Review TRENDS in Cognitive Sciences Vol.9 No.3 March 2005



# Connectionist models of cognitive development: where next?

**Jeffrey L. Elman**

Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093-0515, USA

**Over the past two decades, connectionist models have generated a lively debate regarding the underlying mechanisms of cognitive development. This debate has in turn led to important empirical research that might not have occurred otherwise. More recently, advances in developmental neuroscience present a new set of challenges for modelers. In this article, I review some of the insights that have come from modeling work, focusing on (1) explanations for the shape of change; (2) new views on how knowledge may be represented; (3) the richness of experience. The article concludes by considering some of the new challenges and opportunities for modeling cognitive development.**

## Introduction

Over the past several decades, computational models have played an increasingly important role in the developmental literature. Among these, connectionist models stand out as offering some of the most exciting (and also controversial) insights into cognitive development. The controversy has for the most part been healthy, and has stimulated new research that has led to important empirical discoveries. More recently, as the modeling work has matured, the relationship between neural-network models and other computational approaches – including Bayesian models, information-theoretic analyses, statistical learning, among others – has helped to reveal the deeper computational principles at work.

So what have we learned from these models? And what phenomena remain to be explored? The goal of the present article is to present a critical – and admittedly, very personal – review of what I believe have been the major advances that have come from attempts to model developmental phenomena (for additional reviews of the modeling literature, see [1,2–4]). My assessment is positive. But I will also argue that important new challenges exist. The work to be done might be tougher, but will provide a greater yield.

## What have we learned?

We have learned a great deal about development from computational modeling, and more than can be covered in this review. I therefore focus on what I see as some of the central issues: (1) explaining the shape of change; (2) new views on how knowledge may be represented; (3) the richness of experience.

### *The shape of change*

The simplest and possibly most natural pattern of development might seem to be one in which, over time, performance increases in a linear fashion. In fact, very few developmental patterns illustrate such linear tendencies. Development usually proceeds in fits and spurts, sometimes interrupted by long periods where little appears to change, and even sometimes by periods where performance temporarily deteriorates.

Noteworthy examples of such nonlinearities abound in the realm of language acquisition, and have played a major role in theorizing about the mechanisms that make language acquisition possible. The special ability of children to learn languages (during a Critical Period) motivated the hypothesis of a specialized neural mechanism, the Language Acquisition Device, that operates only during childhood [5]. The rapid increases in word comprehension, production, and knowledge of grammar that occur in young children during their second year of life have led many theorists to conclude that these bursts arise because something new has appeared in the child.

In non-linguistic domains, the apparent stage-like changes in children's behavior (e.g. with balance-beam problems, in which children are shown a see-saw device with two weights at opposite sides of a fulcrum, and are asked which side, if either, will go down) has encouraged the supposition that children are passing through qualitatively different stages of cognitive development. And most dramatic of all are so-called 'U-shaped' performance curves. Perhaps the best known, and theoretically most consequential, has involved the acquisition of the English past tense. The temporary decline in correct performance, characterized by what looks like over-application of the 'add *-ed*' pattern, has seemed to many to provide dramatic evidence for the psychological reality of rules. Computational models have been proposed that offer alternative interpretations for all of these phenomena.

*The Critical Period and stage-like learning*. McClelland and colleagues [6,7] demonstrated that critical period effects for phoneme perception could be achieved in a neural network, even when the mechanism underlying learning remained unchanged. O'Reilly and Johnson [8] found similar effects in a model of imprinting in chicks. Marchman [9] obtained a similar result in networks that learned the past-tense mapping, and found that the

*Corresponding author:* Elman, J.L. (jelman@ucsd.edu).

Available online 1 February 2005

www.sciencedirect.com 1364-6613/$ - see front matter  doi:10.1016/j.tics.2005.01.005

greater plasticity in 'young' networks allowed them to recover from artificial lesions to a much greater extent than when the lesions occurred later in learning. The implication here is that even when the learning mechanism remains constant, the cumulative effects of learning can result in entrenchment, and this reduces the plasticity of the system [4,10].

A second possibility (not necessarily incompatible with the first) has been proposed by Newport [11] and Elman [12]. In these models, the advantages of a critical period for language learning result – counterintuitively – from processing limitations that occur in early childhood (see also [13,14] for related accounts). In addition to suggesting alternative hypotheses for important developmental phenomena, many of these models have led to empirical research that appears to support their predictions (e.g. [6,7,15,16]).

Stage-like effects in learning the balance-beam problem can be shown to arise naturally in neural networks as a consequence of simple nonlinearities in the response functions of processing units, because small quantitative changes in a learning system can lead to qualitative changes in behavior [17]. Appearance of the latter might not be evidence that a new internal mechanism has appeared on the scene.

However, the actual empirical data regarding the balance-beam problem are more complex than initially described. Children's performance is quite variable, both for an individual child and across ages, and performance can be affected by perceptual variables. Symbolic approaches have difficulty accounting for such effects. A class of models known as 'cascade correlation networks', in which development is modeled by a gradual increase in processing resources, turns out to capture these and other subtleties quite nicely [18].

*U-shaped curves*. Even more striking than nonlinear changes in behavior is when development actually appears to suffer a temporary decline. Probably the best known example of this is the U-shaped performance curve associated with children's learning of the English past tense. Although this has seemed to many to be a prima facie example for the acquisition of a rule, Rumelhart and McClelland [19] were able to show through simulation that U-shaped behaviors might arise in the course of item-based learning. That is, the process of abstraction and generalization need not involve symbolic rules, and non-monotonic changes in performance can be a consequence of the statistical structure of the domain to be learned and the incremental nature of learning [20].

Rumelhart and McClelland's proposal that neural networks provide an alternative to symbolic rules has remained controversial. It has also been enormously productive and has stimulated a large body of research – both computational and behavioral – that has considerably refined our understanding of U-shaped behaviors [20–30]. Indeed, many of the studies that will be discussed in the remainder of this paper were motivated by issues raised in Rumelhart and McClelland's seminal work.

### *New views about how knowledge can be represented*

In symbolic theories, representations are typically discrete in form and all-or-nothing in functionality. This reflects what has been called the 'algebraic mind' [31]. Yet there are many examples, both in development and adult behavior, in which behavior seems to suggest a richer view of what it means to know something. Consider the variability displayed by children in balance-beam and conservation tasks, where behavior often depends crucially on task variables [32,33]. Such behavior is not easily accommodated by symbolic representations and deterministic processing.

Computational models have been particularly useful in suggesting new ways to think about the source and form of knowledge. One example concerns the apparent paradox in which very young children seem to possess knowledge about the existence of hidden objects in one task, yet in another task in which this knowledge must be used more actively, appear no longer to be aware of its existence. Munakata and colleagues explored such behavior both experimentally as well as through computational simulations [34,35]. The simulations led to a mechanistic account for what one might call 'partial knowledge', and to insights regarding a distinction between latent and active representations.

Another major contribution has come from models that have emphasized dynamical systems theory (e.g. [36]). These models downplayed the role of explicit representations, and have provided alternative ways to think about how computation might be accomplished by means of dynamics that operate on non-representational internal states.

### *Just what can be learned from the input?*

One of the most contentious areas in development has surrounded the notion that some knowledge must be innate. Although innate knowledge has been claimed for many domains, it has been developed in its most explicit form in the area of language, from what is known as the 'Poverty of the Stimulus' (PoS) hypothesis [37,38–42]. The basic claim is that the input available to children underdetermines what they will eventually come to know about language. A related argument in favor of the 'algebraic mind' has been that connectionist networks are unable to generalize beyond their immediate experience, suggesting that mental processing is both symbolic and benefits from 'pre-knowledge' [31].

It is clear that something about human biology is essential to learning language (despite impressive demonstrations of non-human primates' communication skills, no non-human comes anywhere close to mastering human language). A crucial question then is *what* it is that is species-specific – is it also domain-specific, specialized knowledge that subserves only language? And is it the ability for symbol processing? Or does language emerge as a consequence of multiple phenotypic modifications that are not in themselves specific to language but which collectively make language possible [2,43–45]? Can language be processed by non-symbolic mechanisms? Models have been especially important in fleshing out such an emergentist position by demonstrating ways in which non-domain specific constraints can interact in ways that yield domain-specific outcomes [46–48].

A more direct response to PoS claims has been to test directly the question of what information is available in the input to children by seeing what a computational model can learn, given input of the sort that a child might be expected to encounter.

*Word segmentation.* One of the very earliest things an infant must learn is to identify the words of her language – not only *what* they are, but *where* they are in the speech stream. Unlike written language, there is no 'white space' in spoken language that delimits words. How then might a child learn words, if even their location is opaque?

Several computation approaches have converged on a similar insight, which is that, at least to a first approximation, sequences of sounds that are highly associated are good candidates to be words. The manner in which this hypothesis is implemented varies [49–51] but the essential idea is that word boundaries are locations where the conditional probability of the next sound, given what has preceded it, is low. A network (or child) that attempts to anticipate what it (she) will hear next will tend to do worse at the onsets of words, and better as more of a word is heard. Error maxima thus constitute likely word boundaries.

*Learning categories*. A related problem, at a somewhat higher level of abstraction, is determining the syntactic and semantic categories of words. Here again, strong claims have been advanced that categories cannot be induced but must be innately given [40]. The claim is that a strategy that relies on distributional information available to a child (e.g. words in the same category tend to have similar distributional properties) will fail given the complexity of real language input. However, several computational models have suggested otherwise [52–55]. Again, these models differ in their details, but share the same insight that a word's privilege of occurrence is a powerful indicator of its category. Importantly, there is a growing empirical literature involving learning of artificial languages by infants and young children that is highly consistent with the type of learning embodied in the computational models (see [56,57] for discussion of this work].

But it is in the realm of syntax that the PoS claim has been most vigorously advanced, and which potentially present the most significant challenges for models that emphasize learning. One problem involves the role of negative evidence regarding ungrammatical sentences. It has been claimed that such information is necessary for children to know when to retract incorrect over-generalizations; but children rarely receive this sort of information. In other words, without such information, children have no way of knowing whether a gap in their experience is accidental, or is systematic (i.e. reflects an ungrammaticality). However, models that follow Bayesian principles of inference (e.g. [58–60]) are often able to circumvent this problem by evaluating a gap in experience with respect to its predicted likelihood, and adjusting generalizations so that they are sensitive to the probability that the gap was accidental versus systematic.

*Learning the structure of language*. Another challenge in the area of syntax concerns the ability to learn compositional and hierarchical structure. This also has been claimed to be unlearnable from positive data alone, but there are several models that successfully do just this (e.g. [12,61,62]).

One important problem often held up as the 'parade case' of innate knowledge is the so-called phenomenon of Aux inversion [38]. This involves patterns of the sort shown in question/answer pairs such as '*Is Mary happy*?'/ '*Mary is happy*'. Given such data, a child might be tempted to infer that questions are formed from declaratives by inverting the order of the subject noun and auxiliary (or modal). But this generalization incorrectly predicts that the interrogative that corresponds to '*The boy who is smoking is crazy*' should be (the ungrammatical) '*Is the boy who smoking crazy*?'

The fact that children do not make such errors has been interpreted as evidence that crucial knowledge about the structure sensitivity of grammatical rules must be innate [38]. However, Lewis and Elman [63] demonstrated that a neural network trained on examples of well-formed sentences that mimicked the types and frequencies of sentences found in real language could in fact learn the correct form of complex interrogatives. This happens because there are many other sentences present in the input (to children as well as to these networks) that provided evidence of a set of interacting generalizations which, taken together, allow the network to expect complex sentences of the sort it has never before seen. The important lesson from this is to expand our notion about what counts as positive evidence. To the extent that language behavior reflects interactions of multiple generalizations, networks appear able to learn not only the separable generalizations, but also how to combine them in novel ways.

Summing up, the record is very positive. Connectionist models now offer alternative hypotheses for many important developmental phenomena, and in several cases, appear to provide a richer and more accurate account of those phenomena than hypotheses from behavioral work.

That said, I believe the field is now at a crossroads. To some extent, the modeling work has captured the low-lying fruit. This is not to minimize the accomplishments. But we have arrived at the point where the easy targets have been identified but the tougher problems remain. In what follows I present what I see as the significant outstanding challenges.

## What is to be done

I suggest there remain two major challenges for developmental modeling. One concerns missing pieces: the phenomena that have not, by and large, been addressed. The second concerns strategy: taking development more seriously in a deeper and more profound way.

### *Missing pieces*

To a large extent, computational models of development have focused on cognitive phenomena. There are important exceptions (e.g. [8,46,64,65]) but they are few in number. Little modeling has been done in the realm of social cognition (there is some work that looks at social interactions, but this tends to have an evolutionary focus, rather than developmental), and even less in the development of affect and emotion. Thus, the computational literature has yet to contribute much to our understanding of theory-of-mind phenomena, development of self-identity, let alone exotica such as the development of moral judgment or play. Until recently, there was little computational work in the area of conceptual learning, 'theory theory', and development of causal inference. Happily, these areas at least are now being addressed [66].

The modeling literature has also had little to say about the role that is played by the child itself as an actor in the process of development. Children have their own agendas. They have drives, desires, things that draw their attention and things they ignore. Even young infants are not passive absorbers of external stimuli, completely at the mercy of their environment. Input is not the same as intake, and the effects of learning are obvious. Modeling the role of such endogenous forces is not intractable [67] but is far from easy.

Finally, although models of acquired neuropsychological disorders abound, much less has been done in the modeling of abnormal development. In my view, this reflects a bias on the part of modelers that is inherited from the developmental community, which is to consider cases of developmental disorders from the perspective of a broken adult system. But as Karmiloff-Smith, Thomas, and colleagues have pointed [68–70], when something is 'broken' in infancy (or before), the course of development itself will be altered. The result may be an outcome that differs in profound ways from the typically developing child.

This does not mean that atypical development is uninformative about typical outcomes. Just the contrary is true: such cases offer valuable information about the cascading effects that events in early development have on later stages. Because such indirect effects are difficult to study, computational models should be especially important in exploring these complex and sometimes non-obvious interactions over time, as the few models that have taken this approach illustrate [71–74]. The overall paucity of such models, however, leads to what I view as the more serious shortcoming in the current state of the modeling work, which is that we have yet to fully meet the challenge of taking development seriously.

### Missing development

Connectionist models are, by and large, models of learning. There is good reason for this. We are profoundly adaptive creatures, and our experiences play a huge role in who we come to be. But the neonate who responds to a moving light is not the same creature as the toddler who learns names for things. The toddler, the older child, and the adolescent respond to the same experience in different ways. In other words, the child is a constantly changing target, and models in which learning is the only source of change run the risk of seriously missing the point.

An important source of the differences that appear in early life are maturational (using the term here to denote developmental changes that are primarily endogenous). To be sure, the neonate's view of the world is different from a two-year-old's in part because of differences in experience and learning But developmental differences in the maturity and anatomical spacing of photoreceptors in the retina also give the infant a view of the world that is literally different from that of the older child. This in turn can have important consequences for what is learned [64,65].

*Causal chains in development*. A second factor to be taken into account is that the process of development itself involves a causal chain in which earlier stages shape and constrain what will happen at later stages, in ways that we have only begun to appreciate. The causal chain can be quite complex and have consequences both over time and across domains.

An excellent example of such a complex causal chain is suggested by the research of Karmiloff-Smith and colleagues, who have studied individuals with Williams Syndrome (WS). Both control subjects and WS individuals show superior recognition of upright faces, compared with upside-down faces. But in WS, this difference is attenuated: they are only slightly worse at recognizing upside down faces, and better than controls. One explanation might be that the 'face processing module' in those with WS is not working as it should. But the more likely account, proposed by Karmiloff-Smith and colleagues [68], is more complex and far more interesting. It appears that WS individuals rely to a greater extent on featural rather than configural information. Why should this be? This feature-based strategy might arise from difficulties in focusing attention, which could itself be the result of difficulties in early infancy in controlling saccadic eye movements. On this hypothesis, a relatively specific deficit early in life has a cascading effect on subsequently acquired behaviors: very early difficulties controlling eye movements lead to difficulties in control of attention, which at a later stage leads to difficulties in integrating global information, leading to a greater reliance on featural information, which then mitigates the disadvantage that upside-down faces have relative to upright faces (which are normally perceived using configural cues). Development is a linked chain of cause → effect relationships. Importantly, these relationships cut across not only time but domains (see Figure 1).

Or consider the question of where language comes from. One possibility is that language depends on neural machinery that is exclusive to itself, and on a set of isolated genes that have only to do with language. In this scenario, the acquisition process might still depend on external input to serve as a trigger (or possibly not; see [38]). This story is relatively simple and straightforward.

Contrasted with this is the emergentist view, that language sits at the crossroads of a number (how many – a dozen? thirty?) of small phenotypic changes in our species that interact uniquely to yield language as the outcome [2,43]. Here, language is seen as a domain-specific outcome that emerges through the interaction of multiple constraints, none of which is specific to language (see, for example, Figure 2). This is a harder story to tell because it means identifying, first, what those ingredients are; second, the ways in which they differ from non-human primates; and third, how exactly it is that they come together to make language possible. Crucially, these 'language ingredients' undoubtedly have interactions that occur over time, forming a causal chain, which may be difficult to untangle.

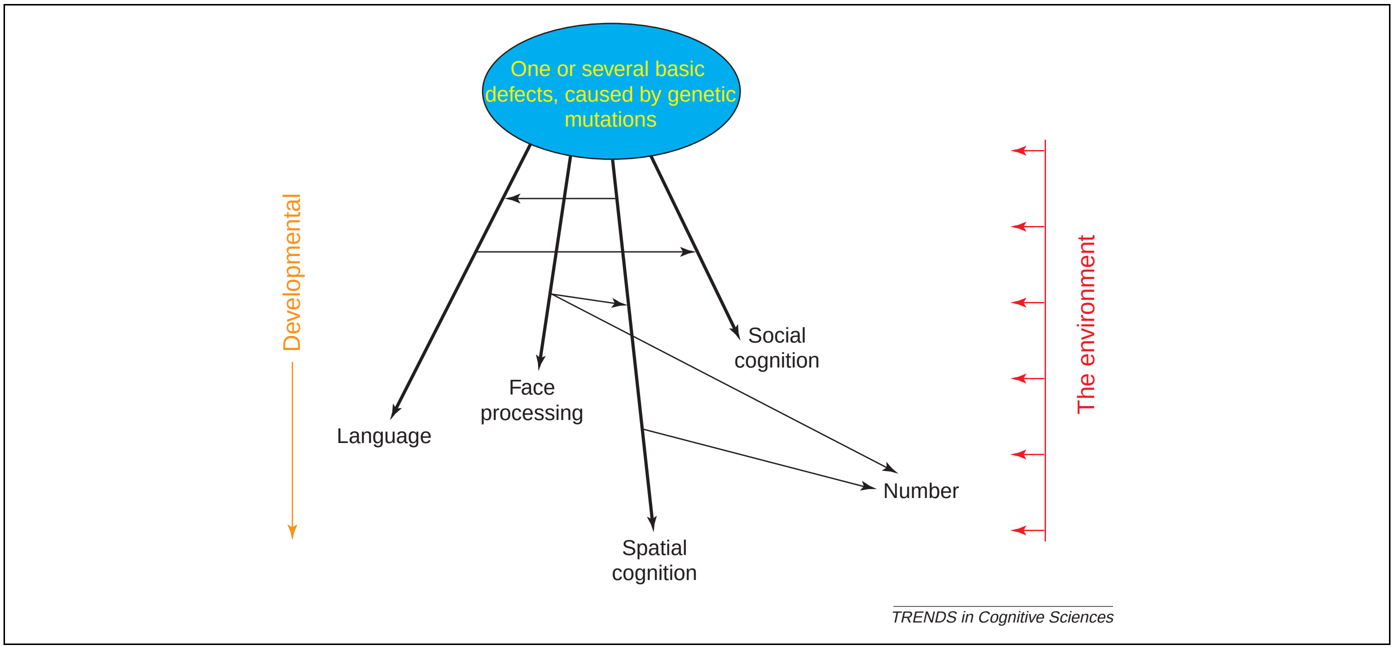


**Figure 1.** Low-level impairments can have cascading effects on the developing system over time. Early genetic mutations might give rise to a complex series of events across developmental time, in which initial deficits have a cascading sequence of outcomes, both across time and across domains. (After A. Karmiloff-Smith, *Kavli Institute for Brain and Mind Distinguished Lecture*, November 17, 2004, with permission).

It is precisely for such accounts that modeling plays a crucial role. But all of this means taking development more seriously. We have begun already to see examples of this, as in models of face recognition, imprinting, or aspects of visual development [8,46,64,65].

## Conclusions

The above barely scratches the surface of where we need to go next. I would argue that we also need to explore the much richer interactions that occur between behaviors as they unfold over time. This is not an easy task and will require models of a scale and complexity that surpass current technology. The 'scaling problem' has often been the bane of modeling. Models that work in single domains might not easily scale up to more complex situations involving development across multiple domains. So building more complex models of development also means solving the scaling problem (see Box 1).

Intriguingly, it may be that development itself provides some clues as to how to build more complex models. We begin, after all, as a single-celled organism. It is through development that we achieve the complexity of form and behavior that marks the mature individual – possibly as many as $100 \times 10^{12}$ cells, each with its own spatial location and connectivity, integrated within a machine of incredible complexity that only now and then do we barely

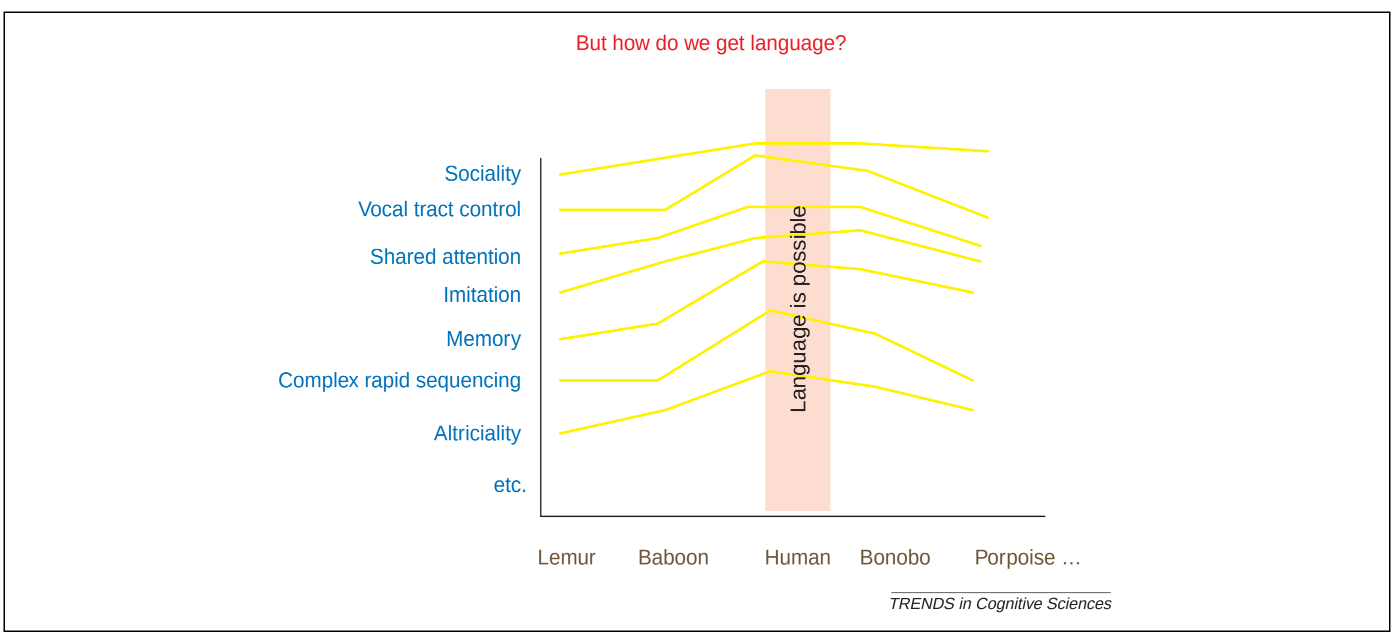


**Figure 2.** An emergentist hypothesis about the prerequisites for language. Humans share many traits with other species, but there are variations in the form of these traits. Human language emerges when a set of language-enabling traits takes a specific form. The interactions of these traits then makes language possible.

**Box 1. Questions for future research**

- *Extending the range of domains*. There remain several important areas of development for which little modeling work has been done and which provide exciting challenges for future research. These include social, emotional, physical and moral development, among many others.
- *Multi-tasking*. Most models focus on single behaviors. But to the extent that, in the developing child, behavior in one domain interacts with behavior in another, it will be important to have models that are capable of multi-tasking.
- *Scaling*. Modeling typically begins by abstracting and simplifying a problem, so that irrelevant details are excluded from study. But in the real world children have to deal with these additional complexities. An important challenge will be to develop models that can scale up to deal with more realistic and detailed behaviors.
- *Cascading effects over time*. Many complex behaviors are the result of cascading processes that depend on earlier stages of development, and which involve complex interactions between different domains. Modeling such interactions over time is the most important and exciting challenge of all, and clearly requires solutions to the first three challenges above.

understand. As we seek to model this process, we might also come to understand deeper principles about how complexity can arise from simple origins. In other words, development may be Nature's solution to its own scaling problem.

## Acknowledgements
This work was supported by grant MH6051701A3. Although they do not necessarily share the opinions expressed here, I am grateful for many discussions about this topic with Liz Bates, Mark Johnson, Annette Karmiloff-Smith, Jay McClelland, and Kim Plunkett.

## References

1 Brent, M.R. (1996) Advances in the computational study of language acquisition. *Cognition* 61, 1–38
2 Elman, J.L. *et al*. (1996) *Rethinking Innateness: A Connectionist Perspective on Development*, MIT Press
3 Elman, J. Computational approaches to language acquisition In *Encyclopedia of Language and Linguistics* (Brown, K., ed), Elsevier (in press)
4 Munakata, Y. and McClelland, J.L. (2003) Connectionist models of development. *Dev. Sci.* 6, 413–429
5 Chomsky, N. (1957) *Syntactic Structure*, Mouton
6 McCandliss, B.D. *et al*. (2002) Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn. Affect. Behav. Neurosci.* 2, 89–108
7 McClelland, J.L. *et al*. (2002) Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiol. Behav.* 77, 657–662
8 O'Reilly, R.C. and Johnson, M.H. (2002) Object recognition and sensitive periods: Computational analysis of visual imprinting. In *Brain Development and Cognition: A Reader* (2nd edn) (Johnson, M.H. *et al*., eds), pp. 392–413, Blackwell Publishers
9 Marchman, V. (1993) Constraints on plasticity in a connectionist model of the English past tense. *J. Cogn. Neurosci.* 5, 215–234
10 Munakata, Y. and Pfaffly, J. (2004) Hebbian learning and development. *Dev. Sci.* 7, 141–148
11 Newport, E.L. *et al*. (2001) Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler* (Dupoux, E., ed.), pp. 481–502, The MIT Press
12 Elman, J.L. (1993) Learning and development in neural networks: The importance of starting small. *Cognition* 48, 71–99
13 Chalup, S.K. and Blair, A.D. (2003) Incremental training of first order recurrent neural networks to predict a context-sensitive language. *Neural Netw.* 16, 955–972
14 Hertwig, R. and Todd, P.M. (2003) More is not always better: The benefits of cognitive limits. In *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (Hardman, D. and Macchi, L., eds), pp. 213–231, John Wiley & Sons
15 Cochran, B.P. *et al*. (1999) Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *J. Mem. Lang.* 41, 30–58
16 Kersten, A.W. and Earles, J.L. (2001) Less really is more for adults learning a miniature artificial language. *J. Mem. Lang.* 44, 250–273
17 McClelland, J.L. (1989) Parallel distributed processing: Implications for cognition and development. In *Parallel Distributed Processing: Implications for Psychology and Neurobiology* (Morris, R.G.M., ed.), pp. 8–45, Clarendon Press
18 Shultz, T.R. (2003) *Computational Developmental Psychology*, MIT Press
19 Rumelhart, D.E. and 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* (Rumelhart, D.E. and McClelland, J.L., eds), pp. 216–271, MIT Press
20 Plunkett, K. and Marchman, V. (1993) From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition* 48, 21–69
21 Rogers, T.T. *et al*. (2004) U-shaped curves in development: A PDP approach. *Cogn. Dev.* 5, 137–145
22 Marchman, V.A. *et al*. (2004) Productive use of the English past tense in children with focal brain injury and specific language impairment. *Brain Lang.* 88, 202–214
23 Marslen-Wilson, W. and Tyler, L.K. (2000) Rules, representations, and the English past tense. In *Exploring Cognition: Damaged Brains and Neural Networks: Readings in Cognitive Neuropsychology and Connectionist Modelling* (Cohen, G. *et al*., eds), pp. 301–317, Psychology Press
24 McClelland, J.L. and Patterson, K. (2002) Rules or connections in past-tense inflections: What does the evidence rule out? *Trends Cogn. Sci.* 6, 465–472
25 Patterson, K. *et al*. (2001) Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia* 39, 709–724
26 Plunkett, K. and Marchman, V.A. (2002) *Learning from a Connectionist Model of the Acquisition of the English Past Tense*, MIT Press
27 Marcus, G.*et al.* (1992) Overregularization in language acquisition. *Monogr. Soc. Res. Child Dev.* Whole Vol. 57
28 Ullman, M.T. (2001) The declarative/procedural model of lexicon and grammar. *J. Psycholinguist. Res.* 30, 37–69
29 Joanisse, M.F. and Seidenberg, M.S. (1999) Impairments in verb morphology after brain injury: A connectionist model. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7592–7597
30 Hare, M. and Elman, J.L. (1995) Learning and morphological change. *Cognition* 56, 61–98
31 Marcus, G.F. (2001) *The Algebraic Mind: Integrating Connectionism and Cognitive Science*, MIT Press
32 Ferretti, R.P. and Butterfield, E.C. (1986) Are children's rule-assessment classifications invariant across instances of problem types? *Child Dev.* 57, 1419–1428
33 Munakata, Y. (2001) Task dependency in infant behavior: Toward an understanding of the processes underlying cognitive development. In *Emerging Cognitive Abilities in EarlyIinfancy* (Lacerda, F. *et al*., eds), pp. 29–52, Erlbaum
34 Munakata, Y. *et al*. (1997) Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychol. Rev.* 104, 686–713
35 Munakata, Y. and Yerys, B.E. (2001) All together now: When dissociations between knowledge and action disappear. *Psychol. Sci.* 12, 335–337
36 Thelen, E. and Smith, L.B. (1994) *A dynamic Systems Approach to the Development of Cognition and Action*, MIT Press
37 Baker, C.L. (1979) Syntactic theory and the projection problem. *Linguistic Inquiry* 10, 533–581
38 Crain, S. (1991) Language acquisition in the absence of experience. *Brain Behav.* 14, 597–611
39 Anderson, S.R. and Lightfoot, D.W. (2000) The human language faculty as an organ. *Annu. Rev. Physiol.* 62, 697–722

40 Pinker, S. (1984) *Language Learnability and Language Development (Cognitive Science Series)*, Harvard University Press
41 Pullum, G.K. and Scholz, B.C. (2002) Empirical assessment of stimulus poverty arguments. *Linguistic Rev.* 19, 9–50
42 Scholz, B.C. and Pullum, G.K. (2002) Searching for arguments to support linguistic nativism. *Linguistic Rev.* 19, 185–223
43 Elman, J.L. (1999) The emergence of language: A conspiracy theory. In *The Emergence of Language* (MacWhinney, B., ed.), pp. 1–27, Erlbaum
44 MacWhinney, B. (1999) *The Emergence of Language*, Erlbaum
45 Quartz, S. and Sejnowski, T.J. (1997) The neural basis of cognitive development: A constructivist manifesto. *Behav. Brain Sci.* 20, 537–596
46 Shrager, J. and Johnson, M.H. (1996) Dynamic plasticity influences the emergence of function in a simple cortical array. *Neural Netw.* 9, 1119–1129
47 Elman, J.L. (1994) Implicit learning in neural networks: The importance of starting small. In *Conscious and Nonconscious Information Processing: Attention and Performance Vol. 15* (Umilta, C. and Moscovitch, M., eds), pp. 861–888, MIT Press
48 Goldowsky, B.N. and Newport, E.L. (1993) Modeling the effects of processing limitations on the acquisition of morphology: The less is more hyopthesis. In *Proceedings of the 24th Annual Child Language Research Forum* (Clarke, E., ed.), pp. 124–138, CSLI
49 Brent, M. and Cartwright, T. (1996) Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61, 93–125
50 Christiansen, M.H. *et al*. (1998) Learning to segment speech using multiple cues: A connectionist model. *Lang. Cogn Processes* 13, 221–268
51 Elman, J.L. (1990) Finding structure in time. *Cogn. Sci.* 14, 179–211
52 Cartwright, T. and Brent, M. (1997) Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition* 63, 121–170
53 Elman, J.L. (1995) Language as a dynamical system. In *Mind as Motion: Dynamical Perspectives on Behavior and Cognition* (van Gelder, R.P.a.T., ed.), pp. 195–225, MIT Press
54 Mintz, T.H. (2002) Category induction from distributional cues in an artificial language. *Mem. Cogn.* 30, 678–686
55 Redington, M. *et al*. (1998) Distributional information: A powerful cue for acquiring syntactic categories. *Cogn. Sci.* 22, 425–469
56 Gomez, R.L. and Gerken, L.A. (2000) Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4, 178–186
57 Saffran, J. (2001) The use of predictive dependencies in language learning. *J. Mem. Lang.* 44, 493–515
58 Regier, T. and Gahl, S. (2004) Learning the unlearnable: The role of missing evidence. *Cognition* 93, 147–155
59 Elman, J.L. (1998) Generalization, simple recurrent networks, and the emergence of structure. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, p. 6, Erlbaum
60 Tenenbaum, J. (1999) Bayesian modeling of human concept learning. In *Advances in Neural Information Processing Vol. 11* (Kearns, M.J. and Solla, S.A., eds), pp. 59–65, MIT Press
61 Christiansen, M.H. and Chater, N. (1999) Toward a connectionist model of recursion in human linguistic performance. *Cogn. Sci.* 23, 157–205
62 Boden, M. and Blair, A. (2003) Learning the dynamics of embedded clauses. *Appl. Intell.* 19, 51–63
63 Lewis, J.D. and Elman, J.L. (2001) A connectionist investigation of linguistic arguments from poverty of the stimulus: Learning the unlearnable. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pp. 552–557, Erlbaum
64 Dominguez, M. and Jacobs, R.A. (2003) Developmental constraints aid the acquisition of binocular disparity sensitivities. *Neural Comput.* 15, 161–182
65 Jacobs, R.A. and Dominguez, M. (2003) Visual development and the acquisition of motion velocity sensitivities. *Neural Comput.* 15, 761–781
66 Rogers, T. and McClelland, J.L. (2004) *Semantic Cognition: A Parallel Distributed Processing Approach*, MIT Press
67 Johnson, M.H. and Morton, J. (1991) *Biology and Cognitive Development: The Case of Face Recognition*, Blackwell
68 Karmiloff-Smith, A. (1998) Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci.* 2, 389–398
69 Karmiloff-Smith, A. and Thomas, M. (2003) What can developmental disorders tell us about the neurocomputational constraints that shape development? The case of Williams syndrome. *Dev. Psychopathol.* 15, 969–990
70 Thomas, M. and Karmiloff-Smith, A. (2002) Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling? *Behav. Brain Sci.* 25, 727–787
71 Paterson, S.J. *et al*. (1999) Cognitive modularity and genetic disorders. *Science* 286, 2355–2358
72 Karmiloff, K. and Karmiloff-Smith, A. (2001) *Pathways to Language: From Fetus to Adolescent*, Harvard University Press
73 Karmiloff-Smith, A. (2003) Genotype/phenotype relations: A cognitive developmental approach/Les relations entre genotype et phenotype: Une approche cognitive developpementale. *Enfance* 55, 33–38
74 Thomas, M.S.C. and Karmiloff-Smith, A. (2003) Modeling language acquisition in atypical phenotypes. *Psychol. Rev.* 110, 647–682