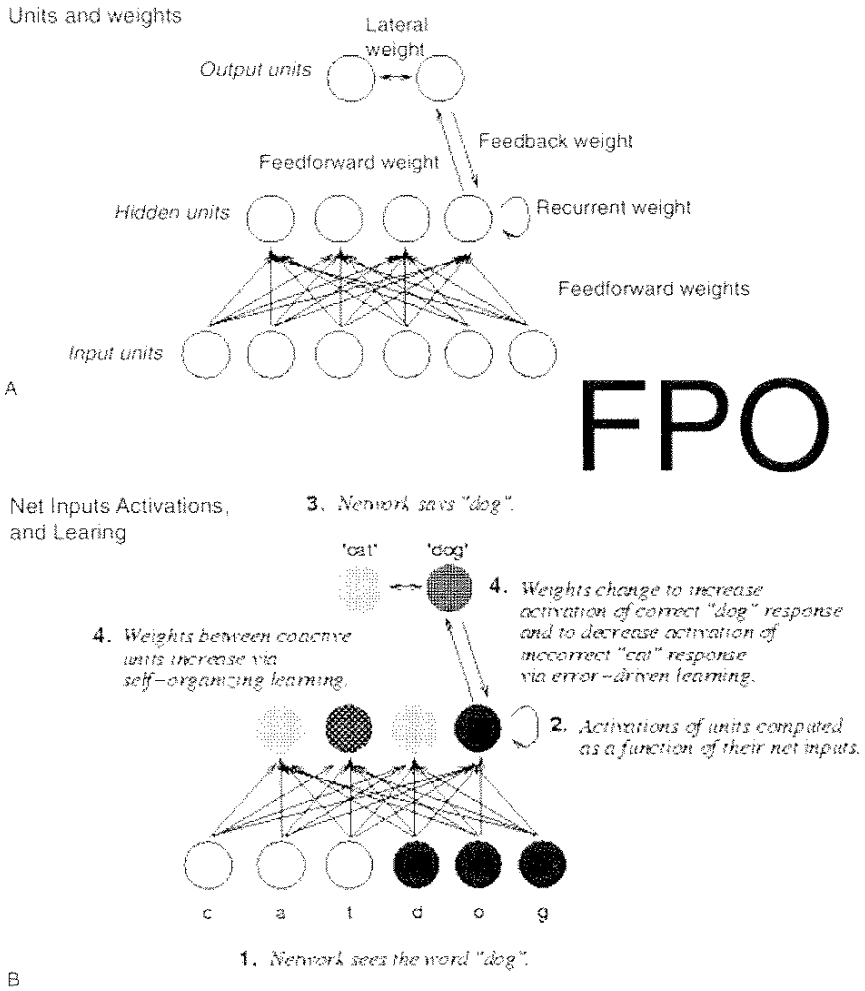

FIGURE 22.1 A diagram of representative neural network architecture (**A**) and processing (**B**). Circles indicate units, and their shading indicates activation levels; arrows indicate weights.

particularly useful methodology for developmental cognitive neuroscience, given the focus of this field on understanding how knowledge is embodied by the brain, and given the parallels between principles of neural communication and relations among units and weights in neural network models.

NET INPUT AND ACTIVATION FUNCTIONS The process of computing a unit's activity is broken down into two steps: computing the net input to the unit and then computing the unit's activity as a function of the net input. The inputs to a unit are weighted by the strength of the connections from the sending units; the stronger the connection, the more the sending unit activity contributes to the net input to the receiving unit. Mathematically, the net input to a unit j (η_j) is expressed as

$$\eta_j = \sum_i w_{ij} a_i$$

where w_{ij} is a weight from unit i to unit j, and a_i is the activity of unit i.

The activation function specifies how the units in a network update their activity as a function of this net input. Activation functions are typically S-shaped (figure 22.2), based on a sigmoidal activation function of the following form:

$$a_j = \frac{1}{1 + e^{-\eta_j}}$$

where a_j is the activation of the unit and η_j is its net input. This S shape reflects two important aspects of neural activity, regarding the nonlinear response of neurons in relation to their inputs. First, the unit is not guaranteed to become active just because it is receiving some amount of input. As indicated by the lower-left part of the S-shaped curve, this net input must get above a certain threshold for the unit to become very active. Second, once the unit is active to some

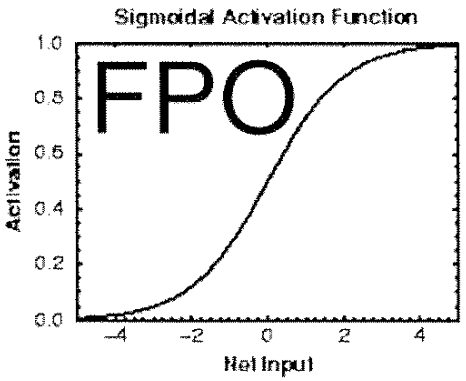

FIGURE 22.2 The sigmoidal activation function, reflecting the nonlinear response of neurons in relation to their inputs.

degree, it is not guaranteed to become much more active with increasing amounts of input. As indicated by the upper-right part of the S-shaped curve, a unit cannot substantially increase its activity level beyond a certain point, even with further net input. This nonlinearity in the activation function allows multiple layers of units to carry out complex computations that are not possible with units using linear activation functions.

LEARNING ALGORITHMS Learning in neural networks takes the form of changes to the weights, which are viewed as corresponding to changes in the efficacy of synapses. Such changes occur as a result of a network's experience with its environment, and they affect how the network responds to subsequent inputs. Because weights may take a value of zero (which is equivalent to no connection), this learning process allows for the possibility of adding new connections (when a zero weight is increased) and pruning away existing connections (when a weight goes to zero) (cf. Shultz, 2006, for more specialized mechanisms for adding and pruning connections). Here, we consider two of the primary types of learning algorithms used in neural network models—self-organizing and error driven.

Self-organizing algorithms are so named because they govern learning without specifying a particular target performance; that is, they lead units to organize their weights themselves based on their local inputs, rather than in terms of meeting particular goals. One of the most common self-organizing algorithms is a Hebbian algorithm (Hebb, 1949), whereby units that are simultaneously active increase the weight between them. Mathematically, the basic form of this learning rule is

$$\Delta w_{ij} = \varepsilon a_i a_j$$

where Δw_{ij} reflects the change in the weight from unit i to unit j, and ε reflects a learning rate parameter. This form of learning has typically been used by modelers focused on biological plausibility (e.g., Miller, Keller, and Stryker, 1989), because the algorithm is grounded in the known biological learning mechanisms of long-term potentiation and long-term depression (Artola, Brocher, and Singer, 1989; Bear and Malenka, 1994). However, the algorithm is not very good at solving complex tasks (whereas humans are), a fact that has led other modelers to turn to more powerful, error-driven algorithms.

Error-driven algorithms are so named because they govern learning based on the discrepancy between a network's performance and its target performance. One of the most common error-driven algorithms is the backpropagation algorithm (Rumelhart, Hinton, and Williams, 1986), whereby the difference between a unit's activity and its target activity is computed and propagated backward through the network, so that the resulting weight changes reduce the unit's error. In this way, the backpropagation algorithm allows networks to learn to solve complex tasks, a necessary criterion for the modeling of human behavior. Mathematically, the backpropagation learning algorithm is

$$\Delta w_{ij} = \varepsilon \delta_j a_i$$

where δ_j reflects the contribution of a given unit to a network's error. Although the backpropagation algorithm has been criticized for being biologically implausible in the details of its implementation (e.g., in the backward propagation of error terms for which there is no neural evidence), biologically plausible versions have been implemented (Hinton and McClelland, 1988; Hinton, 1989; Movellan, 1990; O'Reilly, 1996; O'Reilly and Munakata, 2000). These versions avoid the implausibility problems of backpropagation by indirectly communicating error information through the standard mechanisms of neural communication, the passing of activity signals by means of weights. Further, these activity signals reflect events in the world and networks' expectations regarding the events, such that error information can be computed based on the discrepancies between expectations and outcomes, without requiring an explicit teacher that provides target signals. Such error-driven algorithms thus allow for the continued exploration of simulating performance on complex tasks. Further, the existence of such functionally similar algorithms suggests that models using backpropagation, while biologically implausible in their detailed implementation, should not simply be discounted; lessons from them are likely to prove relevant to the biologically plausible, functionally similar implementations.

As we have shown, learning algorithms can be specified in precise mathematical terms; however, it is important to note that it can nonetheless be difficult to predict exactly how networks will come to solve tasks and how they will develop, given the complex, nonlinear interactions between network units and the environment. Similarly, even with a

precise specification of how synaptic changes occur in the brain, we would not necessarily be able to explain, for example, the complex neural bases of how children learn to read. Understanding changes at the level of the synapse/weight does not translate directly into understanding behavior. Thus, even with a precisely specified learning algorithm, it can be very difficult to predict the behavior of networks of any complexity. We can therefore gain insights into the neural bases of behavior by exploring why networks develop as they do.

How

Now that we have some sense of *why* we might want to use neural network models as a methodology, and *what* they consist of, we are in a position to consider *how* they have contributed to the study of developmental cognitive neuroscience. Neural networks have been used to address many different facets of cognitive development, including individual differences and disorders (Joanisse and Seidenberg, 2003; MacDonald and Christiansen, 2002; Morton and Munakata, 2005; Oliver et al., 2002; Thomas and Karmiloff-Smith, 2002, 2003), constructivist mechanisms of development (Schlesinger and Parisi, 2001; Shultz, 2003, 2006), the coordination of separate specialized brain systems (Jacobs, 1999; Mareschal, Plunkett, and Harris, 1995; Munakata, 2004; Westerman and Miranda, 2004), the influence of early perceptual and motor development on cognition (Jacobs and Dominguez, 2003; Westermann and Mareschal, 2004), and the development of hierarchically organized brain regions (Shrager and Johnson, 1996).

Many neural network models focus more on the role of learning in development than on maturational changes (cf. Shrager and Johnson, 1996). In fact, many aspects of development that may appear to be maturational, such as critical periods, arise in neural networks as a result of learning (Ellis and Lambon-Ralph, 2000; Elman, 1993; McClelland et al., 1999; Rohde and Plaut, 1999; Seidenberg and Zevin, 2006). Similarly, many biological changes that may appear to be hardwired, such as a reduction in the plasticity of synapses across development, have been shown to depend on experience and can be reversed if experience is withheld (E. Quinlan, Olstein, and Bear, 1999).

Here, we focus on neural network explorations of two fundamental issues in cognitive development: the origins of our knowledge and mechanisms of change. We aim to convey an overall sense of how neural network models can speak to these developmental issues, but because of space constraints, we can only briefly cover two examples within each of these areas.

ORIGINS Where does our knowledge come from? Questions of origins (whether of knowledge, life, the universe, etc.) form the basis for some of the most interesting, challenging, and hotly debated issues. In the context of the origins of our knowledge, the debate has taken the form of nature versus nurture, and more recently of specifying the nature of the interactions between them. Neural network models have been used to explore the origins of knowledge in a variety of domains, including language (e.g., Rumelhart and McClelland, 1986; Plunkett and Sinha, 1991; Elman, 1993; Harm and Seidenberg, 1999; Plaut and Kello, 1999; Onnis and Christiansen, 2005), numerical understanding (Dehaene and Changeux, 1993; Verguts and Fias, 2004), and problem solving (McClelland, 1989, 1995; Shultz, Mareschal, and Schmidt, 1994). Here, we focus on models exploring the origins of knowledge of objects, specifically, their continuity and their permanence.

Object continuity. Young infants appear to be sensitive to the continuity of object motion, the fact that objects move only on connected paths, never jumping from one place to another without traveling a path in between. For example, infants as young as 2.5 months look longer at events in which objects appear to move discontinuously than at otherwise similar events in which the same objects move continuously (Spelke et al., 1992). Such longer looking times are taken as an indication that infants find the discontinuous events unnatural, and so possess some understanding of object continuity. What are the origins of such knowledge? Some researchers have concluded that an understanding of object continuity is part of our innate core knowledge, given infants' very early sensitivity to it, and the apparent difficulty in learning such information given that objects are rarely continuously visible in our environment (Spelke et al., 1992). However, as many researchers have noted, it is not clear what the label "innate" really tells us about the nature of the origins of knowledge (Elman et al., 1996; Thelen and Smith, 1994; Smith, 1999). That is, does calling infants' sensitivity to the continuity of object "innate" tell us anything about how infants come to be sensitive to this principle, or about the mechanisms underlying such sensitivity?

In contrast, the neural network approach focuses attention on exactly these kinds of issues, because such mechanisms must actually be implemented in a working model for the account to be considered successful. One such model was devised in the study of imprinting behavior in chicks and of object recognition more generally (O'Reilly and Johnson, 1994, 2002). This model viewed a simplified environment in which objects moved continuously. Based on this experience, the model developed receptive field representations of objects that encoded continuous locations in space, thereby demonstrating a sensitivity to object continuity.

What were the origins of the model's sensitivity to object continuity? First, the network had recurrent excitatory connections and lateral inhibitory connections that allowed

active units to remain active; specifically, active units continued to send activation to themselves by way of the recurrent excitatory connections, and they prevented other competing units from becoming active by way of the lateral inhibitory connections. Thus, when an object was presented as input to the network, certain hidden units became active, and they tended to stay active even as the object moved around in the input. Second, the network learned according to a Hebbian learning rule, which led the model to associate this hidden unit pattern of activity with the object in different locations in the input. Thus, whenever the object appeared in any of these locations, the network came to activate the same units, or the same object representation. In this way, with exposure to events in the world that conformed to the principle of continuity, the model developed receptive field representations of objects that encoded continuous locations in space, and so learned to "recognize" objects that moved continuously in its environment.

One might argue that this model was innately predisposed to understand the continuity of objects (Spelke and Newport, 1997), given that the network was structured "from birth" with recurrent excitatory and lateral inhibitory connections and a Hebbian learning rule—all it needed was the typical experience of viewing objects moving continuously in its environment. That is, the model required experience only in a generic sense, to support an experience-expectant process (Greenough, Black, and Wallace, 1987) that would naturally unfold for all members of a species given the normal environment available throughout evolutionary history. However, again, it is not clear what benefits would be conferred by calling the developmental time course of the model "innate." In contrast, the benefits of the model should be clear in providing an explicit, mechanistic account of the potential origins of our sensitivity to object continuity.

Object permanence. Several models have been proposed to account for infants' apparent sensitivity to the permanence of objects, with very different assumptions about the origins of object-permanence knowledge. At one extreme, such knowledge has been built into a network, with target signals specifying from birth that hidden objects continue to exist when they are hidden (Mareschal, Plunkett, and Harris, 1995). At the other extreme, a model has demonstrated limited sensitivity to the permanence of objects without ever actually developing the ability to represent hidden objects, based on the simple origins of the goal of keeping objects in view (Schlesinger and Barto, 1999). Here, we discuss a model that lies between these two extremes, in which object-permanence knowledge developed without being prespecified (Munakata et al., 1997). The model viewed a simplified environment in which objects disappeared from view behind occluders and reappeared after the occluders were removed. Based on this experience, the model became sensitive to the permanence of objects, continuing to represent objects even after they were hidden.

What were the origins of the model's knowledge of object permanence? As in the object-recognition model described earlier, the object-permanence model had recurrent excitatory connections that allowed active units to remain active. Unlike the object-recognition model, the object-permanence model also had a goal of predicting what would happen next in its environment. Through error-driven learning, the network adjusted its weights if its predictions were incorrect, for example, if the network predicted that an occluded object would not reappear when the occluder was removed, and then the object did in fact reappear. So, when a visible object moved out of view, the network gradually learned to use its recurrent connections to maintain a representation of the object, allowing the network to accurately predict its environment (and the reappearance of such hidden objects). In this way, with exposure to events that conformed to the principle of object permanence, the model provided an explicit, mechanistic account of the potential origins of our sensitivity to the permanence of objects and demonstrated how object-permanence knowledge could develop without being innately specified.

The model also led to the novel prediction that infants should show greater sensitivity to the permanence of familiar objects than of novel objects. The model showed this behavior because it formed stronger representations for familiar objects, based on changes to its connection weights from repeatedly processing those objects. Those changes to the connection weights allowed the model to generalize its knowledge of object permanence to novel objects, but its representations for those novel objects were not as strong as those for familiar objects. This prediction was confirmed in infants, who searched more for familiar objects than for novel objects after they were hidden, despite showing robust preferences for novel objects over familiar objects when they were visible (Shinskey and Munakata, 2005).

CHANGE How does change occur? As many researchers have noted (e.g., Flavell, 1984; Fischer and Bidell, 1991; Siegler, 1989), this question is one of the most fundamental yet unanswered questions in the study of cognitive development. For example, how do children develop complex, higher level cognitive abilities in relatively short periods of time? Why do children sometimes show nonlinear trajectories in their development, such as stagelike progressions, sensitive periods for learning, and U-shaped learning curves? The issue of change is not mutually exclusive from the previously discussed issue of origins. Providing an explicit model of origins entails specifying mechanisms of change (unless the model assumes full-fledged knowledge from the start, an assumption that is inconsistent with the neural network framework and with brain development, not

to mention with the theories of even the most extreme nativists). So all the models described and cited in the previous section also have something to say about change as well as about origins.

In the neural network framework, change can take place at multiple levels, including in the activity of units as activations are propagated through the network, the connection weight changes that occur during learning, and the emergence of new forms that arise from the complex interactions of elements in the network (as described previously for the development of representations of object continuity and permanence). Neural network models have been used to explore the mechanisms underlying many aspects of developmental change, including stagelike progressions (McClelland, 1989, 1995; Raijmakers, van Koten, and Molenaar, 1996; Thomas, 2004; P. Quinlan et al., in press), sensitive periods in learning (Ellis and Lambon-Ralph, 2000; Elman, 1993; McClelland et al., 1999; Rohde and Plaut, 1999; Seidenberg and Zevin, 2006), and U-shaped learning curves (Munakata, 1998; Plunkett, 1991; Rogers, Rakison, and McClelland, 2004; Rumelhart and McClelland, 1986). Here, we focus on models exploring emergent effects in language and conceptual development, specifically how relatively low-level processes can lead to the development of higher-level cognitive abilities and nonlinear developmental trajectories. We consider models exploring how such changes can explain numerous aspects of children's word learning.

Learning about words and semantic categories. How do children learn new words and form appropriate semantic categories for objects in the world? Children show particular behaviors that have led many researchers to posit high-level, conceptual (and possibly innate) structures that guide children's learning; however, neural network models have demonstrated how more basic learning mechanisms could explain these patterns in children's behavior. For example, when learning the name for a new solid object, 2–3-year-olds will reliably extend that name to other solid objects that are similar in shape (a behavior known as the "shape bias"); in contrast, after learning the name for a nonsolid substance, children will extend that name to other nonsolid things that are similar in the material they are made from (the "material bias") (Landau, Smith, and Jones, 1988). Moreover, very early in life, children are able to differentiate abstract semantic categories, such as animals versus artifacts (Mandler and McDonough, 1993). Some researchers explain this behavior in terms of children's early (and possibly innate) high-level, conceptual understanding about different ontological *kinds* of things (Carey, 2000; Keil, 1989), such as animates, objects, and substances (Booth and Waxman, 2002, 2003; Gergely et al., 1995). In contrast, other researchers have used neural network models to explore the possibility that these patterns of behavior could result from more basic learning mechanisms that can extract higher-order regularities, among stimuli with the same labels (e.g., Colunga and Smith, 2005) and among objects from the same category (e.g., Mareschal and French, 2000; McClelland and Rogers, 2003; Quinn and Johnson, 1997; Rogers and McClelland, 2004, 2005).

To test the possibility that simple learning mechanisms could lead to the emergence of word-learning biases, a neural network model was trained to associate perceptual stimuli with their labels, and then tested on its shape and material biases for solids versus nonsolids (Colunga and Smith, 2005). The model was presented with stimuli in terms of perceptual inputs that represented their shape, material, and solidity. Through error-driven learning, the model was trained to produce the correct name for each stimulus on the output layer. The vocabulary the model was trained on captured several aspects of children's vocabulary: the number of words for solids was greater than the number of words for nonsolids, nonsolids had a more restricted range of shapes than solids, and there were strong but imperfect correlations between solid objects and names based on shape, and between nonsolid things and names based on material.

After the network learned how to name 24 stimuli, it was presented with novel solids and nonsolids so that its biases to attend to shape or material could be assessed. With each novel stimulus, the network was also presented with groups of other novel stimuli that were either similar in shape to it or similar in material. The network's shape and material biases were measured in terms of the internal representations the network formed for objects in the hidden layer, which was bidirectionally connected to both the word and the perceptual layers and was also recurrently connected to itself. Like children, the network demonstrated a clear shape bias for solids and a clear material bias for nonsolids. Specifically, the network's internal representations for two solid objects with the same shape but different material were more similar than the internal representation for two solids with different shapes but the same material; the opposite pattern was found for nonsolids. Thus, after simply learning to associate specific words and specific perceptual features, the network formed abstract, generalized expectations about the way different stimuli could be characterized, which correspond to the types of biases observed in young children's word learning. These simulations thus demonstrate how basic low-level learning mechanisms could lead to the development of abstract higher-order generalizations, such that one need not invoke possibly innate, conceptual structures.

Similar basic learning mechanisms may support children's acquisition of semantic categories, such as animals versus plants. Neural networks have demonstrated how such semantic categories can be formed through the learning of statistical structure in the environment (e.g., Mareschal and French, 2000; Quinn and Johnson, 1997), in particular,

through patterns of coherent covariation across objects from the same category (McClelland and Rogers, 2003; Rogers and McClelland, 2004, 2005). When objects have similar representations and share many properties (e.g., all animals move and make sound on their own, while plants do neither), the properties shared by these items will be maximally coherent and will be a strong force driving learning, because they drive changes to connection weights in the same direction. In contrast, idiosyncratic properties (e.g., the fact that some animals can fly but cannot swim, and others do the reverse) drive weights in conflicting directions that tend to cancel each other out early in learning. Overall, this process leads the most coherent properties among categories to be learned earliest, and it can explain how children progress from more coarse to more fine-grained levels of differentiation in their category learning. These coherent properties do not need to be perceptually salient (e.g., the fact that animals can grow); as long as they covary coherently, statistical learning mechanisms can use them to guide category learning. In this way, coherent covariation of properties can lead perceptually distinct items (such as birds and fish) to be viewed as part of the same category, without requiring high-level concepts of animacy.

Moreover, these neural networks also illustrate how nonlinear developmental progressions (such as U-shaped learning curves) can develop simply through sensitivity to statistical regularities. In many complex tasks, such as learning to correctly apply regular or irregular verb past-tense construction, children often seem to "unlearn" a correct behavior (e.g., saying "goed" after correctly saying "went") before eventually achieving complete mastery (Ervin, 1964). These U-shaped patterns of development have elicited explanations in terms of qualitative shifts between different abstract, high-level rule-based systems (Marcus et al., 1992). However, basic learning mechanisms that detect coherent covariation between properties of different objects can also lead to such U-shaped patterns of development (Rogers, Rakison, and McClelland, 2004). For example, the semantic categorization networks described earlier show U-shaped progressions in how they categorize unusual animals, such as bats (which are unusual because they do not have feathers, unlike other exemplars that fly). The networks first correctly categorize bats as animals with fur, then incorrectly characterize them as animals with feathers, and ultimately characterize them correctly again. This U-shaped behavior reflects the coarse-to-fine property of category development. Early on, the networks learn that most animals have fur, thus attribute this coherent characteristic to all animal exemplars, and correctly identify bats as animals with fur. As the networks learn progressively finer distinctions, such as the fact that some animals fly, they again attribute coherent characteristics of flying (e.g., flying animals have feathers) to all flying exemplars. Thus, at this point, the networks incorrectly categorize bats as flying animals with feathers. Finally, as the networks learn idiosyncratic properties of each animal, they become able to correctly identify bats as flying animals with fur and not feathers.

Thus these neural networks show how simple learning mechanisms that pick up on statistical regularities in the environment can lead to complex nonlinear behaviors and to the development of high-level conceptual categories, like those observed in children.

Age-of-acquisition effects. As in the category-learning example, language acquisition is frequently characterized by nonlinear developmental trajectories. In contrast to theories that invoke innate or high-level structures in learning, neural network models have demonstrated how these developmental trends can be explained by relatively simple, low-level mechanisms. Another example of nonlinear change in language acquisition concerns the age of acquisition (AoA) effect, a phenomenon in which words learned early in life are recognized and pronounced faster than words learned later in life. Such patterns might suggest the existence of specialized, time-sensitive learning systems. However, the age at which children learn particular words is typically confounded with word frequency and word length, making it difficult to interpret AoA effects and their implications for how children learn (Zevin and Seidenberg, 2002). Neural network modeling, however, allows for factors such as cumulative frequency and frequency trajectory of words to be manipulated independently of one another, thereby permitting a detailed analysis of each factor's potential impact on lexical development.

To assess whether words learned early in life might enjoy processing benefits not shared by words learned later, a series of recurrent backpropagation neural networks were trained to read nearly 3,000 monosyllabic words derived from natural language corpora (Zevin and Seidenberg, 2002). The networks consisted of three feedforward layers: an orthographic input layer, a hidden layer, and a phonological output layer. A fourth layer, bidirectionally connected with the phonological layer, served to improve the accuracy of the network's phonological outputs. Training consisted of 1 million word presentations, in which the network might receive an input like "FIST" and would then be required to produce the phonemes corresponding to that word. The frequency of some words in the training set was manipulated, such that some words were most frequently presented early in training, while other words were most frequently presented later in training. Critically, the network's cumulative exposure to words on both the "early" and the "late" lists was equivalent by the end of training.

In these models, age of acquisition was predicted by frequency trajectory, such that words from the "early" list were learned more quickly than those in the "late" lists, which

were less prevalent at the beginning of training. However, by the end of training, both lists were learned equally well, and accuracy was at ceiling, thus yielding no lasting AoA effect. Why should this outcome occur? During training, the network learned to extract regularities in the orthography-to-phonology mappings that are characteristic of English. As a result, any benefits conveyed by learning a word early were also passed on to words learned later in training. This finding provides support to the idea that the AoA effects seen in behavioral research may result from the confound of cumulative word frequency with age of acquisition.

However, an extreme reduction in the similarity between the orthography-to-phonology mappings of "early" list words to "late" list words did yield a reliable AoA effect. In this case, the regularities extracted by the network based on the "early" list—and the resulting changes in connection weights—actually *dis*advantaged the learning of words with a very different orthography-to-phonology mapping. The network was never able to produce the later words as accurately as it had learned the earliest words, because the regularities extracted by early learning could not be passed on to words prevalent later in training. In effect, the connection weights in the network became specialized for representing the early set of orthography-to-phonology mappings; despite later experience with a very different set of mappings, this early specialization could never be completely overcome.

In this case, neural network models allowed for an examination of linguistic factors that are normally very difficult to dissociate: cumulative word frequency and frequency trajectory. The results suggested that for natural languages with reliable orthography-to-phonology mappings, cumulative frequency should influence ultimate levels of skilled reading, and frequency trajectory should affect age of acquisition without any lasting AoA effects on ultimate levels of skilled reading. In other words, children should first acquire those words they encountered with the highest frequency. However, this early learning will convey benefits to many words experienced later, such that this later learning is also facilitated by the child's earlier experience. This process serves to wash out age-of-acquisition effects, such that words acquired later share the same processing benefits enjoyed by words acquired at earlier ages. Children should only show a disproportionate advantage for producing words learned early in life if they are experienced with greater total frequency.

These predictions were subsequently confirmed in behavioral research where cumulative frequency and frequency trajectory were explicitly dissociated from other characteristics that might influence the ease with which words are pronounced (Zevin and Seidenberg, 2004). Ultimately, these models allowed for an initial investigation of the factors underlying purported AoA effects before they were cleanly dissociated in a behavioral experiment. More specifically, they permitted direct insight into how, and in what particular situations, early experience with language might result in lasting effects on linguistic behavior.

SUMMARY OF *HOW* NEURAL NETWORK MODELS HAVE CONTRIBUTED As preceding sections have illustrated, relatively basic principles of neural network modeling can have profound implications for a variety of developmental questions, including questions of the origins of knowledge and mechanisms of developmental change. As we have seen, neural network models demonstrate how knowledge of object permanence and object continuity—knowledge that is sometimes considered to be innate—can actually arise naturally from an interaction between early life experiences and the basic beginning state of a system (e.g., in terms of initial excitatory and inhibitory connectivity). We have also described models that reproduce specific features of language learning, including word-learning biases and nonlinear developmental trajectories, using simple learning mechanisms that pick up on statistical regularities in the environment. The relevance of neural network modeling to such vastly different phenomena is a testament to this framework's flexibility and importance. Furthermore, these models have demonstrated how specific neural mechanisms can account for a variety of developmental phenomena, in addition to generating testable (and subsequently confirmed) predictions about children's behavior.

Challenges to the why, what, and how

As in all active areas of science, each of the aspects of neural network modeling that we have discussed has been challenged in some way. Here, we focus on one important criticism within each of the areas of why, what, and how (see also discussion in Elman et al., 1996; McClelland and Plaut, 1998; O'Reilly and Munakata, 2000; Seidenberg 1993; Seidenberg and Zevin, 2006).

CHALLENGES TO WHY MODELS ARE IMPORTANT A common criticism of neural network models is that they can do anything, solve any task, and so on; therefore, their ability to simulate human behavior is uninteresting. That is, there are so many parameters that can be manipulated in a network that it is guaranteed to work eventually. Because getting it to work is guaranteed, this process tells us nothing. Further ammunition for this criticism comes from the fact that several *different* neural network models may succeed in simulating the same human behavior. They all work, and yet they can't all be right, indicating that neural networks are simply too powerful, so a successful simulation proves nothing.

Before countering this criticism using specific examples of neural network models, we first emphasize a general

response: *Criticisms about too much power and too many parameters are relevant to any attempts at scientific theorizing, and are not unique to the neural network modeling endeavor.* One could easily level the same criticisms at verbal theories of behavior, for example. Across a range of domains (attention, memory, language, etc.), multiple competing theories can account for the same behavioral data. And, these verbal theories are typically powerful enough to encompass any new piece of behavioral data that comes along, thanks to the vagueness of constructs and the existence of multiple free parameters (in the form of new limitations or capabilities that can be incorporated into the theory). Thus verbal theories can be constructed to explain anything, and multiple competing theories can account for the same data, so the process of developing theories tells us nothing. Most people probably would not accept this conclusion in the domain of scientific theorizing, and yet many believe it to pose a fundamental problem for neural network models.

We believe that the counterargument to this criticism as applied to neural network models is similar to the counterargument to this criticism as applied to scientific theorizing more generally. *Competing theories and models can be evaluated by many criteria other than simply by accounting for a set of data.* People generally know when a theory feels unsatisfying, even if it is able to account for some data. For example, if a theory needs to add a new component to account for each new piece of data, it will seem more arbitrary than a more unified theory that requires no such adjustments. Or if a theory accounts for data by relying on unspecified constructs, it will seem less compelling than a more fully specified theory. In this way, the plausibility and specificity of underlying assumptions, as well as the ease with which data can be accounted for and predicted, can be evaluated to compare competing theories. The same holds true for evaluating competing models. The neural network framework may support relatively rapid progress along these lines, because the models require the underlying assumptions to be made explicit and because the assumptions are constrained by both bottom-up (biological) and top-down (psychological) information.

A second counterargument to this criticism is that the number of parameters in neural network simulations may accurately reflect the diversity of underlying mechanisms that contribute to behavior, so that neural network models provide a useful tool for exploring these mechanisms. For example, in the context of developmental disorders, the same behavioral deficit may arise from any of a number of distinct underlying causes (Thomas, 2003). Moreover, this problem is not specific to disorders; *any* group of individuals may behave similarly and yet differ in how those behaviors are produced. The capacity of neural network models to simulate this phenomenon can thus be viewed as an important strength. Such models allow us to formally analyze multiple causality in a way that is not possible with purely behavioral measures, by allowing us to independently manipulate and assess the factors contributing to emergent behavior.

Finally, it is important to note that many neural network models have made their contributions by *not* working, that is, by not simply simulating a particular behavior that they were designed to simulate. For example, neural network models have been lesioned to simulate (and provide insight into) the behavior of patients with brain damage (Plaut et al., 1996; Farah, O'Reilly, and Vecera, 1993; Cohen et al., 1994; Farah and McClelland, 1991; Plaut, 1995; Allen and Seidenberg, 1999; Heinke and Humphreys, 2003). Such lesions are performed by removing or damaging units or their connections. In these cases, the models are *not* trained to simulate such atypical performance. Instead, the models are trained to perform correctly, and then they are lesioned. Altered patterns of performance emerge from the basic properties of the models following damage. Lesions or other alterations can also be performed during the course of models' development; such models have elucidated the possible causes of various developmental disorders (Thomas and Karmiloff-Smith, 2002, 2003; Thomas, 2003; Triesch et al., 2006; Williams and Dayan, 2005). In addition, failures of neural network models (e.g., in remembering both specific events and generalizing across multiple events) have provided insight into neural divisions of labor (e.g., between the hippocampus and neocortex) (McClelland, McNaughton, and O'Reilly, 1995).

Thus neural network models, though powerful, can nonetheless fail and can provide insights when they do, and they (like purely verbal theories of behavior) can be evaluated on grounds other than simply accounting for a set of data.

Challenges to What Models Comprise Many challenges have been issued regarding the nuts and bolts that we have described, specifically how well the various elements of neural models map onto elements in the brain. Critics argue that the elements of models are simplistic, missing essential aspects of neural communication that render their use misguided at best. We believe that there are no quick and definitive answers to this challenge, but rather preliminary responses to be further tested and elaborated over the coming years, as part of important progress in the neural network framework.

One response is simply, "Simple is good." That is, the simplified elements of neural network models capture the essential aspects of neural communication, thereby providing a critical methodological tool for exploring the complexities of the relation between brain and behavior. We would otherwise get bogged down in details not particularly relevant to understanding cognition. An analogy may be found in the technique of creating mosaic images from a large collection of smaller images. The details of each of the smaller

images (one is a flower, another is a landscape, etc.) are not particularly relevant, and in fact, one could easily lose sight of the point of the image by focusing on these details. Instead, it is more appropriate to stand back and see the overall image at a simplified level. Neural networks may similarly provide a useful simplification of details, to allow an understanding of neural function and its relevance to cognition. For example, we can understand the efficacy of a synapse in terms of a simplified, single value of a connection weight, much as we can understand a collection of small images in terms of the simplified, overall mosaic. Without such a simplification, we might otherwise get bogged down in all the details of how synaptic efficacies are determined at the biological level (the number of vesicles of neurotransmitter released by the presynaptic neuron, the alignment and proximity of release sites and receptors, the efficacy of channels on the postsynaptic neuron, etc.). This degree of biological detail might cloud the picture of the brain-behavior relation, which is instead clarified by the simplifications of the neural network framework.

Of course, the simple-is-good argument assumes that the neural network simplifications capture the essential computational properties of the biological details. This assumption can be tested by including further details into models and exploring their computational significance. In addition, the appropriateness of the simplifications can be tested by developing models at more than one level of complexity. For example, one simplification in most neural network models is the units' continuous-valued activation term (computed from the net input as described in the "What" section), meant to approximate the rate of firing of discrete spikes. Comparisons of this simplification with more detailed models that actually fire discrete spikes have indicated that the continuous-valued activations do in fact closely approximate the firing rates of the more detailed models (O'Reilly and Munakata, 2000).

Thus the simplified nature of neural network models may allow for a clearer picture of the brain-behavior relation, and the validity of these simplifications can be tested by developing models at more than one level of complexity and by testing the functional relevance of biological details.

Challenges to How Models Have Contributed to Developmental Cognitive Neuroscience Various aspects of the specific models we have elaborated, together with their associated claims, have been challenged (e.g., Baillargeon and Aguiar, 1992; Marcus, 1998; Smith et al., 1999; Stadthagen-Gonzalez, Bowers, and Damian, 2004; Ghyselinck, Lewis, and Brysbaert, 2004; Booth, Waxman, and Huang, 2005). In general, we believe that many of these challenges will lead to progress in developing better models. Further, the issuing of such challenges points to a strength of the modeling framework—instantiated models can be subsequently tested on a range of measures, highlighting their potential limitations and suggesting necessary elaborations and revisions, as well as suggesting critical empirical tests to contrast competing models. Here we focus on one criticism that has been applied to a range of models, namely, their failure to generalize (Marcus, 1998; Pinker and Prince, 1988).

According to this criticism, neural network models may mimic some aspects of human performance, but the bases for human and network behavior differ vastly. Specifically, humans use rules to govern their behavior (e.g., to form the past tense of most words, add "ed"), and so they can generalize to new instances (e.g., to know that the past tense of "blicket" must be "blicketed"). In contrast, neural network models use associations to govern their behavior (e.g., "walk" is associated with "walked"), and so they cannot generalize to new instances. Therefore, although neural networks may mimic certain aspects of human performance across a range of domains, these models fail to generalize to new instances in these domains in the ways that humans can, indicating a fundamental limitation to the models.

We discuss three responses to this generalization criticism (see also McClelland and Plaut, 1999; Munakata and O'Reilly, 2003; Seidenberg and Elman, 1999). The first two responses suggest that the discrepancy between human and network generalization has been exaggerated, and the third response highlights important mechanisms for generalization in the neural network framework.

First, it is not clear how much of cognition is driven by rules as we have just described. Although one might sometimes be able to characterize a person's behavior in terms of rules, this fact does not mean that those rules are explicitly instantiated in and consulted by the person (McClelland and Plaut, 1999; McClelland, 1989; Rumelhart and McClelland, 1986; Munakata et al., 1997; Thelen and Smith, 1994). This point is particularly relevant for developmental cognitive neuroscience, where test populations are often preverbal, nonverbal, or limited in their linguistic skills, and therefore unable to explicitly indicate whether they are in fact using rules to govern their behavior. Thus the assumption that rules govern behavior, and that the neural network framework must therefore incorporate rules to be considered valid, is questionable.

Nonetheless, with or without rules, humans are certainly able to generalize their knowledge to new instances, so failures of neural network models to do so would seem damning. However, the claim that neural networks cannot generalize to new instances has been based predominantly on misguided testing methods (Marcus, 1998). In such tests, neural networks are trained on a particular task, but one set of input units are never activated during this training. At test, those units are activated for the first time, and the network is tested on its ability to generalize what it has learned from training

(i.e., to respond appropriately to these new instances). This type of test is misguided for two reasons. First, this test assumes that when we are presented with new instances (e.g., the word "blicket"), this action activates neurons in our brains that have never fired before. No evidence supports this assumption. Instead, extensive evidence indicates that neural patterns of firing reflect the similarity of inputs (e.g., Desimone and Ungerleider, 1989; Tanaka, 1996), suggesting that generalization to new instances would occur through the overlap between patterns of firing to the new instances and patterns of firing in previous experiences. Second, as described in the "What" section, the basic nuts and bolts of neural network (and neural) processing dictate that units must become active to support learning and meaningful behavior. Therefore, it is not particularly informative to run simulations to test the performance of units that have never become active. In sum, no evidence supports the idea that generalizing an existing ability to a new stimulus involves the activation of a pool of never-fired neurons and that neurons must become active to support learning. However, tests of network generalization have assumed such never-fired pools and consisted of testing the performance of units that have never become active.

Under more plausible testing conditions, networks can generalize to new instances. For example, networks can be presented with a set of stimuli and then tested on their ability to generalize to new instances that activate novel combinations of units that have been active before. Neural networks have been shown to generalize to new instances under such circumstances across a range of domains (Colunga and Smith, 2005; Hinton, 1986; Munakata et al., 1997; O'Reilly and Munakata, 2000; Plaut et al., 1996; Rougier et al., 2005; cf. Marcus, 1998). A key factor in networks' successful generalization (and presumably in humans' as well) is the overlap in representations, or the extent to which a new instance is represented in a way that overlaps with previously experienced instances, guiding how to respond to the new instance. Importantly, this overlap may be present in the input-level representation to the network (e.g., as one might expect in the auditory input patterns for the new instance of "blicket" and the familiar instance of "picket") or in higher-level representations of the input (e.g., in patterns of activity indicating that a word is a verb). Such higher-level representations can function like categories, such that once a new instance is represented appropriately at these higher levels, the network can generalize all its knowledge about the category (verbs, males, objects, etc.) to the new instance. In this way, the learning mechanisms that build on associations in neural network models support more than simple stimulus-response kinds of learning; higher-level representations allow stimuli to be encoded in more abstract and meaningful ways. Further progress in this area will likely depend upon the exploration of the factors that influence networks' abilities to form systematic representations at appropriate levels of abstraction, which can then be used to support meaningful generalizations across different tasks.

Conclusions

In this chapter we have considered why neural network modeling is an important methodology for developmental cognitive neuroscience, what neural network models are, and how neural network models have contributed to addressing two fundamental issues in the study of development—the origins of knowledge and how change occurs. In addition, we have covered criticisms of neural network modeling within each of these areas of why, what, and how. In this section we will briefly review how models have offered a unique opportunity to gain insight into cognitive development. We will close with thoughts about the most productive avenues for future work in neural network modeling.

As described in the "Why" section, models provide many potential advantages, including (1) allowing control, (2) helping us to understand behavior, (3) dealing with complexity, and (4) being explicit. All the models in this chapter tap each of these advantages; here, we highlight one example for each of these advantages. First, the ability to control the frequencies of words that a model was exposed to provided insight into sources of apparent age-of-acquisition effects in children's word learning (Zevin and Seidenberg, 2002). This ability to manipulate the training environment in such a controlled manner and to observe the long-term effects on language learning is unique to the modeling framework. Second, the ability to watch representations develop in a model provided an understanding of how children might progress from more coarse to more fine-grained semantic categories and how this process could lead to U-shaped patterns of development (McClelland and Rogers, 2003; Rogers & McClelland, 2004, 2005; Rogers, Rakison, and McClelland, 2004). This ability to watch learning unfold in networks can help us to understand behavior at a more mechanistic level than would otherwise be possible. Third, the ability to deal with complexity allowed a model to provide a principled account of the potential origins of infants' sensitivity to object continuity (O'Reilly and Johnson, 1994, 2002). A purely verbal description of the complex process of developing receptive fields that encode continuous locations in space would probably appear vague; the model instead shows how this process can emerge naturally in a network. Finally, the need to be explicit about various assumptions in implementing a working model led to the deconstruction of the object permanence concept into specific learning mechanisms and resulting representations (Munakata et al., 1997) and motivated novel behavioral predictions that were subsequently confirmed (Shinskey and Munakata, 2005). Without the forcing function of

explicitness found in the modeling framework, such constructs often remain only black boxes in purely verbal theoretical accounts.

Of course, all these advantages of the neural network modeling methodology rely on the existence of careful empirical studies, which lay out the important phenomena to be addressed and help test competing models. Models cannot stand alone and are meant to be put forth as complementary (rather than superior) to empirical studies, for the reasons elaborated previously. While this point may seem obvious, some criticisms of modeling have seemed to assume that the modeling methodology must be held to a higher standard than empirical work. Specifically, one criticism is that each parameter is not varied and systematically tested in neural network modeling, so that it can be hard to know which parameters are crucial to a network's behavior (McCloskey, 1991; Mandler, 1998). However, the same criticism can be applied to empirical work. Typically the parameter of interest (e.g., delay in a memory task) is varied and its effects measured. Other parameters (e.g., the size of the testing room) are viewed as less relevant and are not varied. In both modeling and empirical methodologies, further progress can be made by subsequently testing such assumptions about which factors are relevant. Such progress has been made more rapidly with empirical methodologies, because the same testing paradigms are often employed by multiple different researchers, helping to isolate which factors are relevant to behavior. As the field of modeling continues to develop, with new models replicating and building on prior models, similar progress in isolating critical factors should result.

This argument brings us to our final point, which focuses on the most productive way to proceed with neural network modeling as a methodology. We believe it will be most fruitful if researchers appreciate both the strengths and the limitations of neural network models (and recognize that some of the limitations are equally applicable to empirical work and to verbal theorizing), such that subsequent models can be developed that build on the strengths and begin to address the limitations. Although, again, this point may seem obvious, the field has tended to miss this kind of balance, instead oscillating between extreme hype (models should be fully accepted simply because they work) and extreme skepticism (models should be completely rejected simply because someone shows some limitation in them). As a caution against extreme hype, we have emphasized specific contributions from neural network models to our understanding of the processes of cognitive development (not simply touting the fact that a model works), and we have tried to underscore the need to evaluate models (like theories) on a range of criteria other than simply working. As a caution against extreme skepticism, we note that all models involve simplifications and, in turn, limitations, so it is not particularly constructive to simply point out limitations and argue that models should thus be discounted. Rather, it will be most productive if an understanding of limitations can support the development of alternative models, which can then be evaluated on similar grounds. Again, it may be useful to consider the parallels with more traditional empirical work and verbal theorizing. Researchers rarely critique theories without providing alternatives or run studies simply to disprove others' theories. Rather, researchers typically put forth alternate theories to account for the data, theories that are on the same playing field as the original theories, equally susceptible to criticism, testing, and so on. We believe that this same process would greatly benefit progress in the modeling endeavor. That is, we will make the most progress by specifying alternative models that build on existing strengths and begin to address limitations. In this way, better models will be developed that tap the unique advantages of this methodology, continuing to advance our understanding of developmental cognitive neuroscience.

ACKNOWLEDGMENTS Preparation of this chapter was supported by research grants from NIMH (MH59066-01), NICHD (HD37163), and NSF (IBN-9873492). We thank Eliana Colunga, Randy O'Reilly, Rob Roberts, Marshall Haith, and members of the Cognitive Development Center for useful comments and discussions.

REFERENCES

ALLEN, J., and M. S. SEIDENBERG, 1999. The emergence of grammaticality in connectionist networks. In B. MacWhinney, ed., *The Emergence of Language*, 115–151. Mahwah, NJ: Lawrence Erlbaum.

ARTOLA, A., S. BROCHER, and W. SINGER, 1989. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, 347:69–72.

BAILLARGEON, R., and A. AGUIAR, 1998. Toward a general model of perseveration in infancy. *Dev. Sci.* 1:190–191.

BEAR, M. F., and R. C. MALENKA, 1994. Synaptic plasticity: LTP and LTD. *Cur. Opin. Neurobiol.* 4:389–399.

BOOTH, A., and S. WAXMAN, 2002. Word learning is "smart": Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition* 84:B11–B22.

BOOTH, A., and S. WAXMAN, 2003. Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives. *J. Cogn. Dev.* 4:357–381.

BOOTH, A., S. R. WAXMAN, and Y. T. HUANG, 2005. Conceptual information permeates word learning in infancy. *Dev. Psychol.* 41(3):491–505.

CAREY, S., 2000. The origins of concepts. *J. Cogn. Dev.*, 1:37–42.

CASE, R., 1985. *Intellectual Development: A Systematic Reinterpretation*. New York: Academic Press.

COHEN, J. D., R. D. ROMERO, M. J. FARAH, and D. SERVAN-SCHREIBER, 1994. Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *J. Cogn. Neurosci.* 6(4):377–387.

COLUNGA, E., and L. B. SMITH, 2005. From the lexicon to expectations about kinds: A role for associative learning. *Psychol. Rev.* 112:347–382.

DEHAENE, S., and J.-P. CHANGEUX, 1993. Development of elementary numerical abilities: A neuronal model. *J. Cogn. Neurosci.* 5:390–407.

DESIMONE, R., and L. G. UNGERLEIDER, 1989. Neural mechanisms of visual processing in monkeys. In F. Boller and J. Grafman, eds., *Handbook of Neuropsychology*, vol. 2, chap. 14, pp. 267–299. New York: Elsevier Science.

ELLIS, A. W., and M. LAMBON-RALPH, 2000. Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *J. Exp. Psychol.* [*Learn. Mem. Cogn.*] 26:1103–1123.

ELMAN, J. L., 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48(1):71–79.

ELMAN, J., E. BATES, A. KARMILOFF-SMITH, M. JOHNSON, D. PARISI, and K. PLUNKETT, 1996. *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press.

ERVIN, S., 1964. Imitation in children's language. In E. H. Lenneberg, ed., *New Directions in the Study of Language*, 163–198. Cambridge, MA: MIT Press.

FARAH, M. J., and J. L. MCCLELLAND, 1991. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *J. Exp. Psychol.* [*Gen.*] 120:339–357.

FARAH, M. J., R. C. O'REILLY, and S. P. VECERA, 1993. Dissociated overt and covert recognition as an emergent property of a lesioned neural network. *Psychol. Rev.* 100:571–588.

FISCHER, K. W., and T. BIDELL, 1991. Constraining nativist inferences about cognitive capacities. In S. Carey and R. Gelman, eds., *The Epigenesis of Mind*, chap. 7, pp. 199–236. Hillsdale, NJ: Lawrence Erlbaum.

FLAVELL, J. H., 1984. Discussion. In R. J. Sternberg, ed., *Mechanisms of Cognitive Development.* New York: Freeman.

GERGELY, G., Z. NADASDY, G. CSIBRA, and S. BIRO, 1995. Taking the intentional stance at 12 months of age. *Cognition*, 56: 165–193.

GREENOUGH, W. T., J. E. BLACK, and C. S. WALLACE, 1987. Experience and brain development. *Child Dev.* 58:539–559.

GHYSELINCK, M., M. B. LEWIS, and M. BRYSBAERT, 2004. Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychol. (Amst.)* 115:43–67.

HARM, M. W., and M. S. SEIDENBERG, 1999. Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychol. Rev.* 106:491–528.

HARM, M. W., and M. S. SEIDENBERG, 2004. Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111:662–720.

HEBB, D. O., 1949. *The Organization of Behavior.* New York: John Wiley and Sons.

HEINKE, D., and G. W. HUMPHREYS, 2003. Attention, spatial representation and visual neglect: Simulating emergent attentional processes in the Selective Attention for Identification Model (SAIM). *Psychol. Rev.* 110:29–87.

HINTON, G. E., 1986. Learning distributed representations of concepts. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, 1–12. Hillsdale, NJ: Lawrence Erlbaum.

HINTON, G. E., 1989. Deterministic Boltzmann learning performs steepest descent in weight-space. *Neural Comput.* 1:143–150.

HINTON, G. E., and J. L. MCCLELLAND, 1988. Learning representations by recirculation. In D. Z. Anderson, ed., *Neural Information Processing Systems, 1987*, 358–366. New York: American Institute of Physics.

INHELDER, B., and J. PIAGET, 1958. *The Growth of Logical Thinking from Childhood to Adolescence.* New York: Basic Books.

IVANCHENKO V., and J. RA, 2003. A developmental approach aids motor learning. *Neural Comput.* 15(9):2051–2065.

JACOBS, R. A., 1999. Computational studies of the development of functionally specialized neural modules. *Trends Cogn. Sci.* 3:31–38.

JACOBS R. A., and M. DOMINGUEZ, 2003. Visual development and the acquisition of motion velocity sensitivities. *Neural Comput.* 15(4):761–781.

JOANISSE, M., and M. S. SEIDENBERG, 2003. Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain Lang.* 86:40–56.

KEIL, F., 1989. *Concepts, Kinds, and Cognitive Development.* Cambridge, MA: MIT Press.

LANDAU, B., L. B. SMITH, and S. S. JONES, 1988. The importance of shape in early lexical learning. *Cogn. Dev.* 3(3):299–321.

MACDONALD, M., and M. CHRISTIANSEN, 2002. Reassessing working memory: A reply to Just and Carpenter and Waters and Caplan. *Psychol. Rev.* 109:35–54

MANDLER, J. M., 1998. On theory and modeling. *Dev. Sci.* 2(1):196–197.

MANDLER, J. M., and L. MCDONOUGH, 1993. Concept formation in infancy. *Cogn. Dev.* 8:291–318.

MARCUS, G. F., 1998. Rethinking eliminative connectionism. *Cogn. Psych.* 37:243.

MARCUS, G. F., S. PINKER, M. ULLMAN, M. HOLLANDER, T. J. ROSEN, and F. XU, 1992. Overregularization in language acquisition. *Monogr. Soc. Res. Child Dev.* 4 (Serial No. 228).

MARESCHAL, D., and R. M. FRENCH, 2000. Mechanisms of categorization in infancy. *Infancy* 1:59–76.

MARESCHAL, D., K. PLUNKETT, and P. HARRIS, 1995. Developing object permanence: A connectionist model. *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, 170–175. Hillsdale, NJ: Lawrence Erlbaum.

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, UK: Oxford University Press.

MCCLELLAND, J. L., 1995. A connectionist perspective on knowledge and development. In T. J. Simon, and G. S. Halford, eds., *Developing Cognitive Competence: New Approaches to Process Modeling*, 157–204. Hillsdale, NJ: Erlbaum.

MCCLELLAND, J. L., B. L. MCNAUGHTON, and R. C. O'REILLY, 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. *Psychol. Rev.* 102:419–457.

MCCLELLAND, J., and D. PLAUT, 1999. Does generalization in infant learning implicate abstract algebra-like rules? *Trends Cogn. Sci.* 283~~77–80~~.

MCCLELLAND, J. L., and T. T. ROGERS, 2003. The parallel distributed processing approach to semantic cognition. *Nature Rev. Neurosci.*, 4:310–322.

MCCLELLAND, J. L., A. THOMAS, B. D. MCCANDLISS, and J. A. FIEZ, 1999. Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. *Prog. Brain Res.* 121:75–80.

MCLOSKEY, M. 1991. Networks and theories: The place of connectionism in cognitive science. *Psychol. Sci.*, 2(6):387–395.

Miller, K. D., J. B. Keller, and M. P. Stryker, 1989. Ocular dominance column development: Analysis and simulation. *Science* 245:605–615.

Morton, J. B., and Y. Munakata, 2005. What's the difference? Contrasting modular and neural network approaches to understanding developmental variability. *J. Dev. Behav. Pediatr.* 26: 128–139.

Movellan, J. R., 1990. Contrastive Hebbian learning in the continuous Hopfield model. In D. S. Touretzky, G. E. Hinton, and T. J. Sejnowski, eds., *Proceedings of the 1989 Connectionist Models Summer School*, 10–17. San Mateo, CA: Morgan Kaufman.

Munakata, Y., 1998. Infant perseveration and implications for object permanence theories: A PDP model of the AB task. *Dev. Sci.* 1:161–184.

Munakata, Y., 2004. Computational cognitive neuroscience of early memory development. *Dev. Rev.* 24:133–153.

Munakata, Y., J. L. McClelland, M. H. Johnson, and R. Siegler, 1997. Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychol. Rev.*, 104(4):686–713.

Munakata, Y., and R. C. O'Reilly, 2003. Developmental and computational neuroscience approaches to cognition: The case of generalization. *Cogn. Stud.* 10:76–92.

Newport, E. L., 1988. Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Lang. Sci.* 10:147–172.

Newport, E. L., 1990. Maturational constraints on language learning. *Cogn. Sci.* 14:11–28.

Oliver, A., M. H. Johnson, A. Karmiloff-Smith, and B. Pennington, 2002. Deviations in the emergence of representations: A neuroconstructivist framework for analysing developmental disorders. *Dev. Sci.* 3:8–40.

Onnis, L., and M. H. Christiansen, 2005. Happy endings for absolute beginners: Psychological plausibility in computational models of language acquisition. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, 1678–1683. Mahwah, NJ: Lawrence Erlbaum.

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

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

O'Reilly, R. C., and M. H. Johnson, 2002. Object recognition and sensitive periods: A computational analysis of visual imprinting. In M. H. Johnson, R. O. Gilmore, and Y. Munakata, eds., *Brain Development and Cognition: A Reader*, 2nd ed. Oxford, UK: Blackwell.

O'Reilly, R. C., and Y. Munakata, 2000. *Computational Explorations in Cognitive Neuroscience*. Cambridge, MA: MIT Press.

Piaget, J., 1952. *The Origins of Intelligence in Childhood*. New York: International Universities Press.

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

Plaut, D. C., 1995. Double dissociation without modularity: Evidence from connectionist neuropsychology. *J. Clin. Exp. Neuropsychol.* 17(2):291–321.

Plaut, D., and C. Kello, 1999. The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney, ed., *The Emergence of Language*, 381–415. Mahwah, NJ: Lawrence Erlbaum.

Plaut, D. C., J. L. McClelland, M. S. Seidenberg, and K. E. Patterson, 1996. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychol. Rev.* 103:56–115.

Plunkett, K., and C. Sinha, 1991. Connectionism and developmental theory. *Psykologisk Skriftserie Aarhus* 16(1):1–34.

Quinlan, E. M., D. H. Olstein, and M. F. Bear, 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.

Quinlan, P. T., H. L. J. van der Maas, B. R. J. Jansen, O. Booij, and M. Rendell, in press. Rethinking stages of cognitive development: An appraisal of connectionist models of the balance scale task. *Cognition.*

Quinn, P. C., and M. H. Johnson, 1997. The emergence of category representations in infants: A connectionist analysis. *J. Exp. Child Psychol.* 66:236–263.

Raijmakers, M. E., S. van Koten, and P. C. Molenaar, 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. *Cogn. Sci.* 20:101–136.

Rogers, T. T., and J. L. McClelland, 2004. *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.

Rogers, T. T., and J. L. McClelland, 2005. A parallel distributed processing approach to semantic cognition: Applications to conceptual development. In L. Gershkoff-Stowe and D. Rakison, eds., *Building Object Categories in Developmental Time: 32nd Carnegie Symposium on Cognition*. Mahwah, NJ: Erlbaum.

Rogers, T. T., D. H. Rakison, and J. L. McClelland, 2004. U-shaped curves in development: A PDP approach. *J. Cogn. Dev.* 5:137–145.

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

Rougier, N. P., D. Noelle, T. S. Braver, J. D. Cohen, and R. C. O'Reilly, 2005. Prefrontal cortex and the flexibility of cognitive control: Rules without symbols. *Proc. Natl. Acad. Sci. USA* 102:7338–7343.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986. Learning representations by back-propagating errors. *Nature* 323:533–536.

Rumelhart, D. E., and J. L. McClelland, 1986. PDP models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, and PDP Research Group, eds., *Parallel Distributed Processing*, vol. 1: *Foundations*, chap. 4, pp. 110–146. Cambridge, MA: MIT Press.

Schlesinger, M., and A. Barto, 1999. Optimal control methods for simulating the perception of causality in young infants. *Proceedings of the 21st Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum.

Schlesinger, M., and D. Parisi, 2001. The agent-based approach: A new direction for computational models of development. *Dev. Rev.*, 21:121–146.

Seidenberg, M., 1993. Connectionist models and cognitive theory. *Psychol. Sci.* 4(4):228–235.

Seidenberg, M. S., and J. L. Elman, 1999. Do infants learn grammar with algebra or statistics? *Science* 284:435–436.

Seidenberg, M. S. and J. D. Zevin, 2006. Connectionist models in developmental cognitive neuroscience: Insights about critical periods. In Y. Munakata and M. H. Johnson,

eds., *Processes of Change in Brain and Cogn. Dev.: Attention and Performance XXI*, 585–612. Oxford, UK: Oxford University Press.

Shinskey, J. L., and Y. Munakata, 2005. Familiarity breeds searching: Infants reverse their novelty preferences when reaching for hidden objects. *Psychol. Sci.* 16:596–600.

Shrager, J., and M. H. Johnson, 1996. Dynamic plasticity influences the emergence of function in a simple cortical array. *Neural Net.* 9:1119.

Shultz, T. R., 2003. *Computational Developmental Psychology*. Cambridge, MA: MIT Press.

Shultz, T., 2006. Constructive learning in the modeling of psychological development. In Y. Munakata and M. H. Johnson, eds., *Processes of Change in Brain and Cognitive Development: Attention and Performance XXI*, 62–86. Oxford, UK: Oxford University Press.

Shultz, T., D. Mareschal, and W. Schmidt, 1994. Modeling cognitive development on balance scale phenomena. *Machine Learn.* 16:57–86.

Shultz, T., W. Schmidt, D. Buckingham, and D. Mareschal, 1995. Modeling cognitive development with a generative connectionist algorithm. In T. J. Simon and G. S. Halford, eds., *Developing Cognitive Competence: New Approaches to Process Modeling*, 157–204. Hillsdale, NJ: Erlbaum.

Siegler, R., 1976. Three aspects of cognitive development. *Cogn. Psych.* 8:481–520.

Siegler, R., 1989. Mechanisms of cognitive development. *Annu. Rev. Psych.* 40:353–379.

Smith, L. B., 1999. Do infants possess innate knowledge structures? The con side. *Dev. Sci.* 2(2):133–144.

Smith, L. B., E. Thelen, B. Titzer, and D. McLin, 1999. Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychol. Rev.* 106:235–260.

Spelke, E., K. Breinlinger, J. Macomber, and K. Jacobson, 1992. Origins of knowledge. *Psychol. Rev.* 99:605–632.

Spelke, E., and E. Newport, 1997. Nativism, empiricism, and the development of knowledge. In R. M. Lerner, ed., *Theoretical Models of Human Development*. In W. Damon, series ed., Handbook of Child Psychology, 5th ed. New York: Wiley.

Stadthagen-Gonzalez, H., J. S. Bowers, and M. F. Damian, 2004. Age-of-acquisition effects in visual word recognition: Evidence from expert vocabularies. *Cognition* 93(1):B11–26

Tanaka, K., 1996. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19:109–139.

Thelen, E., and L. B. Smith, 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Thomas, M. S. C., 2003. Multiple causality in developmental disorders: Methodological implications from computational modeling. *Dev. Sci.* 6(5):537–556.

Thomas, M. S. C., 2004. How do simple connectionist networks achieve a shift from "featural" to "correlational" processing in categorization? *Infancy* 5(2):199–208.

Thomas M. S. C., and A. Karmiloff-Smith, 2002. Are developmental disorders like cases of adult brain damage? Implications from connectionist modeling. *Behav. Brain Sci.* 25:727–788.

Thomas M. S. C., and A. Karmiloff-Smith, 2003. Modeling language acquisition in atypical phenotypes. *Psychol. Rev.* 11(4):647–682.

Triesch, J., C. Teuscher, G. O. Deák, and E. Carlson, 2006. Gaze following: Why (not) learn it? *Dev. Sci.* 9(2):125–147.

Verguts T., and W. Fias, 2004. Representation of number in animals and humans: A neural model. *J. Cogn. Neurosci.* 16(9):1493.

Westerman, G., and D. Mareschal, 2004. From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy* 5(2):131–151.

Westerman, G., and E. R. Miranda, 2004. A new model of sensorimotor coupling in the development of speech. *Brain Lang.* 89(2):393–400.

Williams, J., and P. Dayan, 2005. Dopamine, learning, and impulsivity: A biological account of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychopharmacol.* 15(2): 160–79.

J. Child Adolesc. Psychopharma