Psychological Bulletin
2014, Vol. 140, No. 1, 224–255

0033-2909/14/$12.00 DOI: 10.1037/a0032150

Connectionist Modeling of Developmental Changes in Infancy: Approaches, Challenges, and Contributions

Yevdokiya Yermolayeva and David H. Rakison
Carnegie Mellon University

Connectionist models have been applied to many phenomena in infant development including perseveration, language learning, categorization, and causal perception. In this article, we discuss the benefits of connectionist networks for the advancement of theories of early development. In particular, connectionist models contribute novel testable predictions, instantiate the theorized mechanism of change, and create a unifying framework for understanding infant learning and development. We relate these benefits to the 2 primary approaches used in connectionist models of infant development. The first approach employs changes in neural processing as the basis for developmental changes, and the second employs changes in infants' experiences. The review sheds light on the unique hurdles faced by each approach as well as the challenges and solutions related to both, particularly with respect to the identification of critical model components, parameter specification, availability of empirical data, and model comparison. Finally, we discuss the future of modeling work as it relates to the study of development. We propose that connectionist networks stand to make a powerful contribution to the generation and revision of theories of early child development. Furthermore, insights from connectionist models of early development can improve the understanding of developmental changes throughout the life span.

Keywords: connectionist modeling, neural network modeling, parallel distributed processing, infant development

Since the groundbreaking work of Rumelhart and McClelland in the 1980s, there has been an increasing interest in and research on the application of connectionist models to early human development. Connectionist models are instantiations of theories about the mechanisms that underpin particular behaviors. Building computational models allows for the exploration of the interaction of numerous factors both internal and external to the organism that typically contribute to a behavior, which often can be too complex to specify through verbal theory alone (Shultz, 2003). These computational models provide researchers with a number of ways to explicitly test theoretical assumptions and develop novel and testable predictions.

However, in our view, for many developmental scientists the contribution of models to an integrated understanding of development is far from clear. Although a given model might provide output that is similar to the behavior of infants, it often remains to be seen whether the model's results and the behavioral results occur for the same reasons. For example, a network may simulate effectively infants' ability to discriminate between two objects that differ along multiple features, but it is possible that the features used by the network for discrimination are different from those used by infants. Computational models have also been criticized for using overly technical terms and notations that may be off-putting to nonexperts (Klahr, 2004); for not being explicit about the source of their starting states (Oakes, Newcombe, & Plumert, 2009); and for instantiating developmental changes through external manipulation—for example, manually setting different learning rates to simulate the behavior of older and younger infants—rather than allowing them to emerge naturally within the system (Younger, Hollich, & Furrer, 2004). Further, in cases where the modeler demonstrates comparable performance between a computational model and an infant without any further mechanistic analysis, the contribution is minimal beyond what is apparent from the behavioral data.

These issues highlight two key questions that are of interest to developmental scientists. First, how can a model successfully instantiate developmental change? Second, what does developmental research gain from a computational model of empirical data? In the current article, we discuss these questions with respect to *connectionist* models, also known as *neural network* models or *parallel distributed processing* (PDP) networks (Rumelhart, McClelland, & the PDP Research Group, 1986), that model children's competencies in the first 2 years of life.

Why Focus on PDP Models?

Several types of computational models have been applied to human development, the most prominent of which are connectionist or PDP models, dynamic systems models, rule-based models, and Bayesian models. Our focus in this article is on connectionist

This article was published Online First March 11, 2013.

Yevdokiya Yermolayeva, Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University; David H. Rakison, Department of Psychology, Carnegie Mellon University.

We thank David Plaut, Erik Thiessen, and Blair Armstrong for their helpful comments on previous versions of this paper.

Correspondence concerning this article should be addressed to Yevdokiya Yermolayeva, Department of Psychology, Baker Hall 342C, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: yyermolayeva@gmail.com

models of infant development, for two reasons. First, PDP models have an established history of success in simulating development. From a broad perspective, an informal analysis by Shultz (2003) showed that approximately 75% of the published developmental computational models have been connectionist. These models have accounted for some key developmental phenomena, such as trade-offs in brain development (McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & Munakata, 2000), nonlinear patterns of development (Plunkett & Marchman, 1993; Rogers & McClelland, 2004; Rumelhart & McClelland, 1986), perseveration (Munakata, 1998; Stedron, Sahni, & Munakata, 2005), and the emergence of semantic knowledge (Rogers & McClelland, 2004). Second, this article focuses on connectionist models in the interest of depth over breadth. Instead of providing a shallow survey of a variety of approaches, we chose to examine how a single type of model can be used to simulate developmental change in the first 2 years of life. For reviews that take on a comparative approach of different types of modeling frameworks, we refer the reader to articles by Mareschal (2010) and Munakata (2006); a book edited by Spencer, Thomas, and McClelland (2009) comparing connectionist and dynamic systems approaches; the 2010 special issue of *Trends in Cognitive Sciences* comparing connectionist and probabilistic approaches; and the 2003 special issue of *Developmental Science* on connectionist and dynamic systems approaches.

Why Focus on Infancy?

This article focuses on the first 2 years of life for three reasons. First, fewer empirical research tools are available to test very young children, and many research questions must be asked indirectly. Thus, process analyses and novel predictions generated by computational models may be particularly valuable for behavioral phenomena found in infancy. Second, connectionist models of infant development are especially numerous compared to those of other phases of development. An informal search in the PsycINFO database of the titles and abstracts of published articles, books, and book chapters that contain the terms *connectionism, connectionist network, connectionist model, PDP, parallel distributed processing,* or *neural network model* yielded 31 results when restricted to infancy, 21 when restricted to preschool age, 9 when restricted to school age, and 0 when restricted to adolescence. Finally, as mentioned above, our interest was in providing a deep review, which necessitated a limitation in scope. A focus on just the first 2 years of life allowed us to address the key issues in modeling neural- and experience-based developmental changes.

Despite the focus on the models of early development, in our view, the review presented here can benefit an audience beyond developmental researchers. The modeling techniques that are presented can be used to model change across the life span; for example, the approach to model neurogenesis can be reversed to simulate neural death in the elderly. Similarly, the challenges that have emerged in models of early development can be found in nondevelopmental models as well; for instance, insufficient research can make it difficult for a modeler to generate appropriate training examples for a model, regardless of the age that is simulated. Finally, the contributions that are outlined in this review are applicable to a wide range of connectionist networks including those in areas of clinical psychology (e.g., Siegle, Steinhauer, & Thase, 2004), cognitive neuroscience (e.g., Plaut & Behrmann, 2011), and social psychology (e.g., Monroe & Read, 2008).

Overview of the Article

Our main goals in the article are to answer two questions: Why should connectionist networks be used in the study of development? How has development been modeled in these networks? To answer the first question, we address the key theoretical contributions that connectionist models make: novel predictions, a concrete instantiation of the theorized mechanism of change, and a unified understanding of disparate experiments. Combined, these contributions impact significantly the generation and revision of developmental theories. To answer the second question, we review the two primary approaches to simulating development in connectionist models: focusing on the neural changes and focusing on the changes in experience as the primary force behind development. The neural-change-based approach emerges from a theory that changes in the brain underpin developmental changes. Within this approach, four specific neural changes have been modeled most commonly: those in perception, information integration, maintenance of information, and neural plasticity. The experience-based approach emerges from a theory that exposure to the surrounding environment underpins developmental change. This approach has typically been modeled by varying the amount or type of experience over developmental time. We conclude the article with a discussion of some shared challenges that these approaches face.

Our organization of the review into neural-based and experience-based approaches to modeling development should not be taken to suggest that there is a fundamental divide, such that the two approaches are incompatible and only one can be used in a given model or theory of development. The interplay between experience and neural development has been well established (Johnson, 2000, 2001; Quartz, 1999). Initially, cortical pathways are very weakly specialized, such that broad responses to various stimuli are observed in the brain. However, slight biases in processing strengthen over time as different pathways selectively process particular types of stimuli; thus, experience drives developmental changes in neural architecture and processing. In turn, these changes in the brain can drive infants' experience by constraining what infants attend to, process, and store. One example of this type of interplay between neural changes and experience can be seen in the early development of face processing. Scott, Shannon, and Nelson (2006) used event-related potential (ERP) markers of face processing to show that 9-month-olds exhibit more specialized processing of human faces than of monkey faces, which suggests that greater experience with the human faces alters neural processing. In turn, this change in processing alters how infants experience faces: Although 6-month-olds can discriminate faces of monkeys, 9-month-olds cannot (Pascalis et al., 2005). However, targeted experience with individuating monkey faces between 6 and 9 months not only allows 9-month-olds to preserve this ability (Pascalis et al., 2005) but also causes 9-month-olds to show ERP markers of specialization in processing monkey faces (Scott & Monesson, 2010). Thus, there are bidirectional influences between the infants' experience and neural change.

Although there is interplay between the neural- and experience-based changes in development, we have chosen to separate our discussion of these changes for two reasons. First, the implementation of neural-based changes in a connectionist network is different from the implementation of experience-based changes. That is not to say that networks can only implement one or the other; both can be implemented a single model (e.g., Li, Farkas, & MacWhinney, 2004). Rather, the ways in which one models neural development and changes in experience are different. Our goal in this article is to discuss how development can be modeled; because the "how" is different for models that implement neural and experiential changes, a clear review of each necessitates their separation. Second, although some behavioral researchers have moved toward an investigation of both neural and experiential influences on development, as in the study of face processing discussed above, others are still primarily interested in one or the other. For example, researchers interested in infants' perception and categorization of animals have focused on the role of prior experience with animals (e.g., Hurley, Kovack-Lesh, & Oakes, 2010; Kovack-Lesh, Horst, & Oakes, 2008). In contrast, those interested in the development of infants' memory capacity have attributed changes primarily to neural development (e.g., Káldy & Leslie, 2003, 2005). Therefore, a separate discussion of neural and experiential changes can make modeling more intuitive to many behavioral researchers, because their research may have a similar focus on just one type of change. However, before we proceed with the primary goals in this article, it is necessary to provide an overview of connectionist modeling that will familiarize the reader with the terminology and principles that are used throughout the article.

Overview of Connectionist Modeling

Basic Principles

Connectionist models were designed with a neural inspiration originally to solve cognitive problems. There are several basic features that are characteristic of these models (Rumelhart, Hinton, & McClelland, 1986). Connectionist systems operate in an environment instantiated by the researcher: This is specified by a particular distribution of training patterns that the network is expected to learn. These systems contain a set of processing units that can represent particular input stimuli (*localist* coding) or features of those input stimuli (*distributed* coding). These units all have some state of activation that indicates the current information that the system represents. For example, a unit can represent a whole image of a dog or a single feature (e.g., a tail), and the unit's activation could indicate that the image or feature is visible. Every unit's net input is transformed into an output via a prespecified function, which is typically nonlinear. This output is passed on to other units to which the given unit has outgoing connections. The pattern of connectivity, or weight structure on the connections between units, determines how outputs are propagated. Negative and positive weights indicate inhibitory and excitatory connections, respectively. There is an integration rule that calculates the net input the unit receives from other units based on the outputs of those units and the weights on the connections; this rule typically involves calculating the weighted sums of the excitatory and the inhibitory inputs. There is an activation rule that calculates the new activation of the unit based on the net input from other units in the system. There is a rule that dictates how connections should be modified as a result of experience. For example, the network may employ Hebbian learning, in which the connection between two units is strengthened when both are active. Alternatively, the network may employ a supervised learning rule, in which weights are altered to minimize error on future trials after feedback is received on the current trial.

A crucial feature of connectionist networks that distinguishes them from some other computational models—rule-based systems, for example—is the fact that individual learning episodes are not stored independently in the system (Rumelhart, McClelland, & the PDP Research Group, 1986). Rather, all knowledge is superimposed over the same connections within the system, and these are adjusted as the network learns the interdependencies in the input. A pattern of connections that encodes knowledge typically takes a long time to develop and can often be seen as representing long-term memory (Shultz, 2003). In contrast, activations of specific units change more rapidly and can often be thought of as representing working memory (Shultz, 2003).

We take the above properties to be our operational definition of a connectionist network that was used to select models for review. It should be noted that these properties can be satisfied in a variety of ways. As mentioned above, learning rules for adjusting weights can vary, so that some networks employ unsupervised Hebbian learning (e.g., self-organizing maps), and others employ error-correcting learning. Similarly, although most connectionist structures maintain a nonchanging architecture over the course of a single simulation, some variants have a dynamic component in which units are added throughout the simulation (e.g., cascade-correlation networks). We elaborate on these variations in the next section. Crucially, despite the variability, all of the networks satisfy the above criteria.

When one examines connectionist networks in the context of development, it is important to consider their ability to instantiate three time scales: long-term developmental time of months and years, smaller scale time of individual learning episodes, and short-term activations during the performance of a particular behavior. Connectionist models can encompass all three (McClelland & Vallabha, 2009; O'Reilly & Munakata, 2000; Thomas, McClelland, Richardson, Schapiro, & Baughman, 2009). The developmental time scale can be represented by weights that gradually change with experience (McClelland & Vallabha, 2009; Shultz, 2003) or by structural or parameter changes to the model that simulate neural development (Sirois & Shultz, 2003). Individual learning episodes can also be simulated with weight adjustment: This provides continuity in the model by which discrete learning instances add up to long-term changes. Finally, moment-to-moment task performance is represented by the activations of specific units (McClelland & Vallabha, 2009; Shultz, 2003). For example, a child's expanding vocabulary may be represented by adding units to increase neural capacity, an individual case of learning a word can be represented by adjusting weights between units, and the child's production of a word may be represented by the activation of an output unit that represents that word. These time scales are intertwined in networks: The current activation is a function of the weights that have been developed over time, and this activation leads to further weight changes (McClelland & Vallabha, 2009). In terms of behavior, this can be understood as infants' prior experiences affecting their current actions and their current actions being incorporated into their experiences.

In general, the handling of time in a given connectionist network reflects the theoretical commitments and the interests of the modeler regarding the factors that contribute to a given behavior. For example, Franz and Triesch (2010) theorized that experience with movement and occlusion can drive infants' performance in object unity tasks. Accordingly, they instantiated an extensive experience phase during which the network adjusted its weights as it learned a set of training patterns designed to represent several months of real-world experience. After this phase, it was trained on patterns that represented the specific experimental occlusion events, and it further adjusted its weights. In the test phase, the network's unit activations to the test stimuli were used as proxy for infants' looking times. In contrast, French, Mareschal, Mermillod, and Quinn (2004) theorized that categorization of cats and dogs in the Quinn, Eimas, and Rosenkrantz (1993) study was driven solely by the online learning in the experimental context. Therefore, the networks' weight adjustment during training reflected learning only about the patterns that represented experimental stimuli. As in Franz and Triesch's simulation, unit activations to test stimuli were used as a proxy for looking time. A comparison of these models demonstrates that developmental time can be handled very differently by connectionist networks, depending on the theory adopted by the modeler: Weight adjustment can reflect long-term learning about daily experience or short-term learning about lab stimuli. In both cases, the learning affects immediate behavioral responses, which are represented by unit activations.

Interest in particular phenomena can also lead to an emphasis of dynamics at a particular time scale. For some, the moment-by-moment changes in behavior within a trial are important, whereas others consider only the overall behavior for a given trial. As an example of the former, consider Schlesinger and Casey's (2003) model of the classic Baillargeon (1986) object permanence experiments. Schlesinger and Casey had specific predictions about the relationship between object movement and infants' attention, so they were interested in the amount of looking to the left, right, and center regions of the display instead of the overall looking to consistent and violation trials, which was the outcome measure used in the behavioral experiment. Accordingly, the model that they implemented simulated changes in eye gaze direction within a single test trial. This can be contrasted with a model such as that of French et al. (2004), which used a single set of activation values to represent the overall looking for a test trial. It is conceivable that had French et al. been interested in the dynamics of looking within a trial, such as scanning of animal features, they could have implemented the model on a finer time scale.

The implementation of the three time scales—development, learning, and activation—utilizes a single framework. Thus, although many options exist for the specific implementation of learning (e.g., weight adjustment can represent long-term learning over months or years or short-term learning during a single study) and of activation (e.g., activation could represent average trial behavior or finer behavioral changes within the trial), the same set of basic principles applies across all of them. Thus, different implementations of time should not be perceived as divisive and arbitrary. Rather, connectionist models unify our understanding of development by demonstrating that the same mechanisms of distributed representation can give rise to moment-to-moment activations and to long-term developmental changes.

Common Model Designs

Connectionist modeling uses an unconstrained framework in terms of model structure. That is, there are no specific requirements for the number of units, groups of units, the learning rule, or the interconnectivity of the network. However, there are particular network structures that have been used commonly to model cognitive development. These are reviewed in this section and summarized in Table 1. This is not intended to be an exhaustive review of all possible modeling approaches, and a more detailed overview can be found in Shultz (2003). It should be noted that, on the surface, the unconstrained nature of connectionist architecture and the numerous types of networks outlined below may imply that there are no principled reasons for choosing a particular structure to model a given phenomenon. However, some network structures are more suitable to model certain tasks over others, due to the features of the task or the theories about information processing during the task; these alignments between structure and modeled phenomenon are highlighted in the overview that follows.

A commonly used network architecture is the standard *multi-layer backpropagation network* (Rumelhart, Hinton, & Williams, 1986). Other network types are often special cases of this network. An example of this type of structure can be seen in Figure 1A. A multilayer backpropagation network consists of input, hidden, and output units. Hidden units are interim units between the information received from the environment (pattern presented on the input units) and the response that is produced (activity of the output units). There may be a single layer of hidden units (a *three-layer* backpropagation network) or multiple layers. As the network learns the training patterns, the connections between units are adjusted based on how well the activation of the output units approaches the desired activation. That is, once the network has activated the output units, it gets feedback that identifies the correct, or target, response. On the basis of this concrete target, the network adjusts the connection weights between all of the units in a way that would minimize error on future trials. For example, if a network is learning object labels, it may activate a unit that represents a word after receiving input about an object. The network would receive feedback regarding which label should have been activated, and the connections would be adjusted in a manner that would move the network in the direction of producing that label. This can be representative of learning over the course of a lifetime or over the course of an experiment. Backpropagation networks are a general type that can be adapted to model a number of tasks by adjusting what the input and output units code. For example, they have been applied to tasks in which the infant has to reach to a particular location, in which case the output units may code object locations (e.g., Munakata, 1998), or to tasks that assess categorization, in which case output units may code global and basic categories (e.g., Quinn & Johnson, 1997). Thus, the backpropagation network is often chosen when an explicit response, such as a reach location, is expected from the system.

A special case of the multilayer backpropagation network that has been commonly used to model infant habituation studies is the *encoder network,* also known as an *autoencoder* (Ackley, Hinton, & Sejnowski, 1985; Rumelhart, Hinton, & Williams, 1986), which is shown in Figure 1B. In the encoder network, the input and output units code the same set of features (specified in the figure by similar shading of input and output units), and there are typi-

Table 1
Overview of Network Types Typically Used in Connectionist Models of Infant Development

Network type	Common features	Typical uses	Example
Multilayer backpropagation network	Units organized into input, hidden, and output layers. Output compared to the target, and weights are adjusted according to the error.	Adaptable to any task by adjusting what the output units code	Quinn & Johnson (1997) Gasser & Smith (1998)
Encoder network	Special case of the multilayer network in which input and output units are identical. Network trained to reproduce the input on the output units via a compressed hidden unit representation.	Habituation studies	Mareschal et al. (2000) Westermann & Mareschal (2004)
Simple recurrent network (SRN)	Input and output units are identical; network predicts the next input based on the current one. Network has a form of memory via the context units that preserve the hidden unit activity from the previous time step.	Habituation studies with a temporal component	Elman (1993) Munakata et al. (1997)
Auto-associator network	Network has a single set of interconnected units that act as both the input and output units. Network settles on an activation pattern over multiple processing cycles.	Tasks in which responses occur over time	Sirois et al. (2000)
Self-organizing map	Grid of units in which the information is organized topographically such that units next to each other respond to similar inputs.	Tasks where the similarity among the items in the input must be discovered; modeling the topographic organization of certain cortical areas	Cohen et al. (2002) Mayor & Plunkett (2010)
Cascade-correlation network	Network has input, hidden, and output layers. Hidden units are added throughout training until the target outputs are approached. Weights trained in successive phases.	Habituation tasks using encoder and SRN variants; instantiating synaptogenesis through hidden unit addition	Shultz & Bale (2001) Shultz & Cohen (2004)

cally many fewer hidden units than input and output units. The network is provided with input patterns and is trained to reproduce the same patterns on the output units. In this process, the network must form a compressed representation over the hidden units because those units are few in number. A primary reason for the use of an encoder network is to model habituation studies, because the network aligns itself with theoretical accounts of habituation (Mareschal, French, & Quinn, 2000; Mareschal, Quinn, & French, 2002). One popular way to conceive of the habituation procedure is that infants perceive a stimulus, construct an internal representation of it, and compare that representation with what is available in the environment (Cohen, 1973). If a mismatch occurs between what is expected and what is observed (e.g., when a new stimulus is presented to the infant), attention is increased, but if there is no mismatch (e.g., when the stimulus is familiar), attention remains low. Similarly, the model is presented with an input pattern that it then re-represents on the output units. The output units' activations are compared to the target activations (identical to the input), which generates error. Error is high in the model when there is a mismatch, comparable to longer looking for infants. Encoder networks present a single input pattern per trial. As a result, they are particularly useful for modeling habituation tasks in which what the child experiences within a trial does not change (e.g., the child sees an image that does not change throughout the trial). However, they would not be appropriate for situations in which the moment-to-moment dynamics within a given trial are of interest.

Habituation has also been modeled with *simple recurrent networks* (SRNs; Elman, 1990), an example of which is shown in Figure 1C. These networks are more useful than encoders for situations in which the child experiences changing information within a given trial (e.g., the child sees moving objects). These networks have a simple form of memory enabled by *context units* that store the hidden unit activation from one time step and feed it back into the hidden units on the next time step. Thus, the internal representation of the input at a given time point is influenced by the representations of the prior inputs. The task of the model is to predict the next input given the current activity of the input and context units and, based on the feedback about the prediction, to adjust its weights to make more accurate predictions. For example, on the basis of an object's current and previous locations, the network may predict the object's location on the next time step (e.g., Munakata, McClelland, Johnson, & Siegler, 1997; Rakison & Lupyan, 2008). This prediction task is appropriate for modeling habituation, because there is a comparison between the internally generated prediction and the external evidence—just as with encoder networks—that generates an error signal that is comparable to infants' increased attention. However, because SRNs predict the upcoming input within a trial, they are more useful than encoders for modeling within-trial dynamics.

Developmental phenomena have also been modeled with *auto-associator networks* (Kohonen, 1977; Rumelhart, McClelland, & the PDP Research Group, 1986). An example of this type of network is shown in Figure 1D. These networks consist of a single set of fully interconnected processing units that serve as both the input and the output; thus, the input ultimately is associated with itself. Processing occurs over multiple cycles during which the network settles into a stable activation state; this adds a temporal component to the task. The settling of these networks has been

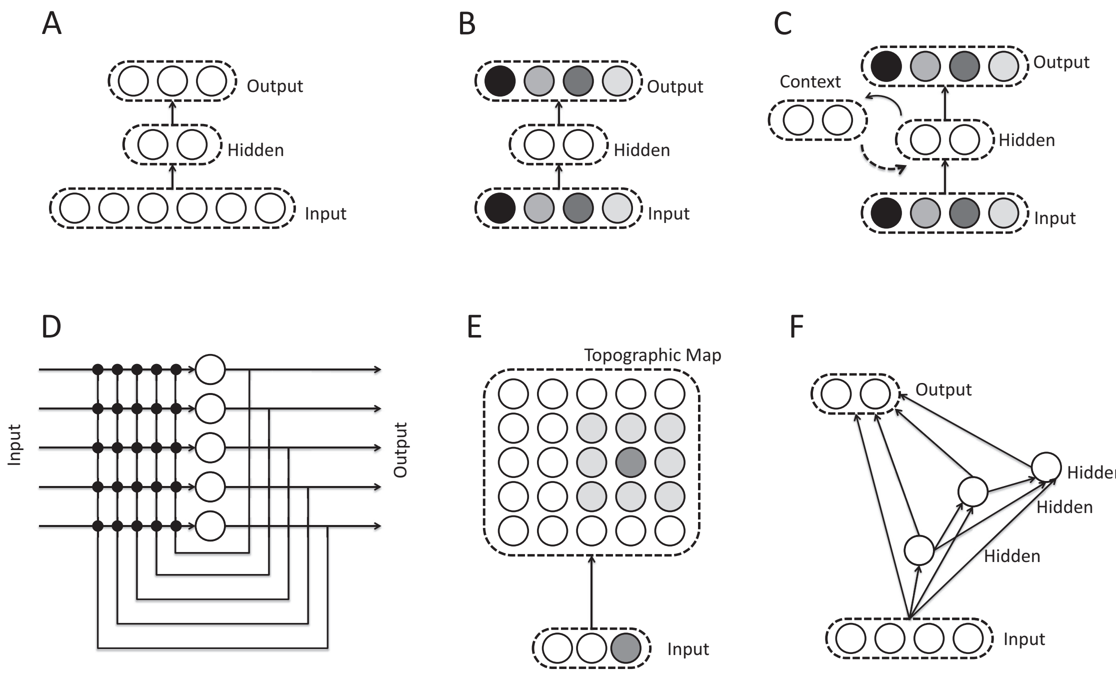

Figure 1. Common connectionist network designs. Units enclosed by a dashed line represent banks of units. Arrows between banks of units represent full connectivity; that is, each unit in one bank is connected to every unit in the other bank. (A) Multilayer backpropagation network with 6 input units, 2 hidden units, 3 output units. (B) Encoder network with 4 input and output units and 2 hidden units. Shading represents the fact that input and output units code the same information. (C) Simple recurrent network with 4 input and output units, 2 hidden units, and 2 context units. Shading represents the fact that input and output units code the same information, and the dashed arrow to the hidden units represents the fact that the context units feed in an exact copy of the previous activation and that the connection is not trained. (D) Auto-associator network with 5 units. The system can receive external input and produce external output. Units feed activation to all other units and themselves. (E) Self-organizing map with 3 input units and a map of 25 units. Shading represents the topographic organization of the map such that one unit responds strongly to the input and neighboring units respond weakly. (F) Cascade-correlation network with 4 input units, 3 layers of hidden units, and 2 output units.

related to the processing dynamics in a typical habituation experiment at the level of looking time (Sirois, Buckingham, & Shultz, 2000). The network requires some number of cycles to settle on a stable activation state when the habituation stimuli are presented. This number is representative of the amount of time that an infant requires to habituate, or the amount of time it takes the infant to form a stable internal representation of the stimuli. In the test phase, the network takes more cycles to settle after being presented with a novel item than a familiar item, because the former is inconsistent with the statistics of the habituation items. Similarly, infants look longer at novel than familiar items because the former are inconsistent with their internal representation. Researchers interested in the way in which habituation unfolds over time may opt to use auto-associators because of this coupling between settling cycles in the model and processing dynamics in infants. However, auto-associators do not generate an internal representation of the information, so they may not be as useful when researchers are interested in the way that similarity relations between stimuli drive internal representations.

Another common type of architecture used to model development is the *self-organizing map* (Kohonen, 1982), which is shown in Figure 1E. These networks form a topographic organization of the input across a two-dimensional map of units. Units in the map compete to respond to inputs: Winning units, those with the strongest response to the input pattern, become specialized in responding to that input pattern through weight adjustment. Neighboring units get their weights adjusted in the same direction, such that they become specialized in similar input patterns. As a result, nearby units come to respond to similar input items, which creates a topographic map of the input stimuli. This is designated in the figure by graded shading: The unit with the darkest shading responds most strongly, and neighboring units respond more weakly. On this map, similar items are represented as closer to each other than are dissimilar items. For example, in a model of speech perception, units that are near each other may respond to similar phonemes. The learning that occurs in these types of networks is not based on error correction, and it is considered to be representative of long-term learning of statistical regularities in the real world (O'Reilly & Munakata, 2000). Self-organizing maps can be effective for representing the topographic organization of the visual cortex, auditory cortex, or somatosensory cortex (Shultz, 2003). Thus, modelers typically choose this type of network when they are interested in two things: how the similarity structure of the input drives the formation of internal representations, and how these representations can be arranged in a neurally plausible manner. However, the complexity of the similarity relations that can be represented is limited because the information is represented on a two-dimensional surface. Thus, this approach may not be useful for problems that involve high-order similarity relations.

The final type of network that has been widely used is the *cascade-correlation network* (Fahlman & Lebiere, 1990), shown in Figure 1F. This network type is similar to a standard backpropagation network in that it has input, hidden, and output layers. However, it has one crucial difference in that its architecture is dynamic: New hidden units are added to the structure to reduce error throughout training. This contrasts with the other networks that have been discussed, because their architecture remains constant throughout the simulation. A cascade-correlation network alternates between two weight training phases: training input to hidden unit weights and training hidden to output unit weights. Training is continued until the output units' activations are within some threshold of the target activations. Encoder networks can be built in the cascade-correlation fashion such that the input and output units are identical to those in a regular encoder network, but there is additional hidden unit recruitment throughout training. Thus, cascade-correlation networks can also be used to model habituation studies. Some researchers (e.g., Shultz & Bale, 2001) have argued that cascade-correlation networks have an advantage over standard backpropagation networks because the former require fewer trials to train; therefore, the cascade-correlation networks are more comparable than standard encoder networks to the training typically received in an infant behavioral experiment.

This section provides only a brief overview of the various network structures available to modelers. These structures are infinitely adaptable from their general form. For example, modelers may link several self-organizing maps together in a hierarchical fashion to simulate visual processing at different levels of specificity (e.g., Cohen, Chaput, & Cashon, 2002). The flexibility of connectionism allows such modifications to be made, and this can enable modelers to instantiate theories about mechanisms of developmental change more effectively.

This section provided an outline of the basic terminology and principles of connectionist modeling. In the following sections we address our primary goals in the article: to discuss why these networks should be used to study development and to demonstrate the way in which they have been applied in that field.

The Benefits of Connectionist Modeling

As we have discussed, behavioral researchers who study child development are often hesitant about the benefits of connectionist networks. One criticism of the value of modeling is that models do not advance our understanding of a behavior because they simply replicate it (Sloman, 2008). However, we propose that connectionist models are indispensable to developmental research because of the contributions they make to theory building. These models make three primary contributions to research on development: They create novel testable predictions, they instantiate the hypothesized mechanism of change, and they create a unified framework for broader understanding of early developmental change (Mareschal, 2010; McClelland, 2009; McClelland & Rumelhart, 1986; Shultz, 2003). The ultimate goal in behavioral experimentation is not to map out infants' failures and successes in individual tasks. Rather, those individual behavioral experiments are building blocks for the formulation of broader theories about development. In our view, these three contributions of connectionist models can be critical for the construction and evaluation of such theories.

Novel Developmental Predictions

Connectionist models can make new predictions that are either task related or developmental. Task-related predictions concern the expected results of a new task or a new outcome measure applied to a current task. Developmental predictions concern the expected behavior at an age that has not yet been tested. Because our focus is on modeling development, below we discuss two examples of developmental predictions that have been confirmed behaviorally. A more representative list of confirmed developmental and task-related predictions, found in Table 2, demonstrates that connectionist models have a history of making predictions that have been supported behaviorally in a variety of domains.

An example of an empirically supported developmental prediction was generated by Colunga and Smith's (2005) simulation of a noun generalization task. The model predicted that children who are in the initial stage of word learning would extend labels of nonsolid test items but not of solid test items. This prediction was based on the ability of networks with no vocabulary training to extend labels only for nonsolid objects, a behavior that was due to the similarity among training patterns. Representations of nonsolid items of the same material clustered closer together than did representations of nonsolid items of the same shape, whereas representations of solid items of the same material were as closely clustered together as representations of solid items of the same shape. Thus, without any vocabulary training, the network was able to extend labels on the basis of initial pattern similarity. The implication of this prediction was that visual similarity of nonsolid items of the same material can support generalization, but vocabulary exposure is needed for shape-based generalization to emerge. Colunga and Smith (2005) administered the noun generalization task to a sample of younger children and found empirical support for this prediction: Children who had not yet learned many words could generalize the names of nonsolid items by material.

Another clear example of a network that produced a novel developmental prediction is Munakata's (1998) simulation of perseveration in the A-not-B task. Development was modeled by strengthening recurrent connections in the network, and the development of perseveration was plotted by measuring the proportion of correct reaches at different connection strengths. This plot revealed that perseveration followed a *U*-shaped path, such that networks with the weakest and the strongest recurrent weights perseverated less than networks with midlevel recurrent weight strength. This result was due to the role of recurrent weights in building a response bias to the initial hiding location (A). When the recurrent weights were very weak, they did not allow the network to develop an A bias even after multiple presentations of A: Thus, the youngest networks with the weakest recurrent connections were not highly likely to respond to A. As the recurrent weights increased, the activation of the A location persisted over a delay, which led the network to develop a bias to perseverate at that location. Finally, recurrent weights became strong enough to retain the activation of the new hiding location (B), which decreased

Table 2
Representative Sample of Articles That Have Made Concrete Predictions Using a Connectionist Model

Study	Domain	Prediction type	Specific prediction
Colunga & Smith (2005)	Label generalization	Developmental	Children with small vocabularies should be able to extend labels of nonsolids but not of solids
French et al. (2004)	Categorization	Task related	Asymmetry in infants' cat and dog categorization can be reversed by using a cat category with a wide range of features and a dog category with a narrow range
Mareschal et al. (2002)	Categorization	Task related	Asymmetric retroactive interference when learning cat and dog categories sequentially such that learning dogs first interferes with subsequent cat learning, but learning cats first does not interfere with subsequent dog learning
Munakata (1998)	Perseveration	Developmental	The development of perseveration in the A-not-B task follows a *U*-shaped trajectory such that initially babies do not perseverate, then they go through a period of perseveration, and finally they once again do not perseverate
Quinn & Johnson (2000)	Categorization	Developmental	Infants' ability to form global-level categories emerges prior to their ability to learn basic-level categories
Rakison & Lupyan (2008)	Learning correlations	Task related	Infants will learn correlations between a label and the moving parts of an object in a context when both are correlated with global motion. When parts, bodies, and labels are correlated but global motion is not, infants will learn the correlation between the parts and the body
Schlesinger & Casey (2003)	Object permanence	Task related	In Baillargeon's (1986) classic car study on object permanence and solidity, infants should spend the most amount of time looking at the center of the display, followed by the right, and then the left

Note. All predictions reported here have been confirmed behaviorally by the authors, except for Munakata's (1998) prediction, which was confirmed by Clearfield et al. (2006).

perseveration. The network generated a new prediction regarding the presence of a *U*-shaped developmental trajectory for perseveration (see also Thelen, Schoner, Scheier, & Smith, 2001, for a similar *U*-shaped prediction for perseveration based on dynamic field theory). The prediction was supported by Clearfield, Diedrich, Smith, and Thelen (2006), who showed that 5-month-old infants did not perseverate in the A-not-B task, whereas 7- and 8-month-olds did perseverate.

The developmental predictions made by an existing network can serve three purposes. First, those that are confirmed behaviorally provide additional—although not conclusive—evidence that the network is a faithful representation of the infant's behavior. A simulation typically is built to fit an existing set of data, and the simulation has far greater credibility if it can predict an independent set of data. This, in turn, strengthens the verbal theory that inspired the simulation. Crucially, in contrast to predictions purely based on verbal theory, a prediction that emerges from a simulation is based on a mechanism that is instantiated and that has been demonstrated to successfully produce the behavior in question. Second, predictions could be beneficial to developmental scientists, because they are an efficient way to inspire new behavioral experiments. For example, it is costly in terms of time and money to test a large number of age groups behaviorally to plot a continuous developmental trajectory. As an alternative, this task can be simulated in a model through repeated testing as some developmental parameter is varied. Based on the network's prediction of the trajectory, only the key age groups have to be tested. For example, based on Munakata's (1998) predictions, it was necessary to test a younger age than the age at which infants have been shown to perseverate to confirm the *U*-shaped trajectory.

A final purpose of new predictions is their contribution to the evaluation of different theoretical accounts. Shultz and Bale (2001), for example, found that their simulation of the Marcus, Vijayan, Bandi Rao, and Vishton's (1999) rule-learning experiment showed a gradation of error to different types of test items: Error increased for test items as their phonological similarity to the training items decreased, even though all test items followed the learned rule. This performance suggested that infants should also display graded patterns of looking if they were tested with items that varied in their similarity to the training stimuli. A rule-based account would have predicted equal looking to all test items regardless of similarity, because looking would be based only on whether the item follows the trained rule and any phonological variability would be ignored. These contrasting predictions could be tested empirically to determine which theory is supported by behavioral data; however, this remains to be carried out. Thus, in addition to saving the resources of behavioral researchers, predictions based on connectionist modeling work can advance existing theories of behavioral phenomena.

Concrete Instantiation of the Underlying Mechanism

A second contribution of models to developmental research is the instantiation of the theorized mechanism. The observed behavioral results are often explained through a verbal theory that describes the underlying mechanism. However, behavioral data alone cannot confirm that the theorized mechanism is, in fact, correct. Although new predictions can be generated and tested based on the theory, it is only by building the mechanism that researchers can show that it produces the expected behavior. The construction of a connectionist simulation can provide this explicit test of the theory, because it requires scientists to declare and evaluate their assumptions (Mareschal, 2010). If the hypothesized mechanism as instantiated in the network does not yield results that are similar to the behavioral data, it is likely that the mechanism is incorrect. The processes of developing a verbal theory and instantiating that theory in a connectionist network could be thought of as reductionism and reconstructionism, respectively (O'Reilly & Munakata, 2000). The former takes the behavior and identifies the components that underwrite it. The latter takes those components and puts them back together in an attempt to show that they do, in fact, result in the behavior in question. Thus, a network instantiation can be particularly useful to behavioral researchers for verifying a verbal theory about the underlying mechanism.

An example of the use of a connectionist model to verify a theorized mechanism comes from the work of Mareschal, Plunkett, and Harris (1999) on infants' expectations about occluded objects. They theorized that the integration of information from two neural processing streams, one dedicated to object identity and one to location, explains the finding that looking-time measures show earlier development of object permanence than do reaching measures because the former depend only on object location information, whereas the latter depend on the integration of identity and location. It is possible to compare the time course of the neural development of the two streams to the time course of the behavioral development of object permanence. If neural integration occurs close in time to behavioral changes, this may suggest that the two are related. However, a concrete construction of the mechanism in a simulation is necessary to establish the causal relationship between integration and behavior, which is exactly what was done by Mareschal et al. They constructed a connectionist network that contained two processing streams, one that encoded object identity and one that encoded object location. The network could provide responses based on location (akin to looking-time measures) or based on an integration of location and identity by feeding information about location and identity into a shared set of hidden units (akin to reaching measures). The model showed that expectations about occluded objects based on integrated information took longer to develop than responses based on location only due to the additional time required to adjust weights that represent a combination of object identity and location. This result replicated the behaviorally observed lag between looking time and reaching measures and provided critical support for the proposed theoretical explanation.

Theoretical explanations are not always supported by connectionist simulations. In fact, the benefit of instantiating the mechanism in a model is that it can point out areas in which the verbal theory fails to explain the behavior. One example of a model that contributed to theory revision is Schlesinger and Young's (2003) connectionist model of Baillargeon's (1986) study of object permanence and solidity (for a further investigation of similar violation-of-expectation experiments from the dynamic field theory perspective, see Schöner & Thelen, 2006). In the behavioral experiment, infants were habituated to a car that moved along a partially occluded track. In the test trials, a block was placed either on or next to the track and subsequently occluded. Infants looked longer when the car continued to roll along the same trajectory when the block was on the track than when it was behind or in front of the track. Baillargeon (1986) concluded that infants reasoned that one object cannot pass through another and that infants had a notion of object permanence because they could remember that the occluded block was still there. Schlesinger and Young (2003) theorized that performance in Baillargeon's task could be based solely on predictive learning during the experiment (i.e., learning to anticipate object movements and locations) and that prior knowledge about solidity and object permanence was unnecessary. In accordance with this theory, they constructed an SRN, a network that learns through prediction, and trained it only on patterns that represented behavioral stimuli without endowing it with prior knowledge. In Simulation 1, Schlesinger and Young's model showed similar behavior to infants: It exhibited higher error to a test trial in which the block was on the track than one in which the block was in front of the track. However, a follow-up simulation contradicted the behavioral findings: The network exhibited higher error when the block was behind the track than when it was on the track. In contrast, in Baillargeon's study, infants' pattern of looking was the same when the block was in front and behind the track. Schlesinger and Young concluded that a predictive mechanism was insufficient to explain behavioral data and that the theory of infants' behavior needed revision to take into account prior knowledge.

A final example of a connectionist model that contributed to theory building through the instantiation of the hypothesized mechanism is the work of Colunga and Smith (2005) on novel noun generalization for solids and nonsolids based on shape and material, respectively. The theoretical account for the development of generalization emphasized the role of the statistics of the language to which children are exposed: the correlations between solidity, shape, and material for nouns in the language were theorized to drive performance in the novel noun generalization task. Differential performance in the task has been observed for English- and Japanese-speaking children: The former extend labels for both complex and simple solids based on shape, whereas the latter extend labels for complex solids based on shape and labels for simple solids based on material (Imai & Gentner, 1997). As a test of the verbal theory that children's native language influenced their behavior, a connectionist network was constructed, trained on the noun regularities of either English or Japanese, and compared to the behavioral data of English- and Japanese-speaking children. In the test phase, the network exhibited the same behavior for both languages: It generalized labels for complex solids based on shape and for simple solids based on material. This matched Japanese but not English behavioral data. However, when the noun regularities in the English training set were supplemented by count-mass syntax information (information that is not present in Japanese), the network accurately simulated novel noun generalization for both languages. Thus, the initial theory had to be revised to include not only the statistics on how nouns correlate with solidity, shape, and material but syntactic data as well. This addition may not have been apparent from the behavioral data and a corpus analysis of each language alone. Although highly similar, Japanese and English have slight variations in their noun distributions. Without a concrete instantiation, it may have been hypothesized that these variations could be sufficient to generate the behavioral differences between English- and Japanese-speaking children.

Taken together, these three examples demonstrate the importance of a model instantiation of the proposed mechanism for theory development. In many cases, it is difficult to confirm a verbal theory solely by behavioral evidence. Verbal theories can make predictions for behavioral performance, but confirmation of those predictions does not indicate that the hypothesized causal relationship exists. It may be the case that the theory has missed a critical neural process that operates in concert with those included in the theory. However, when all proposed neural structures and processes are explicitly instantiated in a model, one can see if they produce the behavior in question. This provides powerful support for a theory or highlights a need for revision.

Creation of a Unifying Framework

A final contribution that models can make to developmental psychology is to create a unified framework for understanding development. This can occur on two levels. At the lower level, a single connectionist architecture can be applied to model several behavioral results, underlining their commonalities. At the higher level, different models that all obey the principles of connectionism can highlight broader aspects of development that are shared across domains.

Connectionist models have a history of successfully integrating individual experiments within the same domain. This type of integrated account of performance contributes to theory building by joining disparate pieces of empirical data into a unified understanding of a behavioral phenomenon (Sun, 2008). Several labs may study a particular behavior and generate a number of findings in different experimental contexts (e.g., infants' categorization studied by Mandler, Bauer, & McDonough, 1991; Quinn & Eimas, 1996; and Rakison & Butterworth, 1998; among others). Typically, it may be difficult to compare across these studies to formulate broad conclusions about the behavior. However, if the results of multiple studies can be simulated within the same network, this suggests that a common underlying mechanism may underpin them. Rogers and McClelland (2004) accomplished this in a model of the development of semantic cognition. They used a single basic architecture in their simulations in which the network was provided with an item and a context and was trained to output the appropriate features of the item. With only minor variations of the architecture or training that were designed to explore particular aspects of representation and processing, Rogers and McClelland were able to simulate a broad range of phenomena such as categorization in infancy, learning labels at different levels of inclusiveness, and inductive generalization. They demonstrated that one of the primary driving forces for all of these phenomena was *coherent covariation* of features, which is the grouping of training items based on the common clusters of features among these items.

Similarly, Franz and Triesch's (2010) network was a unified simulation of 12 studies from different laboratories that explored two domains of perception: object unity and continuous trajectory. The same pretraining phase representative of prior experience was needed to simulate the tasks on unity perception and those on trajectory perception; this consisted of static and moving objects that occasionally became occluded. Networks that did not receive an adequate amount of pretraining failed both types of tasks. The fact that the same pretraining was necessary to model 12 distinct experiments from two domains suggests a potential common underlying basis for the two areas of research: Infants' exposure to moving and occluded objects in the world is a driving force behind their perception. Additional examples of unified simulations are the works of Munakata (1998) and Stedron et al. (2005) on perseveration; Li et al. (2004) and Li, Zhao, and MacWhinney (2007) on language acquisition; and Van Overwalle (2010) on teleological reasoning.

In addition to shedding light on common mechanisms across different behavioral tasks, the creation of a unified model can help to address a criticism raised against connectionist networks regarding the arbitrary nature of their starting states, which do not intuitively match the characteristics of children (Oakes et al., 2009). That is, the structure of the network prior to the simulation (e.g., the starting weights, unit connectivity patterns) does not clearly parallel the state of the child at the start of the experimental observations (e.g., memory capacity, prior developmental history). Furthermore, starting states vary across simulations, which highlights their apparent arbitrary nature. This problem can be remedied in part by the use of a single framework to model numerous behavioral results. If the same starting state is adopted to model a range of data (e.g., the same unit connectivity), this suggests that the starting state is not arbitrary and may, in fact, provide an accurate depiction of the child at the start of an experiment.

Beyond the more narrow unification at the level of a single architecture modeling several experiments, there is a wider level of unification that has been created by the use of connectionist approaches across domains (e.g., categorization, theory of mind, language). At this level, connectionist models have extended our theoretical understanding of learning and development beyond particular tasks and domains by highlighting the emergence of complex behaviors from simple components (McClelland & Vallabha, 2009). It has been argued that infants' environments contain an overwhelming amount of information, which requires biases or constraints to guide learning and prevent the learning of arbitrary associations (Keil, 1981; Murphy & Medin, 1985). Across domains, connectionist models have been used to show that such biases may not be necessary. Connectionist networks have highlighted the importance of experience in driving development (Elman, 2005) and have shown that a system based purely on associative learning principles without any prespecified biases can learn the relevant information, exhibit complex behaviors, and show nonlinear patterns of development (Munakata & McClelland, 2003). This has been referred to as emergence: Observed behaviors emerge as a result of the interaction of various mechanisms; the mechanisms are formulated in terms of equations, but their joint operation gives rise to something more complex that is on a different level of description (McClelland & Vallabha, 2009). That is, the specific equations that determine the network processing (e.g., how the inputs to a unit are integrated to determine its activation) reflect the mechanistic dynamics. Based on these dynamics, behaviors such as labeling objects or reaching to certain locations can emerge. Furthermore, these dynamics are not limited to typical development: Small differences in starting states can lead to large individual differences and developmental impairments, without the need for specialized module damage (Morton & Munakata, 2005; Munakata & McClelland, 2003). One case of emergent dynamics is the nonlinear pattern of development that can arise from linear mechanistic changes. For example, Munakata (1998) has shown that linear changes in connectivity strength can cause nonlinear changes in perseveration such that infants' development is characterized by a *U*-shaped curve from no perseveration, to perseveration, to no perseveration. Similarly, Li et al. (2007) have shown that an exponential vocabulary spurt can occur based on linear changes in the amount of exposure to words.

Another example of emergent dynamics is the interaction of behaviors on different time scales (McClelland & Vallabha, 2009). Weight changes that are a product of long-term learning influence the current activations of the network, which, in turn, affect further weight changes. For example, Rakison and Lupyan's (2008) simulation showed that long-term experience with objects in the real world can constrain the feature correlations that children can learn in the lab. Similarly, Colunga and Smith (2005) have shown that linguistic experience, as evidenced by vocabulary size, can constrain the manner in which children generalize novel nouns such that they learn to use different properties to generalize about different types of objects. Both simulations highlight the emergence of complex constraints on behavior from domain-general associative learning as long-term accumulated weights affect the moment-to-moment behaviors. Taken together, the examples discussed above demonstrate the sufficiency of the same basic principles of connectionism for acquiring a range of information. This puts in question the need for biases, innate constraints, and specialized mechanisms to explain developmental changes.

Summary

Connectionist models can be a powerful tool in the formulation of theories of infant development through three contributions. First, predictions made by connectionist models can be an efficient way of identifying the specific ages that should receive focus in behavioral research and can be instrumental in disambiguating between different theoretical accounts. Second, connectionist models instantiate verbal theories of the mechanisms that underlie developmental changes and provide an explicit test of whether the mechanism can, in fact, produce the expected results. Finally, models can create a unified framework that highlights similarities in learning and development within and across domains. Taken together, these three contributions of connectionist networks are critical for the building of developmental theories. We now turn to a discussion of the two ways in which development has been modeled: based on changes in neural processing and based on changes in experience. We relate this discussion to the contributions discussed above to provide concrete evidence of the added value of connectionism.

Modeling Developmental Changes Through Changes in Neural Processing

One common way in which infant development has been modeled is through changes in neural processing. This approach assumes that the developmental changes that are observed in behavioral experiments are underpinned by development in the brain structure of the infant. As this development unfolds, it leads to a change in the way in which information is processed by the brain, which in turn leads to a change in behavior.

Commonly Modeled Phenomena

There are four basic neural developments that have been simulated in connectionist networks to account for changes in behavior. The first is the developmental improvement in visual perception. The second is the increased integration of information: As children develop, information that is processed in different areas of the brain becomes more integrated. The third is the developmental improvement in maintenance of information in short-term memory. The last neural development that has been modeled is neural plasticity, or the changes in brain structure through the addition of new neurons, removal of old neurons, or changes in the interconnectivity of the brain.

Perceptual development. One way in which neural processing changes throughout infancy is via improvements in visual perception (for a review, see Kellman & Arterberry, 2006), which impact neural processing because they impact the amount and the quality of information that can be processed. In terms of behavior, visual acuity has typically been measured with preferential looking procedures that assess both spatial frequency sensitivity and contrast sensitivity. Spatial frequency sensitivity is a measure of the highest number of vertical gratings per degree of visual angle that can be discriminated from a solid block of color. At birth it is approximately 40 times poorer than in adulthood and requires another 4 to 6 years to reach adult levels (Maurer & Lewis, 2001). Contrast sensitivity is a measure of the ability to see items of different luminance in an image. Similar to spatial frequency, contrast sensitivity is also below adult levels at birth and can take up to 15 years to reach adult levels (Maurer & Lewis, 2001). Such visual constraints can severely limit the amount of information that infants extract from the environment.

There are several examples of models that have implemented developmental changes in visual perception to account for changes in behavior. These networks can be broadly understood as implementing a perceptual and not a cognitive change, because the change occurs in the information that enters the system: Development is simulated by allowing the model to obtain more or higher quality information from the environment while the processing of that information remains the same.

One example of the implementation of perceptual improvement comes from Gureckis and Love's (2004), simulation of infants' performance in the classic Younger and Cohen (1986) study on learning correlated attributes. In the study, 4-, 7-, and 10-month-old infants were presented with line drawings of animals in which two features correlated consistently, and the third feature varied (e.g., the features of specific animals may be coded as 111, 112, 221, 222, where the first two features go together, and the third varies). They were then tested on new animals with features that retained or violated the correlation (e.g., 112 and 121, respectively). By 10 months, infants looked longer when the correlation was violated than when it was retained. Gureckis and Love (2004) hypothesized that this change was a result of improvement in visual acuity. As a test of this hypothesis, the model was trained on inputs designed to represent the behavioral stimuli, and Gureckis and Love manipulated the noise in the network to simulate development of visual acuity. The input for younger networks contained relatively high noise, which distorted the individual training exemplars. This resulted in a failure to learn the training categories because all of the inputs were indiscriminable. In contrast, the older networks contained less added noise and therefore less distortion, which meant that the relevant features that distinguished category members could be encoded.

A somewhat different perceptual improvement approach was taken by Shultz and Cohen (2004) and Shultz (2010). Their simulations of the same task implemented an adjustment of the threshold parameter in a cascade-correlation network, which sets how closely the network's outputs have to match the targets before training stops. The match had to be closer for older networks than for younger networks, so the older network underwent much more training than younger networks to obtain a closer match between the output and the target. This was considered a perceptual improvement because, according to the authors, the older networks extracted more information than the younger networks did from the stimuli. Subsequently, older and younger networks engaged in the same cognitive process of comparing internal representations to the external input, which suggests that no developmental change was implicated in cognition.

Finally, Dominguez and Jacobs (2003) used yet another approach that explicitly built different levels of visual acuity into their network to examine how changes in visual acuity can help in the detection of binocular disparity. The model was presented with an object in the left and right visual fields and was trained to identify the amount of disparity between the two images. Dominguez and Jacobs used a feed-forward network that included three groups of units that filtered visual input according to low, medium, or high spatial frequency. On the basis of these filters, the model activated a single output unit whose activation magnitude reflected the amount of disparity that the model detected between the two visual fields. Throughout training, the model adjusted its weights based on feedback. Development was modeled by activating only some of the filters and manually turning off others. Dominguez and Jacobs found that binocular disparity was acquired only when the filters were activated in a coherent progression (low, medium, then high; or high, medium, then low). Those networks for which the filters were all activated at once or were activated at random did not perform as well. Once again, this model implements a perceptual development, because there is a change in the amount of visual detail that is processed. These results suggest that the development of visual acuity in humans from low spatial frequency to high spatial frequency sensitivity may be functional for the acquisition of binocular disparity.

These three approaches to modeling development in visual perception all assume two basic tenets: first, perception improves throughout infancy, and second, these improvements account for changes in behavior. Improvements in perception are implemented in similar ways through a change in the quality or amount of encoded information. The difference in these approaches comes from the level at which the mechanism is specified. Gureckis and Love (2004); Shultz and Cohen (2004); and Shultz (2010) left it relatively unspecified: The assumption was that visual input improves with age, with no specification of how this may happen. In contrast, Dominguez and Jacobs (2003) provided more detail for this mechanism, suggesting different levels of information filtering. It may have been necessary for Dominguez and Jacobs to be more specific in their mechanisms than the others have been because they were interested in the emerging properties of the visual system.

What have these three models contributed to our understanding of the link between perceptual development and changes in behavior? First, the models of the Younger and Cohen (1986) work generate a testable prediction: If it is theorized that visual acuity hampers stimulus discriminability, infants should succeed in the task when stimuli with more distinct features are used. A behavioral test of this prediction could provide supporting or disconfirming evidence for the theory. Second, all three models show the sufficiency of the proposed mechanism of change for producing the target behavior. Although verbal theory can specify the hypothesized mechanism, it is possible that a concrete instantiation of this mechanism would not result in the expected behavior because of a missing component. For example, Dominguez and Jacobs' (2003) verbal theory of the relationship between visual acuity improvement and the emergence of binocular disparity could have missed a factor, such as infants' increased mobility, that may alter visual experience and affect binocular disparity. However, the model's instantiation of this verbal theory demonstrated that the hypothesized components were, in fact, sufficient to produce the behavior, suggesting that the mechanism operates in the theorized manner.

Integration of information. Another common neural development that has been implemented in models is the increase in integrated processing of information in the brain. That is, different pieces of information in the input initially may be processed by separate areas of the brain but over developmental time come to be integrated into a unified internal representation. Recent work in neuroscience has shown that, over development, functional connectivity in the brain changes from being segregated, such that spatially close regions show strong correlations, to being integrated, such that spatially distant regions show strong correlations (Field et al., 2009). One example of the development in information integration is the model constructed by Mareschal et al. (1999), which used the development of dorsal and ventral stream integration as the underlying basis for changes in infants' interaction with objects.

The dorsal and ventral streams are the two processing pathways of the visual system (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982; Van Essen, Anderson, & Felleman, 1992). The dorsal stream processes information related to visually guided attention and action and ultimately feeds into the posterior parietal region, whereas the ventral stream processes object features necessary for object identification and feeds into the inferotemporal cortex. In terms of development, local responses to individual scene components emerge earlier in the ventral stream than the dorsal stream. For example Braddick, Birtles, Wattam-Bell, and Atkinson (2005) have shown that visual evoked potential responses to orientation reversal, processed by the ventral stream, are present in infants starting at 4 weeks of age. In contrast, responses to direction of motion reversal, processed by the dorsal stream, were observed in less than 25% of the infants prior to 7 weeks of age, and reliable responses were not found in most infants until 11–13 weeks of age. In contrast, global responses based on the integration of these components emerge earlier in the dorsal stream than the ventral stream. For example, Braddick and Atkinson (2007) found that infants can respond to motion coherence in a visual display by 8 weeks, but it takes several extra weeks for responses to global form to develop.

Mareschal et al.'s (1999) model implemented the dorsal and ventral pathways and their integration. The model received visual input on a "retina," which then connected to two separate banks of units: the object recognition module and the trajectory prediction module. The former represented the ventral stream, and its units came to respond selectively to particular objects regardless of their location on the retina. The latter represented the dorsal stream, and its units learned to predict the next object location. Mareschal et al. did not make theoretical commitments regarding their choice to label these systems as "modules"; this label may have been used to specify that sets of units processed a different type of information, and they did not communicate directly with each other. According to Mareschal et al., infants' looking and reaching responses both require the dorsal stream because object location must be tracked. However, reaching for an object also involves the ventral system, because the voluntary retrieval of an object requires an evaluation of its identity and desirability. The results showed that dorsal-stream-based responses emerged earlier than responses requiring both streams, suggesting that the lag between looking and reaching measures observed in behavioral studies may be due to prolonged development of the integration between the two systems. Both systems were immature and produced imprecise outputs. When the behavior relied on only the dorsal system, as was the case with looking, an accurate response could be produced early in training because only a single imprecise system was implicated. However, when the behavior relied on the integration of dorsal and ventral streams, more training was necessary to produce an accurate response because it required the coordination of two imprecise sources of information.

This instantiation of the two visual processing streams is consistent with the classic understanding of their information-processing characteristics: The ventral stream codes features of objects in a position-invariant manner; the dorsal stream codes object location and supports actions toward objects (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982; Van Essen et al., 1992; for more recent evidence showing that ventral stream neurons are sensitive to position, see Aggelopoulos & Rolls, 2005; DiCarlo & Maunsell, 2003; Op De Beeck & Vogels, 2000). Furthermore, it is consistent with the protracted time course of the development of dorsal and ventral stream integration (Braddick & Atkinson, 2011). Mareschal et al.'s (1999) model provides an alternative explanation for infants' behavior in object permanence tasks, because it suggests that information integration, and not short-term memory as suggested previously (Munakata et al., 1997), contributes to the emergence of successful performance. Thus, the model demonstrates that the integration-based theoretical account can explain the data, thereby prompting further behavioral experimentation that can disambiguate between the two theories.

Another example of the development of integrated processing comes from Mareschal and Johnson's (2002) model of unity perception experiments (e.g., Johnson & Aslin, 1996; Kellman & Spelke, 1983), in which infants are habituated to a partially occluded rod and then tested on a complete rod or a broken rod. The network received input about the movement of the rod behind the occluder that was then sent to a layer of separate banks of units, referred to as encapsulated feature detection modules by Mareschal and Johnson. Each module detected the presence or absence of one of seven features, such as parallel object parts, comotion of object parts, and background texture. The modules detected the presence of these features but did not adjust connections through learning. They were encapsulated in the sense that there were no interconnections between the modules within that layer of the network; instead, each module fed into the next layer of hidden units. According to Mareschal and Johnson, these modules were "analogues" of the visual system but they were not designed to represent specific anatomical features. From the hidden units, information was sent to the output units that coded whether there were one or two rods as a binary output. This was not meant to represent explicit counting of the rods but rather a perception of the display as being unified or not unified. Over training, the model learned to assess the unity of different displays that were similar to the behavioral displays of one or two rods.

Mareschal and Johnson (2002) argued that the ability to integrate the information from these different modules explains behavioral changes. Initially, the model perceived the individual components of the visual information, but it could not combine them coherently to perceive the unity of the display. However, over time it developed weights that were effective in integrating information and creating a representation of an occluded single rod. Similarly, infants can detect all of the features that are present in the habituation display (e.g., that the two parts of the rod lie on the same line; two parts of the rod are moving together), but these features are not used together in perception. Over time, infants learn to integrate the features to generate a coherent internal representation (e.g., perception of a single rod based on collinearity and comovement).

Taken together, Mareschal and Johnson's (2002) and Mareschal et al.'s (1999) work illustrates that developmental changes in behavior can be accounted for by changes in the integration of information: Over time, infants proceed from basing responses on individual pieces of information to basing them on a joint representation of all available information. These models illustrate the connectionist contribution to a unified understanding of development. Connectionist networks naturally integrate the available information by gradually developing a weight structure that takes into account the similarity structure of the input. The success of these networks in modeling different phenomena suggests that this type of gradual learning can explain emergent behaviors across domains. Furthermore, it is consistent with findings on increased functional integration among brain regions over development (Field et al., 2009).

Maintenance of information. Another common way to simulate developmental changes in processing has been through the improved ability to maintain information over a delay, or increases in short-term memory, typically attributed to the prefrontal cortex in behavioral research. In now classic research, Goldman-Rakic (1987) showed that neurons in the monkey's dorsolateral prefrontal cortex maintained their activity over a delay period during which the monkey had to remember a target; this activity was interpreted as an internal representation of the target. Similarly, Braver et al. (1997) found that activity in the human prefrontal cortex correlated with the difficulty of a working memory task. Káldy and Sigala (2004) suggested that the brain regions that support working memory in infancy, particularly object-location memory, are more widespread and include the entorhinal, perirhinal, parahippocampal, and posterior parietal areas in addition to the prefrontal cortex. Further, the frontal cortex develops more slowly than other areas of the brain such as the visual and the auditory cortices (Huttenlocher, 1990; Huttenlocher & Dabholkar, 1997). For example, adult levels of neuronal density in the visual cortex are reached at about 5 months of age; in the prefrontal cortex, these levels are not reached until about 7 years of age (Huttenlocher, 1990). Thus, areas that support short-term memory develop throughout early childhood.

Short-term memory is often represented by the sustained activity of individual units in connectionist networks. Development is modeled by increasing the network's ability to sustain activity over time. For example, in a model of the A-not-B error, Munakata (1998) and Stedron et al. (2005) manipulated the persistence of hidden unit activity by changing the strength of the recurrent weights on these units. These recurrent weights were designed to reactivate the hidden units whose activity decayed over time. As a simulation of development of working memory, stronger recurrent weights were instantiated in networks that simulated older children than in those that simulated younger children. These stronger recurrent weights enabled the hidden units to overcome the decay in activation and maintain representations of hiding locations over a delay. The older models were able to sustain the memory trace of the new hiding location B and did not revert to responding at the old hiding location A. In contrast, the younger models lost the trace of the new location and continued to respond at the old location. Working memory improvement has been observed in empirical work, which showed that infants' ability to maintain information over a 300-ms delay improved between 6.5 and 7.5 months (Oakes, Ross-Sheehy, & Luck, 2006; for a dynamic field theory model of these data, see Perone, Simmering, & Spencer, 2011). Strengthening recurrent connections to simulate improvement in working memory is consistent with the available neural evidence as well as with behavioral evidence. Synaptogenesis in the prefrontal cortex continues well past the first year of life (Huttenlocher & Dabholkar, 1997), so changes in working memory over time could result from more developed prefrontal synapses of older children, akin to the stronger weights in the above models.

The development of working memory was also modeled—albeit in a different fashion—by Elman (1993) in an SRN of language learning and sentence processing. The network was trained on artificial sentences that were representative of the English language. As in a standard SRN, the hidden unit activity that represented the current word being processed was also influenced by the recurrent hidden unit activity that represented all of the previous words in the sentence. However, the simulation varied from a standard SRN in that this recurrent input was eliminated at random time intervals. The amount of time between these eliminations increased with age, which enabled the network to have a larger memory span. This developmental increase in memory span is consistent with empirical work showing that working memory capacity increases throughout development (e.g., Káldy & Leslie, 2003, 2005).

Munakata's (1998) work and Elman's (1993) work focus on two aspects of developmental improvement in working memory: strength and span, respectively. The former models stronger connections that enable the maintenance of information over longer periods of time. The latter models larger memory capacity, such that a greater amount of previous input is maintained in memory while a new piece of input is processed. Analyzing the two models together reveals a common underlying framework that explains the two types of developments in working memory. The same basic principles of connectionism can explain infants' increased ability to maintain a representation of a hiding location and infants' increased memory span of recently presented verbal information. This provides a more unified understanding of working memory development in the first 2 years of life.

Neural plasticity. The final way in which developmental changes in processing have been implemented in connectionist networks is through the plasticity of the brain, or the change in brain structure throughout an individual's lifetime. The brain remains flexible well after birth through three primary processes: neurogenesis (birth of new neurons), synaptogenesis (creation of new synapses between neurons), and synaptic pruning (elimination of synapses). Postnatally, neurogenesis is restricted to the subventricular zone and the dentate gyrus of the hippocampus (Stiles & Jernigan, 2010). Synaptogenesis and pruning, however, are more widespread, although their time course varies by brain area (Casey, Tottenham, Liston, & Durston, 2005; Huttenlocher & Dabholkar, 1997). Synaptogenesis peaks earlier in sensorimotor areas than in the prefrontal cortex, and in accord with this, synaptic pruning also begins earlier in those areas.

Several modeling techniques have been used to account for behavioral changes due to neural plasticity. One common approach has made use of cascade-correlation algorithms (Fahlman & Lebiere, 1990); these are well suited for modeling neural plasticity because of their dynamic network structure, in which new hidden units are installed throughout training. This approach typically is interpreted as modeling synaptogenesis (Shultz, 2003). However, it should be noted that although the hidden units are added as a natural part of network training, the algorithm that specifies when a unit should be added may not reflect neural processing dynamics that underlie synaptogenesis. Thus, this instantiation should be taken to represent broad principles of neural plasticity, not the specific mechanism by which synaptogenesis occurs. These models have been applied to studies on category learning (Shultz, 2010; Shultz & Cohen, 2004), false belief detection (Berthiaume, Shultz, & Onishi, 2013), and rule learning (Shultz & Bale, 2001).

In a model of vocabulary acquisition, Mayor and Plunkett (2010) implemented synaptic pruning in a self-organizing maps, which learned associations between visual and auditory input. They modeled synapse proliferation between two self-organizing maps by creating a high number of random connections. At the end of training, connections with small weights were eliminated. This is akin to the processes of synaptogenesis and pruning: The former forms some maximum number of synapses in early childhood, and the latter eliminates unnecessary synapses to eventually settle at adult levels (Huttenlocher & Dabholkar, 1997). Another example of the use of self-organizing maps to model neural plasticity can be found in models of language acquisition by Li et al. (2004) and Farkas and Li (2002).

A final way in which neural plasticity has been modeled is through the manual addition of hidden units in a static network. In contrast to cascade-correlation frameworks, which recruit additional hidden units as training proceeds, this approach involves the use of different model structures at different ages. For example, Rakison and Lupyan (2008) constructed a network that had fewer hidden units to simulate younger infants than to simulate older infants.

A key aspect of implementing neural plasticity in a network is to compare the time scale of the target behavioral phenomenon to the time scale of the hypothesized neural process. According to Shultz and Bale (2001), the cascade-correlation algorithm is representative of synaptogenesis and not neurogenesis in simulations of habituation studies because only the former can take place within the short span of a typical experiment. Synapse formation between individual neurons can occur in minutes, but large-scale changes in connectivity between distant brain areas take months or years to accumulate (Stiles & Jernigan, 2010). Mayor and Plunkett's (2010) model on vocabulary acquisition (see also Farkas & Li, 2002; Li et al., 2004) and Rakison and Lupyan's (2008) work on infants' learning of motion properties concern more protracted developmental changes on the order of months. Neurogenesis, large-scale synaptogenesis, and pruning would all be appropriate candidates for modeling the findings of those studies. Thus, slightly different implementations of neural plasticity may be appropriate depending on the amount of time over which the underlying neural changes are thought to occur. However, despite the variations in implementation, these networks provide a broad level of unification across specific experiments. They demonstrate that a change in neural connectivity is sufficient to explain behavioral results, or, more generally, that mechanistic dynamics of the system can be a foundation for more complex emergent dynamics of behavior.

Issues in Modeling Development of Neural Processing

Modeling behavioral changes as being based on neural development is not without its challenges, and two issues are particularly prominent: balancing neural plausibility and simplicity, and providing the source for the changes in processing. The first issue emphasizes simulations that reflect the fundamentals of neural processing and include only the necessary and sufficient neural constraints. The second issue emphasizes the instantiation of a mechanism of change that is appropriate given the extent of the neural research on the phenomenon in question.

Neural plausibility versus simplicity. Connectionist networks are inspired by principles of neural processing. However, how accurately is this processing represented in a given model or the connectionist approach as a whole? In our view, the instantiation of neural processing in these networks balances neural plausibility and simplicity. This balance is evident in two areas: which neural components are included in the model and how they are instantiated.

Building as many neural constraints as possible into a model can be intuitively appealing, because the model appears to be a better approximation of the infant. However, neural plausibility must be balanced with simplicity in the course of selecting which components to include. When neural constraints are hypothesized to impact behavior, these constraints should be instantiated in the model to verify this hypothesis (e.g., Munakata's 1998 instantiation of strengthening prefrontal connections as the causal factor behind reduction in perseveration). However, neural constraints that are unrelated to the target phenomenon should not be built into the system simply to make the model more biologically plausible (e.g., building a prefrontal system in a model of auditory processing). This is because incorporating such assumptions will only make for a cluttered and less comprehensible model. If multiple systems are built into the model, it becomes less clear what each system contributes to the outcome because there are numerous, constantly changing connections between units. When the unnecessary components are removed, the processing in the model becomes more transparent because the number of interconnections is reduced. Additionally, when an overly complex system fails to simulate a particular behavior, it can be difficult to determine if a single component caused the failure—and if so, which component it was—or if the interaction of the components was responsible. Finally, there are practical constraints on large-scale models such that often there may not be enough computational power to include all possible neural components.

One solution to the challenge of including just the necessary neural components is to rely on research in neuroscience that specifies which components are directly involved in the target behavior. Furthermore, modelers can assess the necessity of a given neural assumption in the model by removing the assumption and measuring performance. If the model continues to simulate behavioral data, this assumption may be superfluous. The application of these criteria is particularly useful for theory revision: The model can test which components of the theory are necessary and sufficient for the behavior.

It should be noted that, according to these criteria, there is no fixed cutoff for what defines a simple or a complex model. That is, there is no specific number of units, connections, or training patters that determines if a model is simple or complex. Rather, the simplicity of a model is defined relative to the behavior being simulated. If the behavior is very complex (e.g., performing a multisensory task) or if the network provides a unified simulation of several experiments, it may be necessary to build a larger network with more units or connections than what would be necessary for a simpler behavior or a simulation of a single experiment. In absolute terms, the former network would be more complex than the latter. However, if both networks balance simplicity and neural plausibility as discussed above, such that only the neural components implicated in the behavioral phenomenon are included in the simulation, then both can be considered to be appropriately simple relative to the behavior in question.

Once the relevant neural components have been selected for modeling, the accuracy of their instantiation in the model must be assessed. To answer that question, we draw once again on the principle of balancing neural plausibility and simplicity. Connectionist models have largely focused on cognitive phenomena, and this focus requires some amount of simplification because the phenomena are at a much higher level than the underlying cellular processes (Thomas & McClelland, 2008). This simplification is necessary to constrain the size of the model, so that the result is a model that is understandable and computationally feasible. Thus, although an exact replication of the brain may be desirable, the usefulness of such a model would be minimal because it would be just as difficult as the brain to understand (O'Reilly & Munakata, 2000). Crucially, simplification should be used in a way that maintains the key properties of the target phenomenon.

Connectionist networks have adopted a variety of simplifications of neural processing. For example, training with backpropagation of error is not biologically plausible because signals cannot be sent backward through neurons. However, this simplification enables computational efficiency in training and has been shown to be equivalent to a biologically plausible way of training a network

(Xie & Seung, 2003). Thus, although not plausible in itself, it captures the key properties of learning in a biologically plausible system. Another example of simplification comes from unit activations. Although neural processing can be modeled at the fine-grained level of individual neural spikes, such that the activity of a single network unit represents a single neuron, a population rate code typically is adopted, such that a unit represents the average output of a population of neurons (O'Reilly & Munakata, 2000). In employing this simplification, the model loses the details of interaction between individual neurons. However, the resulting unit activity is representative of the activity over a whole brain region, a level of approximation that is sufficient for modeling a cognitive task. A final example of simplification can be seen in the simulation of neural plasticity as described above. Neurogenesis, synaptogenesis, and pruning have been modeled through the addition of hidden units or the proliferation or elimination of unit connections. These instantiations clearly omit many of the details of the processes. For example, they disregard the migration of neurons from the place of formation to their final destination, myelination of axons that increases their conduction velocity, or the processes by which axons are guided to their targets (Stiles & Jernigan, 2010). Furthermore, it may not always be clear which specific process is represented: For example, the addition of a hidden unit can represent the birth of a new neuron or a newly formed connection to an existing neuron. Despite the simplifications, this representation does capture some basic features of neural plasticity, such as the developmental growth of processing capacity and changes in connectivity. Taken together, these three examples demonstrate that connectionist models are simplified to capture the relevant features of neural processing in a manner that allows for a network that is computationally feasible with a transparent and accessible mechanism of operation.

Source of the neural development. Another challenge that modelers must face is how to implement the development in processing that may emerge naturally within the system or may be applied externally by the modeler. Some developmental simulations employ the former approach. For example, cascade-correlation networks automatically add hidden units throughout model training (Fahlman & Lebiere, 1990). However, in a large number of developmental simulations the modeler interferes with network training to administer the developmental changes by hand rather than having these changes occur without any further manual input after the initialization of the simulation. One example of the manual approach can be found in Shultz and Cohen's (2004) simulation of a correlated attribute learning task, in which the score threshold parameter value, the criterion for how closely network outputs had to approach the targets before training stopped, was manually altered by the modelers to simulate development.

On the surface, the manual approach might appear to be problematic for models of developmental phenomena, and it has been criticized by several researchers (e.g., Dehaene, 1998; Smith & Scheier, 1998; Younger et al., 2004). It may seem that a model focused on development should encompass the full chain of events that bring about a behavioral change. For example, if increases in memory capacity are thought to underwrite the developmental change, the model should be able to grow its own memory without the modeler interfering and adding hidden units. One response that has been provided to this criticism is that the mechanism of processing changes often has not been adequately specified by neuroscientists and therefore should not be included in a model (Westermann & Mareschal, 2004). If there is no consensus about how the processing develops, making changes externally instead of building in an incorrect mechanism of change is appropriate. Moreover, how the neural development occurs may be beyond the scope of interest; instead, the interest may be primarily in how that neural change brings about the behavioral change. Modeling this latter question is challenging in itself and is an important first step in understanding the behavioral change. Simultaneously adding the mechanism by which the neural change itself occurs may make it difficult to comprehend how the various components of the network bring about the behavior. This may render the model less useful, especially for understanding the causal link between neural and behavioral changes.

Our goal is not to argue that neural processing changes should always be made by hand. Rather, initial efforts to instantiate a behavioral phenomenon in a connectionist network should do so, because the first step is to understand whether the hypothesized neural change can bring about the behavioral change and how that process can occur. These questions can be answered most adequately in a model that does not include extraneous mechanisms. However, once these questions have been addressed in a simplified network, one can proceed to the next step in the chain and simulate the process by which the neural changes occur. In our view, the models with hand-administered neural changes described in this review have made important contributions to our understanding of the behavioral changes in infancy. They also provide solid basis for further exploration into the way in which neural changes can emerge naturally within a simulation.

Summary

There are four common types of neural developments that have been implemented in networks to account for changes in behavior: those in perception, integration of information, working memory, and neural plasticity. Taken together, these four approaches face two primary challenges: balancing neural plausibility and network simplicity and identifying the source of the changes in processing. All four neural developments are occurring in infancy. However, it is rare for modelers to implement all of them in a single network, because usually only the neural change that is hypothesized to be responsible for the behavioral change is included in the model. Once the candidate neural changes are selected for modeling, they are further simplified, which typically allows the modeler to develop a network that can be easily analyzed and understood and that is not computationally overbearing. Furthermore, many modelers and developmental researchers are interested in the interplay between neural and behavioral development, not the process by which the neural development itself occurs. As a result, they simplify further: They omit the mechanism that underwrites the neural change and manually administer it instead, particularly if the mechanism behind the neural change has not been adequately explored in behavioral research. In our view, this provides a critical foundation for understanding the link between neural development and behavior and can spur further modeling efforts that specifically focus on the source of the neural development itself.

Modeling Developmental Changes Through Changes in Experience

Another common approach to modeling developmental change focuses on changes in experience. The theory behind this approach is that developmental changes in infants' behavior are a product of infants' learning about their environment. There is a long tradition of research, starting with Piaget (1952), that studies the role of experience in developmental change. This research spans a broad range of domains such as motor development (e.g., Adolph, 1997), visual expertise (e.g., Scott & Monesson, 2010), and language processing (e.g., Werker & Tees, 1984). Thus, modeling work that employs this theoretical stance is well grounded in empirical research.

Commonly Modeled Phenomena

There are two common phenomena related to changes in experience that have been modeled by changing the training pattern set. The first is based on the total amount of experience: Older children may have more experience than younger children. The second is based on the type of experience: Older children may have different experiences than younger children.

Additional experience. By definition, older children have more exposure than younger children do to particular events in the environment. Accordingly, to model an older child the network must receive more epochs of training with the stimulus set that represents the child's experience than needed to model a younger child. For example, Rogers and McClelland (2004) modeled conceptual development by assessing their network's performance throughout training: Early and late time points corresponded to younger and older children, respectively. Similarly, Franz and Triesch (2010) exposed older networks to more experience with object movement than younger networks received prior to exposing both to the habituation and test stimuli used in the object unity perception experiments. Other examples of the use of additional training to model developmental changes can be found in Munakata et al. (1997); Rakison and Lupyan (2008); and Schafer and Mareschal (2001). Additional training is used to model development when the underlying assumption is that the overt changes in behavior are due to older children's greater experience with a particular domain. The success of this approach across domains (e.g., unity perception, learning of correlations, perseveration) demonstrates that the accumulation of experience can cause a range of complex behavioral changes.

It is important to note that there may not always be a direct correspondence between the number of additional training trials in the model and the difference in age; that is, the amount of training in a network may change nonlinearly as age changes in infants linearly (Franz & Triesch, 2010). This may not be a fault of networks, and it may align with infants' exposure to the environment. As children age they spend more time awake (Halpern, MacLean, & Baumeister, 1995), so it is possible that the amount of experience acquired in the first month of life may be less than that acquired in the fourth month of life. Thus, the total amount of experience would change nonlinearly, because each subsequent month of life adds more experience than did the previous month due to more time spent awake.

Changing the type of experience. As an alternative approach to modeling development, modelers may opt to change the training set. The logic behind this strategy is that children may have different experiences as they become older—for example, by changing what they attend to in a scene—and therefore the experience provided to the older network should reflect this change. One example of this approach is Elman's (1993) model of language learning. As the network was trained, the ratio of complex to simple sentences in its training set increased. However, Elman acknowledged that this may not have been a plausible instantiation of language exposure, because it is unlikely that children's experience would shift from only simple to only complex sentences. Another example of a changing training set can be found in Mayor and Plunkett's (2010) model of the development of joint attention and language. Infants are initially unable to engage in joint attention, so when an adult labels an object, infants cannot use the adult's gaze direction to detect the labeled object (for a review, see Mundy & Van Hecke, 2008). However, once infants can engage in joint attention, they can detect simultaneously the labels and their referents. To model this change, Mayor and Plunkett (2010) initially trained the network separately on objects and labels and then shifted to their simultaneous presentation.

Note that examples cited above addressed language development (see also the language acquisition models of Farkas & Li, 2002, and Li et al., 2004, which employ a growing training set). To our knowledge, no simulations of infant behavior in which a visual training set changed with age have been conducted. It is plausible, however, that such an input change may occur, because once infants sit up, crawl, and then walk they have access to a greater variety of environments (Campos et al., 2000). Therefore, an outstanding issue is how developmental changes in infants' visual environments may contribute to the differences in their behavior.

Issues in Modeling Changes in Experience

As with neural-development-based approaches, simulations that employ changes in experience encounter a number of challenges. Among the most common challenges are those related to the generation of training events, interference during different phases of learning, and the role of feedback in learning. The first and the third challenges require modelers to ensure that there is a tight coupling between the child's and the model's experiences. The second challenge requires an alternative theoretical basis for the way in which infants store learned information.

Generating the appropriate experience. Generating the training stimuli for the network can be challenging because they must reflect the experiences that infants are thought to have, which is particularly critical when those experiences are theorized to be the causal factor behind development. Two factors must be considered when the training set is generated for a given network: whether it is generated based on real measurements of the environment or on approximations, and the amount of training that should be provided. In a way, the problem of generating appropriate experiences for a network parallels the problem of which neural components to instantiate in the network, as discussed above. With respect to neural components, we advocated for the inclusion of only those that are causally relevant to the developmental change. Similarly, it is necessary to select only those experiences that are causally relevant to the development. We return to this parallel later, in the section titled “Common Challenges for Brain-Based and Experience-Based Approaches.”

It is relatively straightforward to generate training examples based on real measurements when specific experimental stimuli are available, because the researcher has simply to recode these stimuli according to some scheme that preserves the features of interest. For example, Christiansen and Curtin (1999; Christiansen, Conway, & Curtin, 2000) used an 11-feature phonological coding scheme to recode the sounds presented by Marcus et al.’s (1999) study on rule learning in infants. Each phoneme used by Marcus et al. was represented across the 11 phonological feature input units according to the presence or absence of each feature in that phoneme. Additional units coded boundaries between syllable strings and the stress of each phoneme. Similarly, Mareschal et al. (2000, 2002) and French et al. (2004) used the Quinn et al. (1993) cat and dog stimuli in their behavioral experiments and took measurements of the stimuli to create training patterns for the model. Feature measurements of the cat and dog photographs (e.g., leg length) were converted to activation values across units coding those features. For example, shorter legs had a lower activation value of the “leg length” unit than did longer legs.

Nonetheless, as cautioned by McClelland and Plaut (1999), one must be careful when coding experimental stimuli because the modeler’s choice of representation may not align with the way that infants represent the stimuli. If a modeler represents the stimuli incorrectly, the model may still reproduce the infants’ behavior but for different reasons, because it would rely on different information. As a result, although the model may match infants’ behavior in the current task, it may fail to generalize to other tasks in which distinct ways of representing the input would result in different behavioral outcomes. For example, various researchers have used alternative coding schemes to those of Christiansen and Curtin (1999; Christiansen et al., 2000) to code the stimuli used in the Marcus et al. (1999) experiment: Shultz and Bale (2001) used a single sonority feature, whereas Gasser and Colunga (2003) used five phonological features. Unfortunately, it is extremely difficult to determine which, if any, of these coding schemes is the most appropriate or accurate. For example, the specific features of the auditory stimulus that are perceived and encoded by infants remain to be seen. As a result, it may be the case that a given model and infants rely on entire different sets of features to dishabituate to the test stimuli. Although in this case the different ways of representing the experimental stimuli may give the same outward behavior, there may be a task in which they would result in a mismatch between the model’s performance and the infants’ performance. For example, Shultz and Bale (2001) suggested that it may be possible to assess whether infants solely encode sonority, or vowel-likeness of a phoneme, in Marcus et al.’s experiment by changing the experimental stimuli such that they consist of different phonemes that have the same sonority value. If infants rely only on sonority, they should be unable to discriminate phonemes with the same sonority and should treat the consistent and inconsistent test items as equivalent. In contrast, if the behavioral results still match the original findings, this would suggest that infants likely rely on other features to discriminate the experimental stimuli. This provides an example of the theoretical contribution of modeling: Models point out areas in which our understanding is insufficient to build the mechanism that underlies the behavior; they make predictions for behavioral outcomes based on different theories; predictions are tested experimentally; and theories receive either supporting or disconfirming evidence.

In contrast to taking measurements of real experimental stimuli as a basis for the training set administered to the model, taking measurements of the general environment to which infants are exposed is much more challenging. Information is limited about the types of daily events that infants experience and learn about. The most extensive database of children’s early environments is CHILDES (MacWhinney, 2000), which covers over 130 corpora of transcripts of naturalistic linguistic interactions that involve children. Each corpus varies with respect to the context, children’s age, number of children studied, and language spoken. For example, the Korman corpus consists of transcripts of British mothers speaking to their infants, and the Hungarian corpus consists of transcripts of five 1- to 3-year-old children who are engaged in free play. Modelers have used CHILDES extensively to endow their models with linguistic experience that is comparable to what children may encounter in the real world (e.g., Christiansen, Allen, & Seidenberg, 1998; Li et al., 2004, 2007).

CHILDES is not the only tool modelers have used to provide linguistic experience to a network that is directly based on children’s experience. Li et al. (2004, 2007) used the MacArthur–Bates Communicative Development Inventory (CDI; Dale & Fenson, 1996) to ensure that the model was trained on the words that children acquire early in life. One advantage of using databases such as CHILDES and the CDI is that they allow the modeler to be specific about the age that is simulated. For example, Li et al. (2004, 2007) used the CDI to model the language capacity of children at a particular age by selecting the words that the average child knows at that age. In contrast, when the training set supplied to the model is based on estimates of the real environment rather than direct measurements, there is less specificity about the age being modeled, which, in turn, makes developmental predictions less precise.

Although measurements of children’s language experience are available for modelers to use, children’s visual environment has been less specified. There have been some recent efforts in this area in which head-mounted cameras were used to track children’s visual experience (e.g., Aslin, 2009; Cicchino, Aslin, & Rakison, 2011; Franchak, Kretch, Soska, Babcock, & Adolph, 2010). But these efforts have been limited, and only a few modelers have employed such data to specify the model’s visual input (e.g., Yu & Smith, 2011; Yurovsky, Hidaka, Yu, & Smith, 2010). However, generating network training sets based on data taken from video would allow for more precision in specifying the role of experience in developmental change. Therefore, it is necessary for modelers to make greater use of the current data available on infants’ visual experience. Similarly, it is necessary for behavioral researchers to study such experiences in a variety of daily contexts, cultures, and developmental periods, so that they may be cataloged in a manner similar to that of CHILDES (MacWhinney, 2000) and used in modeling work.

As a result of the difficulty in obtaining direct measurements of infants’ experiences, a number of modelers have opted to approximate these instead. For example, in a model of language learning, Elman (1993) used an artificial language that had characteristics of English to approximate the linguistic input of a young child. However, Rohde and Plaut (1999) have criticized this approach because the training corpus omitted the crucial semantic constraints that exist in English. In fact, Elman's (1993) results relied on the absence of such constraints, and a different outcome occurred when semantics were added to the input. Similarly, Mayor and Plunkett (2010) used constraints that they believed to be veridical of children's visual and linguistic environments to generate an artificial set of labels and objects for their model of language acquisition. For the labels, Mayor and Plunkett assumed that most words heard by infants are produced by the same speaker (e.g., the infant's mother); therefore, most word tokens given to the model were highly similar in pronunciation. For the objects, the they assumed that object categories have a similarity structure and that these categories sparsely fill the representational field. Accordingly, each category had a prototype that was distorted to different degrees to create category members. However, item generation based on prototype distortion may not be valid because it assumes a symmetric distribution of items around a center, which may not be the case (Johns & Jones, 2010).

One drawback to an artificial training set is that it may contain spurious correlations, or correlations between two things that happen to be correlated in an artificial training set that are not correlated in real life (Mareschal, 2003). They occur without the modeler's intention simply due to the coding of the different features of the training items. Connectionist networks are powerful statistical learners that will pick up on all regularities present in the input, including spurious correlations, which may negatively influence the networks' performance. For example, in Mareschal and Johnson's (2002) model of unity perception, the artificial training set had a weak correlation between the presence of background texture and the presence of two objects due to the way in which the background texture was coded. This spurious correlation negatively impacted the network's ability to learn an event that contained background texture and a single object. Alternatively, the network could rely on such spurious correlations outwardly to match the behavior of infants, but the model and the infants would behave similarly for different reasons. An artificial training set may also miss a correlation that is present in the environment. In such cases, the model may fail to simulate behavior because the necessary correlation that is available to infants is not available in the network's training set.

Once the set of training items has been generated, the modeler must select the amount of training for the model, in terms of both the item distribution (i.e., whether all items are seen with equal frequency) and the amount of training (how many times the entire training set would be presented). If the training examples were generated based on measurements of the environment, selecting the distribution of items can be relatively easy because it can be based on their real-world distribution. However, if the examples were artificially created, the modeler must decide whether the frequency with which individual items are presented would vary. For example, Cohen et al. (2002; see also Chaput & Cohen, 2001) had to infer the frequency of causal and noncausal events in the infant's environment for their models of causal perception. Cohen and colleagues assumed that causal events were more common than noncausal events in infants' daily experiences, so they provided the models with a large corpus of examples of which 85% were causal and 15% were noncausal. Other modelers have opted for an equal distribution of all experiences. For example, Smith, Gasser, and Sandhofer (1997) and Gasser and Smith (1998) randomly generated training items in their simulation of the acquisition of object feature labels, which made it unlikely that some items would appear more frequently than others.

In addition, the overall amount of experience must be specified. Typical infant experiments last just a few minutes and usually administer one to two dozen trials in total. Indeed, it is rare for infants to be exposed to hundreds of training trials in the laboratory. In contrast, models of these experiments vary widely with respect to how many training repetitions are administered. For example, Westermann and Mareschal (2004) presented the training set 1,000 times in their simulation of Younger and Cohen's (1986) habituation experiment. In their simulation of the same experiment, Gureckis and Love (2004) administered the same number of repetitions as the behavioral experiment with infants. However, in our view, a direct correspondence between the number of trials that networks and infants receive may not be necessary, because each network training trial may not be representative of a single exposure to the stimulus. Instead, it could be representative of the additional processing that is supported by the hippocampal system that re-creates previously presented patterns so that they may encoded in the cortex (Foster & Wilson, 2006; McClelland et al., 1995). Although there is no direct evidence that suggests that the infant hippocampus can re-create patterns, there is evidence that the hippocampus cell proliferation starts at the 24th gestational week and is largely complete by the end of the first postnatal year (Seress, Ábrahám, Tornóczky, & Kosztolányi, 2001). This provides some indirect evidence that the infant hippocampus, at least by the end of the first year of life, may be sufficiently developed to re-create information that has been presented and thereby support cortical encoding. Thus, the greater number of trials received by networks may reflect the additional processing supported by the hippocampal system.

Interference in learning. One issue in incorporating changes in experience that modelers commonly face is catastrophic interference, which arises when a model is trained to produce outputs to two distinct sets of stimuli in succession such that new experiences "overwrite" previous ones (McCloskey & Cohen, 1989). All of the weights initially are adjusted to accommodate the first set only; then, all become readjusted to accommodate the second set only, thereby greatly altering any knowledge that was stored about the first set. Catastrophic interference can be problematic when developmental models incorporate a "real-world" experience phase prior to an experimental lab phase or when the experience set is changed with development. For example, to simulate the way in which prior experience with speech could influence an infant's performance in a language-related experiment in the lab, a model would be trained on a set of prior experiences representative of the speech that infants normally hear throughout the day. Following this prior experience phase, the model would be trained and tested on the experimental stimuli. In this example, catastrophic interference could arise between the habituation training and the prior experience training. The weights that were initially developed to accommodate the prior experience stimuli would be altered by the habituation stimuli, such that the model may no longer provide the correct responses to the prior experience stimuli.

One way to avoid catastrophic interference in models is to use a complementary learning systems approach (McClelland et al., 1995; O'Reilly, Bhattacharyya, Howard, & Ketz, 2011). This approach is based on the idea that learning and memory are supported by two systems: the hippocampus and the neocortex.

The hippocampus encodes new information rapidly and reinstates it in the neocortex, which then slowly incorporates it into the current body of knowledge. There is relatively high overlap in the neocortical representations, which supports generalization; thus, information learned about one item extends to other items that have similar representations. In contrast, hippocampal coding is sparse, such that each unit represents a particular combination of features. This sparse coding reduces interference between the current stimulus and previously learned information, which allows the hippocampus rapidly to encode new information without significantly altering prior knowledge. These complementary systems could be used to implement different phases of learning in developmental models. A specific instantiation of this can be found in Rakison and Lupyan's (2008) SRN that simulated conceptual development in infancy. The two systems were represented by separate banks of hidden units: a "fast-learning" set of hidden units representative of the hippocampus and a "slow-learning" set of hidden units representative of the neocortex. All input units connected to all units in the slow-learning and fast-learning systems, and all units in those systems connected to all output units. The fast-learning system weights had a higher learning rate and a higher decay rate than did the slow-learning system weights. Training of the model proceeded in two phases. During pretraining, representative of infants' prior experiences, both the fast-learning and the slow-learning weights were adjusted, but during habituation, representative of the lab experience, the slow-learning weights were fixed and only the fast-learning weights were adjusted. This ensured that the habituation stimuli did not create catastrophic interference with the pretraining stimuli.

With respect to infants, the use of the separate weights was intended to represent the fact that the brief habituation experiment in the lab should not influence long-term memories of prior experiences. However, because the slow-learning weights were maintained, albeit in a fixed state, prior experience influenced the processing of the habituation stimuli. It should be noted that freezing long-term weights during habituation is a simplification, because infants do retain memories of habituation stimuli over a delay; for example, Pascalis, de Haan, Nelson, and de Schonen (1998) have shown that 6-month-olds recognize images of faces to which they were habituated after a 24-hr delay. Despite this simplification regarding the long-term encoding of habituation experiences, the model does consistently represent the hypothesized effect of prior experience on habituation learning. Thus, complementary learning systems may be a useful tool for developmental networks that model the relationship between prior experience and lab experience or that account for developmental changes by changing the set of training events.

Role of error-correcting feedback in learning. Another issue that can arise with respect to connectionist networks relates to the way in which learning occurs, with or without error-correcting feedback. Under the first option, weight adjustment occurs through the minimization of the discrepancy between the produced and the desired output. This is referred to as *supervised learning,* because a specific teaching signal that identifies the correct output is provided to the network. Under the second option, referred to as *unsupervised learning,* there is no teaching signal. The type of learning that is chosen for a network must be in accordance with the infant's experience: Does the infant receive feedback during learning?

A representative example of supervised learning comes from work on categorization by Quinn and Johnson (1997). They trained a three-layer backpropagation network on various animals and furniture that were coded across the input units. The network's task was to produce the basic (e.g., cat) and global (e.g., mammal) category of each item that was presented on the input. However, supervised learning may not be appropriate in this case. It is unclear what the source of the teaching signal would be in the network, because the model simulated the performance of very young infants who did not produce category labels explicitly and therefore could not have been corrected by a parent. In addition, parents rarely provide explicit corrective feedback, and children may ignore such feedback when it is provided (Brown & Hanlon, 1970; MacWhinney, 2004; Marcus, 1993). Thus, supervised learning in the form of an explicit external error signal and desired target may not always be a plausible way of modeling infants' learning.

However, it is possible to provide an error signal to a network that is internal to the infant. One approach has been to use encoder networks and SRNs, which involve supervised learning but attribute the error to an internal comparison between the external stimulus and its internal representation. During training, these networks learn to reproduce the input pattern on the output units. As a result, it can be argued that learning is essentially unsupervised because no agent actively provides an error signal to the child (Mareschal et al., 2002). For example, in an encoder model of infants' categorization of cats and dogs, the network compared its re-creation of the stimulus to targets that were identical to the original inputs (French et al., 2004; Mareschal et al., 2000, 2002). In some sense, this is similar to infants' comparison of their internal representation of the cat or dog with the animal that is displayed on a given trial. Similarly, in SRN simulations of object permanence, the network's prediction of the next input was compared to targets that were identical to that next input (Franz & Triesch, 2010; Munakata et al., 1997). Again, this was comparable to an infant who anticipates some visual input and then compares it to reality. This approach to modeling task performance is consistent with theories of infants' information processing: Infants' attention is driven by the formation of an internal representation of the stimulus and the comparison of this representation with the information present in the environment (e.g., Cohen, 1973; Sokolov, 1963).

Alternatively, it is possible to modify supervised learning algorithms such that feedback is provided in a more realistic manner. In a typical backpropagation network, error signals are provided to all output units. This results in a situation where there is feedback about all possible responses, which may not always be plausible. For example, Gasser and Smith (1998) argued that children do not receive all potential feedback from parents when they learn to label properties of objects; rather, the feedback is limited to the correct response. When the child learns to label colors, the child may consider multiple potential color names for a given object but may verbally produce just one. If the child says that a lemon is red, the parent may say, "No, it is not red, it is yellow," but the parent would not provide feedback about all other colors such as "It is yellow, it is not red, it is not blue, it is not green, it is not purple." In contrast, a backpropagation network receives the latter form of feedback, because it assumes that all of the colors that the child has considered, not just the one that was verbally produced, are evaluated according to their correctness. Gasser and Smith (1998; see also Smith et al., 1997) sought to make feedback administration more plausible when training a network to label object properties. The network learned to label the feature that was queried (e.g., object color) by activating the correct output unit, and an error signal was provided only to the output unit that indicated the correct response and to any other output unit that was active beyond a particular threshold. Thus, the network received two types of teaching signals: When it was correct it was given that feedback (akin to saying, "Yes, the lemon is yellow," if the child correctly identified the color), and when it was incorrect it was instructed that the particular response was incorrect and provided with feedback about the correct response. This type of feedback is more synonymous with a parent correcting only the verbal response of the child and not all of the options the child initially considered.

There are a number of contexts in which it is better to forgo supervised learning algorithms in favor of unsupervised learning because the latter does not use any error signal and therefore does not need to be modified for contexts when there is no explicit source of error correction. The most common unsupervised algorithm is Hebbian learning, which is often used to train self-organizing maps (e.g., Cohen et al., 2002; Li et al., 2004, 2007). Hebbian learning strengthens the weight between two units whose activity is correlated and weakens the weight between two units whose activity is uncorrelated. It is considered as similar to long-term potentiation and long-term depression processes in the brain (Munakata & Pfaffly, 2004). For example, Cohen et al. (2002) used Hebbian learning to modify the connections in self-organizing maps during training on causal and noncausal events. The maps learned the similarity structure of the examples with respect to their spatial, temporal, and causal characteristics. Unsupervised learning was appropriate in this case because it is unlikely that infants experience explicit feedback that specific events are either causal or noncausal. Hebbian learning has also been instantiated in other types of networks that do not form topographic representations but rather learn to activate particular output units given a certain input. For example, Colunga and Smith (2005) used Hebbian learning in a network that was trained to reproduce the names of objects when given input on their properties. However, it is unclear whether learning is completely unsupervised in this case, because it is likely that some feedback would have been provided in situations when the child produced the wrong label for the object.

We should note that, in addition to supervised and unsupervised learning, another option for learning exists in the form of *reinforcement learning* (Barto, 1995). This type of learning involves a system that performs actions that are evaluated by a critic. However, the system receives no input about the optimal action or the way in which its current actions should be changed to reach the optimal action. To our knowledge, in developmental models, this type of learning has been used only to study motor development (Berthier, Rosenstein, & Barto, 2005).

Summary

Two techniques have been used to model the impact of experience on developmental changes in behavior. The first technique focuses on quantity: Older infants' greater experience within a domain changes their behavior. In neural networks, this has been simulated through additional training: Networks that are intended to simulate older infants are given more epochs of training than those that are designed to simulate younger infants. The second technique focuses on quality: Older and younger infants have qualitatively different experiences, which causes different behaviors. This has been implemented in neural networks by changing the set of training examples over time. Both approaches allow researchers to show how experience can give rise to observed behaviors: The verbal theory that the two are related is substantiated by a working mechanism that demonstrates the sufficiency of experience. Furthermore, they highlight commonalities across studies that may initially appear unrelated. For example, accumulation of experience can lead to nonlinear patterns of development in language acquisition (Li et al., 2004) and in learning correlations between object features (Rakison & Lupyan, 2008).

Both techniques face a challenge of selecting the appropriate set of training stimuli and specifying whether learning about these stimuli will occur with or without feedback. Both of these aspects must be in line with infants' real-world experiences. However, there is often a lack of empirical evidence for infants' real-world experiences, particularly visual experiences. Thus, modelers often make general assumptions about infants' learning or experiences (e.g., infants likely get more experience with causal than noncausal events; Cohen et al., 2002), but these assumptions remain to be evaluated empirically. Building them into a network can provide a critical test of whether the hypothesized experiences can explain changes in behavior. An additional challenge is related to the catastrophic interference of new experiences with previous experiences during learning. Catastrophic interference can be avoided through complementary learning systems that are designed for long-term and short-term accumulation of information.

It should be noted that although there are fewer techniques that are used to model changes in experience than there are to model development in neural processing, this should not be taken to reflect the superiority of one approach over another. The two approaches are equally well supported empirically and provide equally strong accounts of developmental change. The discrepancy is simply a result of the complexity of neural changes—there are numerous changes that occur in early childhood, many of which have now been implemented in connectionist models. In contrast, changes in experience occur only in the amount or the type of experience.

Common Challenges for Brain-Based and Experience-Based Approaches

Numerous parallels can be drawn between models that simulate development based on neural processing changes and those that model development based on experience. These parallels relate to the challenges and solutions that arise for both approaches. There are four challenges commonly found: the balance between central and peripheral components in a network, parameter specification, inadequate research on the topic being modeled, and model comparison. These challenges are important to address because they may lead to misperceptions and mistrust of connectionist modeling.

Central and Peripheral Components in a Simulation

A given behavior is determined by a multitude of components, any of which could be instantiated in a model. For example, a model of children's performance in a sequential touching task could include the child's mood, experience with similar objects, the motor neuron activity that governs reaching, or the retinal ganglion cell activity in response to visual input, to name a few. Which of these components, if any, should be included in a developmental model that simulates change over time? We propose that to answer this question the candidate components should be classified as either *central* or *peripheral*. Central components are those that are theoretically relevant to the developmental change; that is, these components are part of the theorized mechanism that causes the change. In contrast, peripheral components are those that are not theoretically relevant to the developmental change; although these components are present in the behavior, they do not cause the change in behavior. It should be noted that the decision to classify components as central or peripheral may not always be straightforward. The components that are deemed central and peripheral can vary depending on the question being asked or the theory behind the developmental change. For example, if the researcher is asking a question at the level of cellular dynamics, different components may be included than if the question is asked at the level of neural systems dynamics. Although the decision as to which components are central and peripheral may be subjective and be driven by the modeler's interests, it should be supported by a clearly articulated theory that identifies the components that are thought to be involved in the mechanism of change.

Once central and peripheral components have been identified based on the theory of development, the modeler must decide which components should be included in the network. One alternative is to include only the central components. This approach to modeling has been referred to as the *fundamentalist* approach (Kello & Plaut, 2003), because it includes only those components that are fundamental to account for the behavioral phenomenon. Models that include all of the proposed central components should be consistent with the behavioral data, because it is those components that are hypothesized to give rise to the data (Seidenberg & Plaut, 2006). Although peripheral components may be added, they would not add to the simulation of the observed behavior. In the above example, one explanation for developmental changes in sequential touching performance could be greater experience with the objects that are presented. Thus, experience would be classified as a central component and included in the simulation. In contrast, if it is assumed that motor neuron activity does not change across ages, motor neurons would be classified as a peripheral component and might be omitted from the simulation.

An alternative to the fundamentalist approach is the *realist* approach, which prescribes that all known components of the behavior should be included, regardless of their potential contribution (Kello & Plaut, 2003). According to this approach, it is impossible to definitively identify which components are necessary to explain the behavior, and thus all candidates should be included. Although it may seem that this would yield the closest approximation of an infant, it is usually not advantageous. The inclusion of all components obscures the mechanism because it becomes less clear which ones cause the change. Therefore, prior to model construction the mechanism of change must be clearly articulated, so that all of its necessary components are apparent. This approach also allows for the most effective theory improvement: If all of the central components are included and the model does not replicate the behavioral data, the theory of development must be revised and different central components should be identified.

Central and peripheral components have to be identified both in neural-based and experience-based developmental models. The importance of identifying and including only the central neural components has been discussed in some detail above, in the section titled "Neural plausibility versus simplicity." A model that appropriately balances plausibility and simplicity includes the neural components that are hypothesized to cause the developmental change, and it is not cluttered with peripheral neural components. For example, Mareschal et al. (1999) hypothesized that changes in infants' behavior in object permanence tasks are based on the development of dorsal and ventral stream integration. Accordingly, Mareschal et al. instantiated those streams in the network. In contrast, Munakata et al. (1997) did not attribute the developmental change to the two streams, so the network did not include them. It is unlikely that Munakata et al. would doubt the existence of that neural division; rather, it was not central to their explanation of development and therefore was not included in the model structure.

Similarly, central and peripheral components must be identified in experience-based models, so that only the experiences that are central to the mechanism of change are used in training. For example, Franz and Triesch (2010) hypothesized that visual experience with linear object motion and occlusion contributes to the development of object unity perception, so the network was trained on these events. Although there are numerous other events that infants see within the first few months of life (e.g., causal events, nonlinear motion), these were not included in the training because they were not hypothesized to be part of the mechanism of change. This allowed for the cleanest demonstration of this mechanism: Older networks received more experience with the set of events and displayed different behavior than did younger networks; therefore, those experiences must have been sufficient to cause the developmental change.

Taking the neural- and experience-based approaches together, it becomes clear that one commonality between the two is the need for a clear theory of developmental change prior to model construction to guide the identification of the central and peripheral components of a given behavior. Crucially, the identification of some components as peripheral does not indicate that they are not present in the behavior. Rather, they are not causally relevant to the mechanism of change. Different groups of researchers may develop different underlying theories for this mechanism, which results in different model structures or sets of training events, as evidenced by the example of Mareschal et al.'s (1999) and Munakata et al.'s (1997) models of object permanence studies. Such alternative theoretical and network approaches to the same behavioral phenomenon can significantly increase the understanding of that phenomenon through the development of novel testable predictions. Furthermore, proposing a clear theory of development and using it to explicitly identify the central and peripheral components can address some of the criticisms leveled against connectionism. For example, the criticism that connectionist terminology may be off-putting to nonexperts (Klahr, 2004) may be in part due to the fact that this terminology is not always grounded in developmental theory. This makes it difficult for behavioral researchers to understand the role of model components in generating behavior. If modelers explicitly identify the central and peripheral components of their network based on a theory of development, researchers who are not experts in modeling may find simulations more accessible.

Parameter Specification

Another challenge that arises in models that implement neural development and in those that implement changes in experience relates to the various parameter settings of the model. A common argument against the usefulness of connectionist modeling and computational modeling more generally is that any behavior can be simulated with sufficient exploitation of free parameters (Roberts & Pashler, 2000). In other words, it has been claimed that any pattern of data can be modeled if free parameters are sufficiently tweaked. A typical connectionist simulation involves the specification of multiple parameters, such as the number of input, hidden, and output units; initial weights; learning rate; and structure of the input, to name just a few. These parameters can instantiate aspects of neural processing or aspects of experience. Although a few modelers have manipulated various model parameters to assess their role in simulating infants' behavior (e.g., Christiansen et al., 1998; Quinn & Johnson, 1997; Rakison & Lupyan, 2008), this kind of parameter checking is not performed systematically. This may lead researchers to be skeptical about the role of these parameters in network performance.

Parameters such as the network architecture, learning rate, and weight decay determine how the network processes information and can be particularly important in models that relate behavioral changes to neural development. Some developmental simulations have manipulated these parameters to demonstrate their effect on the network's performance. For example, Quinn and Johnson (1997, 2000) manipulated the number of hidden units in a network that modeled global- and basic-level categorization. Regardless of the number of hidden units, the network consistently learned global categories first, although when more hidden units were used, basic categories emerged with less additional training. Similarly, Rogers and McClelland (2004) demonstrated that the way that the input units coded the information—either in a localist manner, in which the activity of a single unit coded an object or a property, or in a distributed manner, in which the activity of multiple units coded an object or a property—did not significantly impact the basic phenomena demonstrated by their simulations. Thus, these manipulations show that performance often will remain qualitatively constant within a range of parameter values that specify aspects of processing.

Similarly, the structure of the training set is a crucial parameter in networks that employ an experience-based approach to instantiating developmental change. A number of modelers have varied the training set to demonstrate what features of the input are necessary to match infant performance. Such demonstrations clarify the central components of the training set that are implicated in the mechanism of change. For example, Christiansen et al. (1998) simulated word segmentation development with different types of input cues: phonology, utterance boundaries, and stress. They demonstrated that segmentation was best when all three cues were provided to the network. Similarly, Mayor and Plunkett (2010) showed that networks that received less joint attention experience, or fewer copresentations of objects and their labels, acquired fewer words throughout training. These manipulations show, not surprisingly, that the nature of the training set can impact the model's ability to simulate infant data. Therefore, it is important to consider whether the model's experiences match the infant's experiences.

Despite the impact of parameter values on model performance, criticisms of connectionist networks that are based on single parameter values are not sound, for two reasons. The first reason is that parameter values work in concert to determine the performance of the network. A single parameter value has no meaning; rather, the way the parameter interacts with others determines the model's performance. For example, the number of hidden units must be balanced in relation to the amount and the complexity of the training stimuli (Mareschal et al., 2000). Too many hidden units will cause each unit to respond to a single training item, which will prevent commonalities between items from being encoded, thereby reducing generalization. In contrast, too few hidden units will not allow the network to generate differentiable representations of all of the training items. This suggests that it is not the single parameter of hidden unit number that is important for the modeler to manipulate but instead how that parameter interacts with the type of input provided to the network. Another example of joint parameter effects on network behavior can be found in the work of Rohde and Plaut (1999). They showed that effective learning in a model of language acquisition relied on sufficiently large initial random weights, which allowed for the input activations to be discriminable and for large error signals to be propagated back through the network's layers. When initial weights were too small, the network required many more training trials to achieve the same level of performance. This manipulation shows that the model's success does not rely on a single parameter of initial weights. Instead, that parameter interacts with the amount of training, which must be sufficiently large given the initial starting weights. Thus, the criticism that tweaking of single parameters can result in any pattern of performance is unfounded. In fact, achieving particular performance in a model is a challenging task, because it requires an understanding of the way in which the parameters interact to constrain the network's behavior.

The second reason that networks should not be overly criticized for parameter values is that the same problem is present in behavioral research. Behavioral researchers make a number of decisions during experimental design, such as which stimuli to use, the length of presentation, and the manner of presentation. Decisions typically are made to provide the infant with the most favorable environment to display the effect of interest. Thus, the observed behavior of the infant may be due to the specific settings of the experiment. For example, Oakes and Ribar (2005) have shown that the manner of stimulus presentation can influence infants' ability to form categories. They demonstrated that 4-month-olds formed separate categories of cats and dogs only when the training examples were presented in pairs and not when the examples were presented one at a time. In contrast, 6-month-olds could form categories in both presentation modes. Thus, depending on the parameters of stimuli presentation, different conclusions could be drawn about 4-month-olds' ability to categorize. These results show that just as some simulation results may rely on the specific parameters of the model, behavioral results may also rely on the specific parameters of the experimental procedure that generate a favorable environment to perform a particular task. The key is to determine the degree to which the infant's or the network's behavior can be generalized to other circumstances. Results should generalize to some range of conceptual replications, but it would be unreasonable to expect them to generalize to all experimental contexts. For example, classic causal perception study results (e.g., Cohen & Oakes, 1993; Leslie & Keeble, 1987) should be found for a range of objects that interact in a causal manner. However, the speed at which the objects move could affect the results, because at some point the events may be too fast for infants to process. Similarly, a network's performance should be consistent within a range of parameter values. However, it would be unreasonable to expect a network to maintain the same performance for every single parameter value that can be imagined, particularly if parameter values are varied individually and their interdependence is ignored.

Insufficient Research

Another common challenge across the two approaches is the lack of neural or behavioral research on the target topic, which can make it difficult for modelers to implement the correct neural structures or to design the appropriate training set. With respect to neural processing, although significant progress has been made, our understanding of the infant brain is far from complete. The use of near-infrared spectroscopy (NIRS) holds promise in illuminating infants' information processing. However, this technique is still in its initial stages in its use with infants, and improvement is needed before it can provide definitive and useful information about infants' brains (Aslin, 2012).

With respect to children's experiences, as we have discussed above, CHILDES (MacWhinney, 2000) provides an extensive database of children's linguistic exposure. However, the database does not provide information about the aspects of the environment to which children pay attention and encode. For example, there is insufficient information about the auditory features that are extracted by infants, which creates a challenge to model experiments that provide auditory input, such as Marcus et al.'s (1999) rule-learning experiment. As a result, different modelers adopted a variety of ways to convert Marcus et al.'s behavioral stimuli into network training events: Shultz and Bale (2001) coded them according to a single sonority value; Christiansen and colleagues (Christiansen et al., 2000; Christiansen & Curtin, 1999) coded eight phonological features; and Sirois et al. (2000) used an arbitrary coding scheme unrelated to auditory processing. These various approaches make it difficult to determine which, if any, coding scheme is the best approximation of infants' auditory processing. However, if each approach could generate different predictions, further behavioral testing of these predictions could be helpful for selecting the most appropriate way of coding the auditory input. This illustrates a way in which modeling can spur further theoretically relevant behavioral research.

More critically, as we discussed above, there is no extensive database for children's visual experiences. Consequently, modelers may adopt somewhat arbitrary assumptions about the amount and distribution of infants' experiences. For example, Lupyan and Rakison (2006) simulated infants' learning about animacy and causality by exposing a network to a series of causal and noncausal events. Of the causal events, 75% involved one animate and one inanimate object and 25% involved two animate objects; of the noncausal events, 50% involved one animate and one inanimate object and 50% involved two animate objects. This distribution of the different types of causal and noncausal events was based on conjectures about infants' experiences, because no empirical research was available. However, infants' experiences may be quite different from this surmised distribution. For example, it is conceivable that infants experience some causal events that involve two inanimate objects, such as a rolling ball that knocks over another toy. A model based on inferences about visual experience rather than one based on concrete knowledge of experience may output the same behavior as infants for different reasons, because infants may rely on an entirely different set of experiences.

Two points are worth making that provide hope for surmounting the challenge of inadequate research on early visual experiences. First, instantiations of visual experience in connectionist networks can be treated as explicit predictions that can be tested behaviorally. That is, if a model is trained on a particular set of visual events and successfully replicates infants' behavior, a prediction can be made that these visual events are what infants see, attend to, and encode. This prediction can, in turn, spur behavioral research that examines infants' attention and learning in real environments. Second, progress has recently been made in recording and analyzing infants' early visual experiences through the use of head-mounted video cameras and head-mounted eye trackers. For example, Cicchino et al. (2011) analyzed data from a head-mounted camera and found that infants see more instances of causality than of self-propulsion. Similarly, Franchak et al. (2010) used a head-mounted eye tracker to find that infants rarely look at their mothers' faces during interactions due to physical constraints on their perspective. Therefore, it is apparent that there is an initiative to catalog infants' visual experiences as has already been done for their linguistic experiences.

The examples presented above demonstrate that insufficient behavioral research provides a significant challenge to generating sound models of infants' behavior, which can be overcome through a symbiotic relationship among modeling, behavioral, and neural research. Modeling work relies on empirical research to support the theory about the mechanism of developmental change. In turn, modeling can be a benefit to empirical research, because it can identify areas in which empirical research is lacking and which need more data before an effective theory of developmental change can be developed.

Model Comparison

A final issue that arises in relation to connectionism is the comparison of multiple models of the same phenomenon (e.g., the numerous models of language acquisition, object permanence, or learning of correlated attributes). The different connectionist models of the same topic all appear on the surface to be remarkably successful at replicating the behavioral data. How does one compare them to determine their contribution to our understanding of development? In our view, model comparison involves two steps: determining whether the models are different at the level of implementation or at the level of theory and comparing their strengths and weaknesses. Use of this approach to model comparison can be especially beneficial for the understanding of behavioral data and for the advancement of developmental theory.

When multiple models of the same developmental phenomenon are encountered, the first step is to address the level at which they differ: implementation or theory. Models that differ in implementation agree on the theory behind the developmental change, but this theory is instantiated in the models in different ways. In other words, the causal mechanism that is thought to drive the developmental change is the same across the models. For example, models of the classic Younger and Cohen (1986) study by Gureckis and Love (2004); Shultz and Cohen (2004), and Westermann and Mareschal (2004) all base the development of infants' ability to learn to attribute correlations on the development of visual perception. Thus, the models instantiate the same causal mechanism that relates visual perception to correlational learning. However, the models differ in their implementation of the changes in visual perception. Gureckis and Love modeled improvements in visual perception through decreased noise, such that stimuli appeared more distinct over development. Shultz and Cohen raised the degree to which the internal representation had to match the presented stimulus, so that older networks perceived more details of the stimuli. Finally, Westermann and Mareschal shrunk the receptive fields of neurons in the visual system, which allowed the neurons to be more fine-tuned to specific features of each stimulus. Thus, the three models agree on the mechanism behind the emergence of the ability to learn correlated attributes but build this mechanism in different ways.

Models can also differ at the level of theory such that the mechanism that is hypothesized to support development is different between the models. For example, as discussed previously, both Mareschal et al. (1999) and Munakata et al. (1997) modeled the development of object permanence in infants, but the two simulations differed with respect to the theory of why the developmental change occurred. Mareschal et al. adopted the theory that object permanence develops when object identity and location information are processed in an integrated manner. In contrast, Munakata et al. attributed the change to infants' developing ability to maintain representations in memory. Therefore, these models instantiate different mechanisms to drive development: either integration of two neural processing streams or the strengthening of individual connections. Assessing the level at which models differ is a critical first step in model comparison because it influences the conclusions that can be drawn from the models. In particular, if models differ in terms of theory and further model comparison indicates the superiority of one model over another, this process of model comparison could be helpful in disambiguating between different theories of developmental change.

Once the level of difference has been determined, the second step is to compare the models along several aspects: model assumptions, extent of the modeled behavioral data, and novel predictions. To demonstrate these comparisons, we use two models of Onishi and Baillargeon's (2005) theory of mind tasks: Berthiaume, Shultz, and Onishi (2013) and Van Overwalle (2010). In the behavioral experiment, 15-month-old infants were habituated to an event sequence in which an observer hid an object in one of two boxes; then, they either saw that the object moved to the other box or were absent when this occurred. In the test phase, infants were tested with different events that assessed their expectations about where the person should search for the object based on whether the person was present or absent when the object moved.

First, we compare the extent to which the models' assumptions are backed by neural or behavioral evidence; if such evidence is absent, there is reason to suspect that the mechanism by which infants' behavior is simulated is incorrect. The Berthiaume et al. (2013) and Van Overwalle (2010) models employ different assumptions regarding when learning occurs during an experiment. In the former, learning occurred only during the general experience prior to the experiment, and no learning during an experimental habituation phase was simulated. In the latter, learning occurred throughout all stages, including the test phase. It would be difficult to argue that learning does not occur during an experiment, because the habituation procedure implies that the child's attention decreases as a result of learning. Thus, based on this assumption, Van Overwalle's model may be better grounded than Berthiaume et al.'s.

Second, we compare the extent to which the models yield results that match the observed behavioral patterns. In the case of the Berthiaume et al. (2013) model, there is a match between the model data and the behavioral results with respect to the overall pattern but not the magnitudes of the effects. In particular, infants show about equal differences in looking to correct and incorrect reaches regardless of whether the person's beliefs are true or false. In contrast, the model displays a more robust true belief than false belief effect: The difference in network error (a proxy for looking time) between correct and incorrect reaches is much greater when the person's beliefs are true than when they are false. Van Overwalle's model (2010) of the same behavioral data provides a closer match of the overall pattern and the individual difference magnitudes. Berthiaume et al. do not comment on the discrepancy in magnitude between their data and the infant data. Without a principled reason for this magnitude difference, this imperfect match puts the model at a disadvantage compared to the Van Overwalle model.

Finally, we compare the predictions, if any, that each model makes. The predictions provide a way to falsify the model: If they are not supported by further behavioral (or neural) research, this suggests that revision of the model is needed. Van Overwalle's (2010) model learns the associations between the object, the hiding locations, and the observer. It predicts that the saliency of the observer can affect performance, such that if the observer is not particularly salient, the observer's beliefs will be overridden by the visual information about the actual hiding location of the object. In contrast, the Berthiaume et al. (2013) model relies on the infant's real-world experience that people tend to have more true than false beliefs (i.e., people's beliefs tend to align with reality more often than not). If false beliefs are rare, according to the model, younger infants initially should base their expectations on the state of the world rather than on the observer's beliefs and should fail Onishi and Baillargeon's (2005) task at a younger age than that tested in the experiment. Although the two models do not offer predictions that are in opposition to each other, they provide new avenues for assessing which model, if any, appropriately simulates infant behavior. To our knowledge, Van Overwalle's prediction has not been pursued behaviorally. Berthiaume et al.'s prediction that younger infants should fail the task is inconsistent with behavioral evidence that 13-month-olds succeed in the task (Surian, Caldi, & Sperber, 2007); however, it is conceivable that it could be supported if infants younger than 13 months were tested. Based on a more founded assumption about learning and a closer fit to the behavioral data, it appears that the Van Overwalle model has an advantage over the Berthiaume et al. model. However, it remains to be seen whether the critical prediction of the model would be supported by behavioral data. Furthermore, our comparison of the assumptions and data fit was limited in the interest of providing a brief example that illustrates the main points of the two. Prior to definitively favoring one model over another, a more rigorous analysis of these aspects of models must be conducted.

Comparing multiple models of the same phenomenon is challenging yet necessary for the understanding of the behavioral data and for the advancement of theory. To be clear, we do not want to imply that there is always a unilaterally correct model. Connectionist modeling should be viewed as a continuous exploration of the various components that are hypothesized to contribute to a particular behavior (McClelland, 2009). With respect to model comparison, this means that in all likelihood, all competing models of the same phenomenon have some correct and incorrect components. A comparison of these can highlight the areas where each model is most successful and can spur further modeling efforts that take advantage of these successes.

Summary

Connectionist models of development typically encounter several challenges: the need to balance central and peripheral components of the behavior, the role of specific simulation parameters, inadequate empirical research to support a simulation, and competing theoretical accounts. Before a developmental change can be modeled, the modeler must specify the theory of change to be instantiated in the network. Components that are central to this theory—those that are causally related to the change—should be necessary and sufficient to simulate the change in behavior and therefore be implemented in the model. Peripheral components are typically omitted, because they may obscure the mechanism by which the change occurs in the model. Specifying the theory of change and its central and peripheral components can be particularly challenging if there is insufficient behavioral research on the topic being modeled. In such cases, models can be beneficial for behavioral researchers because they can illuminate areas in need of further study to adequately formulate a theory of developmental change. Although models often are criticized for exploiting too many free parameters, these criticisms ignore the fact that behavioral results rarely generalize to all possible experimental contexts, thereby illustrating a seemingly similar problem. The critical point in relation to both behavioral work and modeling work is that some range of generalization should be expected, but expecting full generalization is unreasonable. Finally, researchers are sometimes faced with multiple models of the same phenomenon, which can make it difficult to determine which account is most accurate. In our view, it is likely that every account possesses correct and incorrect components. However, by comparing models on their assumptions, data fit, and predictions, it is possible to identify areas of success and failure, which is beneficial for the theoretical understanding of behavior and for further modeling efforts.

Conclusion

Developmental researchers primarily ask questions related to what develops, when it develops, and how it develops. Connectionist models can be helpful in answering these questions and are a particularly useful resource in studying development within the first 2 years of life due to the limitations that the participants' young age imposes on behavioral research.

Two main approaches have been used to simulate behavioral changes in a connectionist network: those based on neural developments and those based on experiential changes. It should be noted that although this review has presented the two approaches in separate sections, the intent has not been to suggest that only one or the other operates in theories and models of development. Rather, this was done to clarify the different approaches to modeling and to make better contact with the perspectives adopted by many developmental researchers. The first approach adopts a theoretical stance that behavioral changes are a product of neural changes. Accordingly, models that employ this approach often instantiate developmental changes in perception, integrated processing, memory, and neural plasticity to simulate changes in behavior. Although many of these changes may be occurring simultaneously in infancy, modelers often include only those changes that are implicated in the mechanism of development to increase transparency. As a result, many modelers opt for administering changes in neural processing by hand instead of having the change emerge naturally in the system, because the ultimate interest is in the mechanism that links neural and behavioral changes, not the mechanism by which the neural development itself occurs. The second approach that has been used adopts the perspective that behavioral changes are due to either a quantitative or a qualitative change in experience as infants get older. Development is simulated via additional training epochs or a change in the training events. The importance of the training set creates a pressing need for more accurate empirical data on infants' experiences. Thus far, many modelers have relied on assumptions about what infants may experience because the relevant data have not been available, particularly in the domain of visual experience. Similarly, modelers employing a neural-based approach often face inadequate research on neural development throughout infancy. Furthermore, existing research is often conducted on nonhuman primates or other animals, which can raise questions about the degree to which the results can be employed in models of human cognition.

A central challenge to both approaches has been the balance between creating a model that includes what is known currently about development and one that is simple enough to be useful to the field. This challenge has been met by using a cutoff criterion and including only those components that are directly involved in the mechanism of behavioral change. This approach ensures that models based on changes in neural processing do not incorporate all aspects of infant neural development but only those that are hypothesized to cause the behavioral change. Similarly, it ensures that models based on experience are not exposed to all daily events and objects that infants may encounter but only those that directly impact the change in behavior. Following this cutoff criterion also would be beneficial for making models more generally understandable. Connectionist networks have been criticized for being opaque to nonexperts (Klahr, 2004). However, if all model components are clearly identified and grounded in developmental theory, models would become intelligible to a wider audience.

Connectionist models have thus far provided numerous contributions to the construction and revision of developmental theories. For example, connectionist models have provided coherent accounts for the presence of critical periods in development (McClelland et al., 1999; Zevin & Seidenberg, 2002) and the trade-offs that emerge in the specialization of different brain areas (McClelland et al., 1995; O'Reilly & Munakata, 2000). Such contributions to theory building have primarily emerged from novel predictions made by connectionist models, their use in explicitly testing mechanisms of change, and the unification of disparate behavioral results in a single theoretical framework based on principles of associative learning.

However, behavioral researchers have not always made use of these potential benefits of connectionist models. For example, it took several years for an explicit prediction made by Munakata's (1998) model of the A-not-B task to be tested behaviorally (Clearfield et al., 2006), despite the prediction's potential benefit for the theoretical understanding of memory and perseveration. In our view, this lack of interchange between modelers and behavioral researchers is in part due to the hesitance of the latter toward modeling. For example, behavioral researchers have been critical of the starting states of connectionist networks (Oakes et al., 2009). If models are viewed as adopting arbitrary starting states that are not tightly linked with behavioral work, it is not unreasonable for researchers to be skeptical about the model's usefulness. However, as connectionist networks become more closely linked to developmental theories, some of these concerns may be alleviated. In particular, unifying frameworks that use similar structures to model a range of findings provide a consistent set of starting and processing components, which can suggest a stable correspondence with behavioral data. This may make it more likely that behavioral researchers will make use of the insights offered by connectionist networks.

In addition to the contributions that have already been made by connectionism, what can we expect from these models in the future? We anticipate that over time models will come to have a better correspondence to infants and that more models will be developmental models. The ability of connectionist networks to more closely approximate infants will emerge from two sources. First, computational power will improve, enabling network architectures to be more similar to the brain and allowing for the use of larger training sets that are more representative of real experiences (McClelland, 2009). That is not to say that the principle of balancing simplicity with neural plausibility will be abandoned; modelers will still strive to include only the central components implicated in the mechanism of change. Rather, the increase in computational power will allow for more detail in the specification of this mechanism. For example, it may be possible to simulate behavior starting at the level of interaction between individual neurons rather than neural populations. Second, more research will be conducted that will specify in greater detail the environments of infants and children and will provide data on what events are commonly experienced, perceived, and encoded. Such research will require modelers to make fewer unfounded assumptions about infants' experiences. This will work in concert with increased computational power, such that developmental connectionist models will become even better approximations of children and allow for more precise predictions and theory testing.

We also anticipate that more simulations of infant behavioral data will be conducted from a developmental perspective. The current review focused only on models of development. Consequently, it did not address the numerous nondevelopmental models that simulate infant behavior at a single point in time and do not explore its change over time, typically because the model simulates behavioral data from a study that addressed only a single age. For example, models of infants' categorization of cats and dogs (e.g., French et al., 2004; Mareschal et al., 2000, 2002) explored only the effects of various experimental manipulations at a single age, as did the behavioral studies on which the models were based. However, there is no reason why models should study a single age group. They can be extended to make novel developmental predictions and provide added value to the underlying behavioral work that did not examine developmental change. We anticipate that more modelers will take advantage of this, because it is a cost-effective way of motivating new developmental experiments. Furthermore, we expect that the developmental trajectory for a given behavior will be extended beyond infancy to include the entire spectrum of change from infancy to adulthood. This would allow researchers to understand if the changes that are observed over a few months in infancy and those that are seen between infancy and adulthood result from same the mechanism. If the same mechanism underlies these changes, the same simulation should produce infant and adult behavior. In contrast, if another mechanism is needed to bring infants' performance to adult competence levels, the same model could not produce infant and adult behavior without significant modifications. One example of this in the existing modeling literature is Gureckis and Love's (2004) model of infants' category learning, which adapted a network architecture that had been used previously to model adult category learning. The model was successful in simulating infants' behavior, which suggested that a common mechanism may operate during category learning throughout the life span. This type of modeling that uses the same mechanism to model child and adult behavior has been an active area of research in other modeling approaches as well (e.g., probabilistic modeling: Kemp & Tenenbaum, 2008; dynamic field theory: Perone et al., 2011).

The focus in this article has been on connectionist models of development within the first 2 years of life. Despite this narrowly defined topic, many of the arguments presented in this article can be of interest to a more general audience. With respect to the two approaches of modeling development, they are not unique to early development and can be applied to model age-related changes at any time point. Neural-based approaches to modeling development that simulate neurogenesis, synaptogenesis, and synaptic pruning can be employed throughout the life span, because these neural changes occur past the first 2 years of life (Casey et al., 2005; Huttenlocher & Dabholkar, 1997; Stiles & Jernigan, 2010). Furthermore, the principles behind modeling a process such as neurogenesis can be applied to model the reverse process of neural death that may occur in the elderly. Similarly, experience-based approaches to modeling development are not age restricted. Studies of expertise in adulthood often focus on the role of experience (e.g., Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002); these studies could benefit from insights generated by experience-based models of infant behavior.

Likewise, the common challenges reviewed in this article can arise in nondevelopmental models; therefore, some of the solutions presented can be of use to a broad range of behavioral researchers and modelers, particularly those interested in mechanistic explanations. For example, the identification of the central and peripheral components of a behavior during network construction must be made for all simulations, regardless of the age that is modeled. Similarly, insufficient research can hamper modeling of behaviors that occur throughout the life span. This may lead modelers to develop training sets that do not reflect accurately the experiences of human participants. However, such cases can also be opportunities to explore the consequences of the developed training set (McClelland, 2009). That is, if the modeler uses a training set that is based on some conjectures about participants' experience and the model successfully simulates behavioral data, this could suggest that the training set does, in fact, reflect real-world experience.

Most importantly, the contributions of connectionist models to developmental research can be extended to behavioral research more generally. A connectionist model that successfully simulates behavioral data provides a concrete instantiation of the mechanism that underlies that behavior. This instantiation can be used to generate novel testable predictions, which can be a time- and cost-efficient way to explore new research topics that can provide significant contributions to the theoretical understanding of a target behavioral phenomenon.

In sum, our goals in this article were twofold: first, to demonstrate the added value of connectionist modeling to behavioral research in infant development, and second, to provide concrete examples of how connections networks are used to model development. In our view, the benefit of connectionist modeling can be enhanced significantly if modelers and developmental scientists work in concert on the key questions in early development. Modelers can bring to the table the novel predictions that models can generate and can help to refine existing theoretical accounts of development. However, their ability to do so relies on developmental scientists' exploration of relevant behavioral questions, such as those related to early experience. If both sides meet these expected and practical contributions, developmental science stands to make strong advances in the theories of early development. Furthermore, this symbiotic relationship between behavioral scientists and modelers need not be limited to the study of infant development and can be beneficial for the study of behavior throughout the life span.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9,* 147–169. doi:10.1207/s15516709cog0901_7

Adolph, K. E. (1997). Learning in the development of infant locomotion. *Monographs of the Society for Research in Child Development, 62*(3), 1–140. doi:10.2307/1166199

Aggelopoulos, N. C., & Rolls, E. T. (2005). Scene perception: Inferior temporal cortex neurons encode the positions of different objects in the scene. *European Journal of Neuroscience, 22,* 2903–2916. doi:10.1111/j.1460-9568.2005.04487.x

Aslin, R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science, 86,* 561–565. doi:10.1097/OPX.0b013e3181a76e96

Aslin, R. N. (2012). Questioning the questions that have been asked about the infant brain using near-infrared spectroscopy. *Cognitive Neuropsychology, 29,* 7–33. doi:10.1080/02643294.2012.654773

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition, 23,* 21–41. doi:10.1016/0010-0277(86)90052-1

Barto, A. G. (1995). Reinforcement learning. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 804–813). Cambridge, MA: MIT Press.

Berthiaume, V. G., Shultz, T. R., & Onishi, K. H. (2013). A constructivist connectionist model of transitions in false-belief tasks. *Cognition, 126,* 441–458. doi:10.1016/j.cognition.2012.11.005

Berthier, N. E., Rosenstein, M. T., & Barto, A. G. (2005). Approximate optimal control as a model for motor learning. *Psychological Review, 112,* 329–346. doi:10.1037/0033-295X.112.2.329

Braddick, O., & Atkinson, J. (2007). Development of brain mechanisms for visual global processing and object segmentation. *Progress in Brain Research, 164,* 151–168. doi:10.1016/S0079-6123(07)64008-4

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research, 51,* 1588–1609. doi:10.1016/j.visres.2011.02.018

Braddick, O., Birtles, D., Wattam-Bell, J., & Atkinson, J. (2005). Motion- and orientation-specific cortical responses in infancy. *Vision Research, 45,* 3169–3179. doi:10.1016/j.visres.2005.07.021

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage, 5,* 49–62. doi:10.1006/nimg.1996.0247

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition of child speech. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 11–53). New York, NY: Wiley.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1,* 149–219. doi:10.1207/S15327078IN0102_1

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences, 9,* 104–110. doi:10.1016/j.tics.2005.01.011

Chaput, H. H., & Cohen, L. B. (2001). A model of infant causal perception and development. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 182–187). Mahwah, NJ: Erlbaum.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13,* 221–268. doi:10.1080/016909698386528

Christiansen, M. H., Conway, C. M., & Curtin, S. L. (2000). A connectionist single-mechanism account of rule-like behavior in infancy. In L. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 83–88). Mahwah, NJ: Erlbaum.

Christiansen, M. H., & Curtin, S. L. (1999). The power of statistical learning: No need for algebraic rules. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 114–119). Mahwah, NJ: Erlbaum.

Cicchino, J. B., Aslin, R. N., & Rakison, D. H. (2011). Correspondences between what infants see and know about causal and self-propelled motion. *Cognition, 118,* 171–192. doi:10.1016/j.cognition.2010.11.005

Clearfield, M. W., Diedrich, F. J., Smith, L. B., & Thelen, E. (2006). Young infants reach correctly in A-not-B tasks: On the development of stability and perseveration. *Infant Behavior & Development, 29,* 435–444. doi:10.1016/j.infbeh.2006.03.001

Cohen, L. B. (1973). A two-process model of infant visual attention. *Merrill-Palmer Quarterly, 19,* 157–180.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development, 17,* 1323–1343. doi:10.1016/S0885-2014(02)00124-7

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. *Developmental Psychology, 29,* 421–433. doi:10.1037/0012-1649.29.3.421

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role of associative learning. *Psychological Review, 112,* 347–382. doi:10.1037/0033-295X.112.2.347

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28,* 125–127. doi:10.3758/BF03203646

Dehaene, S. (1998). The development of delayed response: Parallel distributed processing lacks neural plausibility. *Developmental Science, 1,* 194–196. doi:10.1111/1467-7687.00025

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology, 89,* 3264–3278. doi:10.1152/jn.00358.2002

Dominguez, M., & Jacobs, R. A. (2003). Does visual development aid visual learning? In P. T. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 257–277). New York, NY: Psychology Press.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211. doi:10.1207/s15516709cog1402_1

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71–99. doi:10.1016/0010-0277(93)90058-4

Elman, J. L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences, 9,* 111–117. doi:10.1016/j.tics.2005.01.005

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

Farkas, I., & Li, P. (2002). Modeling the development of lexicon with a growing self-organizing map. In H. J. Caulfield (Ed.), *Proceedings of the Sixth Joint Conference on Information Sciences* (pp. 553–556). Durham, NC: Association for Intelligent Machinery.

Field, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., . . . Petersen, S. E. (2009). Functional brain networks development from a "local to distributed" organization. *PLoS Computational Biology, 5,* e1000381. doi:10.1371/journal.pcbi.1000381

Foster, D. J., & Wilson, M. A. (2006, March 30). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature, 440,* 680–683. doi:10.1038/nature04587

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010). Head-mounted eye-tracking of infants' natural interactions: A new method. Proceedings of the 2010 Symposium on Eye Tracking Research & Applications (pp. 21–27). Austin, TX. Association for Computing Machinery.

Franz, A., & Triesch, J. (2010). A unified computational model of the development of object unity, object permanence, and occluded object trajectory perception. *Infant Behavior & Development, 33,* 635–653. doi:10.1016/j.infbeh.2010.07.018

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General, 133,* 382–397. doi:10.1037/0096-3445.133.3.382

Gasser, M., & Colunga, E. (2003). Pattern learning in infants and neural networks. In P. T. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 233–255). New York, NY: Psychology Press.

Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes, 13,* 269–306. doi:10.1080/016909698386537

Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In F. Plum & V. Mountcastle (Eds.), *Handbook of physiology* (pp. 373–417). Bethesda, MD: American Physiological Society.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25. doi:10.1016/0166-2236(92)90344-8

Gureckis, T. M., & Love, B. C. (2004). Common mechanisms in infant and adult category learning. *Infancy, 5,* 173–198. doi:10.1207/s15327078in0502_4

Halpern, L. F., MacLean, W. E., & Baumeister, A. A. (1995). Infant sleep-wake characteristics: Relation to neurological status and the prediction of developmental outcome. *Developmental Review, 15,* 255–291. doi:10.1006/drev.1995.1010

Hurley, K. B., Kovack-Lesh, K. A., & Oakes, L. M. (2010). The influence of pets on infants' processing of cat and dog images. *Infant Behavior & Development, 33,* 619–628. doi:10.1016/j.infbeh.2010.07.015

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia, 28,* 517–527. doi:10.1016/0028-3932(90)90031-I

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology, 387,* 167–178. doi:10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNE1>3.0.CO;2-Z

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word learning: Universal ontology and linguistic influence. *Cognition, 62,* 169–200. doi:10.1016/S0010-0277(96)00784-6

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review, 17,* 662–672. doi:10.3758/PBR.17.5.662

Johnson, M. H. (2000). Functional brain development in infants: Elements of an interactive specialization framework. *Child Development, 71,* 75–81. doi:10.1111/1467-8624.00120

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience, 2,* 475–483. doi:10.1038/35081509

Johnson, S. P., & Aslin, R. N. (1996). Perception of object unity in young infants: The roles of motion, depth, and orientation. *Cognitive Development, 11,* 161–180. doi:10.1016/S0885-2014(96)90001-5

Káldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: Integrating "what" and "where" information. *Developmental Science, 6,* 360–373. doi:10.1111/1467-7687.00290

Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition, 97,* 153–177. doi:10.1016/j.cognition.2004.09.009

Káldy, Z., & Sigala, N. (2004). The neural mechanisms of object working memory: What is where in the infant brain? *Neuroscience and Biobehavioral Reviews, 28,* 113–121. doi:10.1016/j.neubiorev.2004.01.002

Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88,* 197–227. doi:10.1037/0033-295X.88.3.197

Kellman, P. J., & Arterberry, M. E. (2006). Perceptual development. In W. Damon, D. Kuhn, & R. Siegler (Eds.), *The handbook of child psychology: Cognition, perception, and language* (6th ed., pp. 109–160). New York, NY: Wiley.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15,* 483–524. doi:10.1016/0010-0285(83)90017-8

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language, 48,* 207–232. doi:10.1016/S0749-596X(02)00512-0

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, USA, 105,* 10687–10692. doi:10.1073/pnas.0802631105

Klahr, D. (2004). Commentary: New kids on the connectionist modeling block. *Developmental Science, 7,* 165–166. doi:10.1111/j.1467-7687.2004.00334.x

Kohonen, T. (1977). *Associative memory: A system theoretical approach*. New York, NY: Springer. doi:10.1007/978-3-642-96384-1

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43,* 59–69. doi:10.1007/BF00337288

Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorization. *Infancy, 13,* 285–307. doi:10.1080/15250000802189428

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25,* 265–288. doi:10.1016/S0010-0277(87)80006-9

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17,* 1345–1362. doi:10.1016/j.neunet.2004.07.004

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science, 31,* 581–612. doi:10.1080/15326900701399905

Lupyan, G., & Rakison, D. H. (2006). What moves in a mysterious way? A domain-general account of learning about animacy and causality. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 525–530). Mahwah, NJ: Erlbaum.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum. doi:10.1162/coli. 2000. 26. 4. 657

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language, 31,* 883–914. doi:10.1017/S0305000904006336

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology, 23,* 263–298. doi:10.1016/0010-0285(91)90011-C

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition, 46,* 53–85. doi:10.1016/0010-0277(93)90022-N

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999, January 1). Rule learning in seven-month-old infants. *Science, 283,* 77–80. doi:10.1126/science.283.5398.77

Mareschal, D. (2003). Connectionist models of learning and development in infancy. In P. T. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 43–82). New York, NY: Psychology Press.

Mareschal, D. (2010). Computational perspectives on cognitive development. *WIREs Cognitive Science, 1,* 696–708. doi:10.1002/wcs.67

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology, 36,* 635–645. doi:10.1037/0012-1649.36.5.635

Mareschal, D., & Johnson, S. P. (2002). Learning to perceive object unity: A connectionist account. *Developmental Science, 5,* 151–172. doi:10.1111/1467-7687.t01-1-00217

Mareschal, D., Plunkett, K., & Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviors in infancy. *Developmental Science, 2,* 306–317. doi:10.1111/1467-7687.00076

Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cognitive Science, 26,* 377–389. doi:10.1207/s15516709cog2603_8

Maurer, D., & Lewis, T. L. (2001). Visual acuity and spatial contrast sensitivity: Normal development and underlying mechanisms. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 237–251). Cambridge, MA: MIT Press.

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review, 117,* 1–31. doi:10.1037/a0018130

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1,* 11–38. doi:10.1111/j.1756-8765.2008.01003.x

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. doi:10.1037/0033-295X.102.3.419

McClelland, J. L., & Plaut, D. C. (1999). Does generalization in infants learning implicate abstract algebra-like rules? *Trends in Cognitive Sciences, 3,* 166–168. doi:10.1016/S1364-6613(99)01320-0

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. II: Psychological and biological models*. Cambridge, MA: MIT Press.

McClelland, J. L., & Vallabha, G. (2009). Connectionist models of development: Mechanistic dynamical models with emergent dynamical properties. In J. P. Spencer, M. S. C. Thomas, & J. L. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 3–24). New York, NY: Oxford University Press.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation, 24,* 109–165. doi:10.1016/S0079-7421(08)60536-8

Monroe, B. M., & Read, S. J. (2008). A general connectionist model of attitude structure and change: The ACS (attitudes as constraint satisfaction) model. *Psychological Review, 115,* 733–759. doi:10.1037/0033-295X.115.3.733

Morton, J. B., & Munakata, Y. (2005). What's the difference? Contrasting modular neural network approaches to understanding developmental variability. *Journal of Developmental and Behavioral Pediatrics, 26,* 128–139. doi:10.1097/00004703-200504000-00010

Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A PDP model of the A-not-B task. *Developmental Science, 1,* 161–184. doi:10.1111/1467-7687.00021

Munakata, Y. (2006). Information processing approaches to development. In W. Damon & R. Lerner (Series Eds.) and D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2*: *Cognition, perception, and language* (6th ed., pp. 426–463). New York, NY: Wiley.

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science, 6,* 413–429. doi:10.1111/1467-7687.00296

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104,* 686–713. doi:10.1037/0033-295X.104.4.686

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science, 7,* 141–148. doi:10.1111/j.1467-7687.2004.00331.x

Mundy, P., & Van Hecke, A. V. (2008). Neural systems, gaze following and the development of joint attention. In C. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 819–837). New York, NY: Oxford University Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316. doi:10.1037/0033-295X.92.3.289

Oakes, L. M., Newcombe, N. S., & Plumert, J. M. (2009). Are dynamic systems and connectionist approaches an alternative to good old-fashioned cognitive development? In J. P. Spencer, M. S. C. Thomas, & J. L. McCelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 268–284).

New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195300598.003.0013

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7,* 85–98. doi:10.1207/s15327078in0701_7

Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science, 17,* 781–787. doi:10.1111/j.1467-9280.2006.01782.x

Onishi, K. H., & Baillargeon, R. (2005, April 8). Do 15-month-old infants understand false beliefs? *Science, 308,* 255–258. doi:10.1126/science.1107621

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology, 426,* 505–518. doi:10.1002/1096-9861(20001030)426:4<505::AID-CNE1>3.0.CO;2-M

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2011). Complementary learning systems. *Cognitive Science*. Advance online publication. doi:10.1111/j. 1551–6709.2011.01214.x

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

Pascalis, O., de Haan, M., Nelson, C. A., & de Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 249–260. doi:10.1037/0278-7393.24.1.249

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences, USA, 102,* 5297–5300. doi:10.1073/pnas.0406627102

Perone, S., Simmering, V. R., & Spencer, J. P. (2011). Stronger neural dynamics capture changes in infants' visual working memory capacity over development. *Developmental Science, 14,* 1379–1392. doi:10.1111/j.1467-7687.2011.01083.x

Piaget, J. (1952). *The origins of intelligence in children (M. Cook, Trans.)*. New York, NY: Norton. doi:10.1037/11494-000

Plaut, D. C., & Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. *Cognitive Neuropsychology, 28,* 251–275. doi:10.1080/02643294.2011.609812

Plunkett, K., & Marchman, V. A. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21–69. doi:10.1016/0010-0277(93)90057-3

Quartz, S. R. (1999). The constructivist brain. *Trends in Cognitive Sciences, 3,* 48–57. doi:10.1016/S1364-6613(98)01270-4

Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology, 63,* 189–211. doi:10.1006/jecp.1996.0047

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22,* 463–475. doi:10.1068/p220463

Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: A connectionist analysis. *Journal of Experimental Child Psychology, 66,* 236–263. doi:10.1006/jecp.1997.2385

Quinn, P. C., & Johnson, M. H. (2000). Global-before-basic object categorization in connectionist networks and 2-month-old infants. *Infancy, 1,* 31–46. doi:10.1207/S15327078IN0101_04

Rakison, D. H., & Butterworth, G. (1998). Infants' use of parts in early categorization. *Developmental Psychology, 34,* 49–62. doi:10.1037/0012-1649.34.1.49

Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development,* 73(1), vii, 1–130. doi:10.1111/j.1540-5834.2008.00454.x

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367. doi:10.1037/0033-295X.107.2.358

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition, 72,* 67–109. doi:10.1016/S0010-0277(99)00031-1

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized face-like electrophysiological responses. *Psychological Science, 13,* 250–257. doi:10.1111/1467-9280.00446

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 45–76). Cambridge, MA: MIT Press.

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

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press.

Schafer, G., & Mareschal, D. (2001). Modeling infant speech sound discrimination using simple associative networks. *Infancy, 2,* 7–28. doi:10.1207/S15327078IN0201_2

Schlesinger, M., & Casey, P. (2003). Where infants look when impossible things happen: Simulating and testing a gaze-direction model. *Connection Science, 15,* 271–280. doi:10.1080/09540090310001655093

Schlesinger, M., & Young, M. E. (2003). Examining the role of prediction in infants' physical knowledge. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society* (pp. 1047–1052). Boston, MA: Cognitive Science Society.

Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review, 113,* 273–299. doi:10.1037/0033-295X.113.2.273

Scott, L. S., & Monesson, A. (2010). Experience-dependent neural specialization during infancy. *Neuropsychologia, 48,* 1857–1861. doi:10.1016/j.neuropsychologia.2010.02.008

Scott, L. S., Shannon, R. W., & Nelson, C. A. (2006). Neural correlates of human and monkey face processing in 9-month-old infants. *Infancy, 10,* 171–186. doi:10.1207/s15327078in1002_4

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: Data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 25–49). Hove, England: Psychology Press.

Seress, L., Ábrahám, H., Tornóczky, T., & Kosztolányi, G. Y. (2001). Cell formation in the human hippocampal formation from mid-gestation to the late postnatal period. *Neuroscience, 105,* 831–843. doi:10.1016/S0306-4522(01)00156-7

Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Shultz, T. R. (2010). Computational modeling of infant concept learning: The developmental shift from features to correlations. In L. M. Oakes, C. H. Cashon, M. Casasola, & D. H. Rakison (Eds.), *Infant perception and cognition* (pp. 125–152). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195366709.003.0007

Shultz, T. R., & Bale, A. C. (2001). Neural network simulation of infant familiarization to artificial sentences: Rule-like behavior without explicit rules and variables. *Infancy, 2,* 501–536. doi:10.1207/S15327078IN0204_07

Shultz, T. R., & Cohen, L. B. (2004). Modeling age differences in infant category learning. *Infancy, 5,* 153–171. doi:10.1207/s15327078in0502_3

Siegle, G. J., Steinhauer, S. R., & Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *International Journal of Psychophysiology, 52,* 63–76. doi:10.1016/j.ijpsycho.2003.12.010

Sirois, S., Buckingham, D., & Shultz, T. R. (2000). Artificial grammar learning by infants: An auto-associator perspective. *Developmental Science, 3,* 442–456. doi:10.1111/1467-7687.00138

Sirois, S., & Shultz, T. R. (2003). A connectionist perspective on Piagetian development. In P. T. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 13–41). New York, NY: Psychology Press.

Sloman, A. (2008). Putting the pieces together again. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 684–710). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816772.031

Smith, L. B., Gasser, M., & Sandhofer, C. M. (1997). Learning to talk about the properties of objects: A network model of the development of dimensions. *Psychology of Learning and Motivation, 36,* 219–255. doi:10.1016/S0079-7421(08)60285-6

Smith, L. B., & Scheier, C. (1998). Babies have bodies: Why Munakata's net fails to meet its own goals. *Developmental Science, 1,* 202–204. doi:10.1111/1467-7687.00030

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Hillsdale, NJ: Erlbaum.

Spencer, J. P., Thomas, M. S. C., & McClelland, J. L. (2009). *Toward a unified theory of development: Connectionism and dynamic systems theory reconsidered*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195300598.001.0001

Stedron, J. M., Sahni, S. D., & Munakata, Y. (2005). Common mechanisms for working memory and attention: The case of perseveration with visible solutions. *Journal of Cognitive Neuroscience, 17,* 623–631. doi:10.1162/0898929053467622

Stiles, J., & Jernigan, T. L. (2010). The basics of brain development. *Neuropsychology Review, 20,* 327–348. doi:10.1007/s11065-010-9148-4

Sun, R. (2008). Introduction to computational cognitive modeling. In R. Sun (Ed.), *The Cambridge handbook of computational psychology (pp.* 3–20). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816772.003

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18,* 580–586. doi:10.1111/j.1467-9280.2007.01943.x

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,* 1–34. doi:10.1017/S0140525X01003910

Thomas, M. S. C., & McClelland, J. L. (2008). Connectionist models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology (pp.* 23–58). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816772.005

Thomas, M. S. C., McClelland, J. L., Richardson, F. M., Schapiro, A. C., & Baughman, F. (2009). Dynamical and connectionist approaches to development: Toward a future of mutually beneficial co-evolution. In J. P. Spencer, M. S. C. Thomas, & J. L. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 337–353). New York, NY: Oxford University Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical systems. In D. G. Ingle, M. A. Goodale, & R. J. Q. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992, January 24). Information processing in the primate visual system: An integrated systems perspective. *Science, 255,* 419–423. doi:10.1126/science.1734518

Van Overwalle, F. (2010). Infants' teleological and belief inference: A recurrent connectionist approach to their minimal representational and computational requirements. *NeuroImage, 52,* 1095–1108. doi:10.1016/j.neuroimage.2010.05.028

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49–63. doi:10.1016/S0163-6383(84)80022-3

Westermann, G., & Mareschal, D. (2004). From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy, 5,* 131–151. doi:10.1207/s15327078in0502_2

Xie, X., & Seung, H. S. (2003). Equivalence of backpropagation and contrastive Hebbian learning in a layered network. *Neural Computation, 15,* 441–454. doi:10.1162/089976603762552988

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development, 57,* 803–815. doi:10.2307/1130356

Younger, B. A., Hollich, G., & Furrer, S. D. (2004). An emerging consensus: Younger and Cohen revisited. *Infancy, 5,* 209–216. doi:10.1207/s15327078in0502_6

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14,* 165–180. doi:10.1111/j.1467-7687.2010.00958.x

Yurovsky, D., Hidaka, S., Yu, C., & Smith, L. B. (2010). Linking learning to looking: Habituation and association in infant statistical language learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1589–1594). Austin, TX: Cognitive Science Society.

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

Received February 22, 2012
Revision received January 10, 2013
Accepted January 22, 2013 ■