Connectionist Models in Developmental Cognitive Neuroscience:
Critical Periods and the Paradox of Success

Mark S. Seidenberg

Department of Psychology

University of Wisconsin-Madison

Jason D. Zevin

Sackler Institute for Developmental Psychobiology

Weill Medical College of Cornell University

Address for correspondence:

Mark Seidenberg

Department of Psychology

University of Wisconsin

1202 W. Johnson St.

Madison WI 53705

Phone: 608-263-2553

Email: seidenberg@wisc.edu

Abstract

Connectionist models have made significant contributions to understanding developmental phenomena, mainly by providing novel computational accounts of behavioral emergence and change. What is the fate of such models given the increasing interest in and information about the biological bases of development? We consider this issue with respect to the classical idea of a critical period for acquiring language. The standard view is that neurobiological developments on a strict maturational timetable create limits on language learning capacity. Computational analysis suggests the opposite: that learning itself creates neurobiological conditions underlying the "closing" of the critical period. The critical period example suggests how connectionist models can continue to provide a necessary level of analysis intermediate between behavior and brain.

1. Introduction

Our task in being invited to contribute to this volume was to critically assess the contributions of computational models to the study of the developmental issues that were the focus of the conference. This task seemed unmanageable given the variety of computational architectures in the marketplace and the range of issues in brain and cognitive development to which they have been applied. We therefore set a more modest goal: to consider the role of such models in light of the rapidly increasing body of knowledge about the brain bases of behavior and its development. To further simplify our task, we will largely focus on connectionist models of the PDP variety[1] (Rumelhart, McClelland, et al., 1986), for several reasons:

   • because they are intrinsically developmental: they simulate the course of acquiring a skill or type of knowledge. Many other computational models are not;

   • because they are explicitly intended to bridge overt behavior and its brain basis, providing insight about both;

   • because they have been applied to a broader range of phenomena in cognition and development than other frameworks, and so there is a significant body of work to consider, including other work presented at this conference;

   • because many of the issues that arise concerning connectionist models are not specific to this type of model but rather concern the role of computational models more generally;

   • because we happened to have worked extensively with such models.

2. What Have We Learned From PDP Models About Development?

It is nearly 20 years since the publication of the PDP "bible," the two volumes edited by Rumelhart and McClelland that introduced the basic concepts and initial applications of their approach. As one of us said at the time, "The authors have laid out an approach that will be discussed and developed for years to come." (Seidenberg, 1986) and so it has. There have been many applications of this approach to developmental phenomena, not surprising given its emphasis on learning. Any summary of the advances achieved within this framework relevant to development (see also Munakata & McClelland, 2003) would have to include the following. This list is not exhaustive, and not every model instantiates every point.

A. Unifying acquisition and skilled performance within a single theoretical framework. Connectionist models are intrinsically developmental. The models learn to perform tasks, such as predicting the location of a hidden object (Munakata et al., 1997), determining the meaning of a written word (Plaut & Shallice, 1993; Harm & Seidenberg, 2004, or generating the past tense of a verb (e.g., Rumelhart & McClelland, 1986; Joanisse & Seidenberg, 2000). Each model instantiates a developmental trajectory in the course of learning the task. Overt performance changes over time; however, the same principles

govern acquisition and skilled performance which are merely different points on a developmental continuum (Seidenberg & MacDonald, 1999).

The importance of linking acquisition and skilled performance should not be underestimated; in many areas, such as language, acquisition and skilled performance are studied largely separately with different principles used to explain phenomena in the two areas (see Seidenberg & McDonald, 1999, for discussion).   This pattern occurs in many domains, including perception (e.g., object recognition), motor performance (e.g., walking, reaching), and cognition (e.g., numerical knowledge).

B. Identifying mechanisms of developmental change. Connectionist models have provided mechanisms that explain why developmental changes occur, a significant step beyond the often acute descriptions of behavioral change in the developmental literature. Changes in overt behavior are closely tied to – and emerge from – specific computational properties of these models, which affect knowledge representations and their use in performing different tasks.

The classic illustration of this point is the reconceptualization of stage-like behavior in development, initiated by McClelland (1989) whose modeling demonstrated how developmental stages in performance of the balance-beam problem (Siegler, 1976) could arise from nonlinear transitions in network behavior. In standard PDP networks, knowledge acquisition is gradual and continuous. However, the overt behavior of the system may change more abruptly, reflecting the aggregate effects of many small weight changes.

A second example is provided by research on stages in learning to read. Beginning readers have been described as proceeding through a series of stages (Ehri, 1995): logographic, alphabetic, orthographic; pre-alphabetic, partial alphabetic, full alphabetic, consolidated alphabetic; and so on, depending on how finely the stages are stratified. These observations capture some broad facts about reading acquisition, at a level that engages many educators. These developments are usually ascribed to changes in the child's reading "strategy," based on developing metalinguistic awareness of how writing relates to speech. However, children's behavior is not strictly stagelike insofar as at any one point in time they read different words by different strategies (as Ehri has noted). The "strategy" idea suggests a level of conscious intent that may not be present. Moreover, it is not clear what kinds of processes each strategy entails, and the extent to which they differ. Finally the description leaves open the crucial question as to the forces that cause the child to change strategy.

These behavioral changes arise in a different manner in connectionist models of reading (Seidenberg & McClelland, 1989; Harm & Seidenberg, 1999). The phenomena reflect the child's growing sensitivity to the internal structure of the written and spoken forms of words. This occurs in a connectionist model as a byproduct of the learning process. Our reading models initially learn on a word by word basis. As the model is exposed to a broader sample of words, it begins to pick up on statistical regularities over various levels of structure that the orthographic and phonological representations potentially afford.   Thus the model progressively differentiates the orthographic and

phonological structures of words and the mappings between them, ending up with knowledge that spans multiple levels (letters, digraphs, onset, rime, and other subword structures, words). This is a form of implicit learning. There is no homunculus guiding the model's choice of "strategy;" rather the characteristics that the strategy description are intended to capture fall out of basic computational properties of the model. The point here is not to deny that children ever engage in explicit learning or problem solving strategies; rather it is that as applied to this aspect of early reading the strategy notion is at best a broad characterization of developmental processes that have a particular computational basis, which is relevant to understanding both the fine detail of the behavior, why it occurs, and how it could be facilitated or remediated.

C. Determining the extent to which behavior reflects general vs. domain-specific types of learning or processing mechanisms, or types of knowledge representation.   One of the main ideas in the PDP approach is that behavior (and its brain bases) can be explained in terms of a relatively small number of principles concerning knowledge representation, learning, processing, and so on.   This aspect of the approach is clearly related to Piaget's attempt to describe general cognitive principles underlying learning and developmental change. As in Piaget's day (Piatelli-Palmarini, 1980), controversy explodes when this type of approach is applied to language (cf. the Rumelhart and McClelland, 1986, verb learning model). Such models have challenged a core tenet of modern linguistic theory, that language involves highly abstract, domain specific forms of knowledge and thus cannot be derived from general capacities to think and learn (Seidenberg, 1997).

There continue to be strongly divergent views on this matter. The main complaint about the application of Piaget's concepts to language was that they could not account for any interesting aspect of grammar (see discussion in Piatelli-Palmarini, 1980).[2] However, there now are several plausible applications of PDP principles to specific grammatical phenomena, including ones that have been difficult to explain within grammatical theory. For example, Haskell et al. (2003) provide an account of how children acquire a subtle grammatical distinction, the dispreference for plural modifiers in compounds (e.g., *rats-eater*), which had been repeatedly argued (Pinker, 1999, and elsewhere) to provide evidence for a type of innate grammatical knowledge (i.e., the idea that word formation rules operate at ordered levels or strata in the mental lexicon; Kiparsky, 1982; Gordon. 1985).   Inadequacies of the level-ordering theory had been noted almost since its creation (Spencer, 1991), and so the phenomena resisted systematic explanation. Haskell et al. show that the phenomena relate to phonological and semantic properties of the modifier, from which a variety of degrees of modifier acceptability follow. The Haskell et al. work suggests a plausible account of the *rats-eater* phenomena and promising directions for future research. See articles in Christiansen, Chater, and Seidenberg (1999) and citations therein for other applications of connectionist concepts to linguistic phenomena.

This research is part of a broader reconceptualization of the language acquisition problem (e.g., Bates, 1994; Saffran et al., 1996; Seidenberg, 1997) largely motivated by the PDP approach. In the standard linguistic characterization of the "logical problem of language acquisition," the child's goal is to converge on the grammar of a language and the

main problem is the poverty and variability of the input (Pinker, 1989). Within this framework there is indeed a puzzle about how a child could ever acquire a particular language. However, many of the considerations that make language learning difficult and seem to implicate innate grammatical knowledge (e.g., the lack of explicit negative evidence, the lack of consistent feedback about grammaticality) are moot if the acquisition task is reformulated as acquiring the neural networks that support comprehension and production (Seidenberg, 1997, 2003; Lewis & Elman, 2001).

D. Unifying phenomena previously thought to be unrelated. One result of applying a common set of theoretical principles to a broad range of phenomena has been the discovery of unexpected commonalities among them. Consider, for example, the inflectional and spelling systems of English. Inflection is a component of natural language, which has evolved over eons in one species, humans. Spelling is a technology, a method for representing speech invented only a few thousand years ago. Hence one might expect the two systems to have little in common. Parallel programs of research within the connectionist framework suggest otherwise, however.

In brief, there are close commonalities between the two types of knowledge. Both are quasiregular (Seidenberg & McClelland, 1989): they are dominated by rule-like regularities but admit forms that deviate from these central tendencies in differing degrees; the "exceptions" partially overlap with the rule-governed forms and so are rarely arbitrary; the deviant forms tend to cluster among the higher frequency words in the language, and so on. These properties are handled by connectionist networks that learn probabilistic mappings between codes (spelling, sound, meaning). The correspondences between these types of knowledge extend beyond mere similarity: they are both parts of a lexical system that encodes relations among sound, spelling, and meaning (Harm & Seidenberg, 2004). It follows from this view that these types of knowledge should be similarly affected by neuropathology, which recent research has confirmed. Patterson et al. (submitted) show that semantic impairments in a group of patients with a progressive neurodegenerative disorder exhibit similar deficits on six tasks. For example, patients who regularize past tenses (e.g., *leaved*) also regularize the pronunciations of words such as *listen*. The explanation for both is the greater role of semantics in determining the correct forms of atypical lower frequency words (Plaut et al., 1996; Joanisse & Seidenberg, 1999). These results are consistent with earlier observations about connections between speech and reading (Liberman et al., 1989).

Two other major components of this approach that play significant roles in explaining developmental phenomena should be mentioned. One is *the nonlinear combination of probabilistic constraints*. It is a fundamental insight that bits of information that are not very informative in isolation can become highly constraining when taken together. For example, the infant learning to identify the boundaries between spoken words relies on several types of information, no one of which is highly reliable. The connectionist framework provides a mechanism whereby such probabilistic constraints could be learned and combined efficiently. This is a significant insight about a basic characteristic of learning and skilled performance. In the language acquisition literature, it is often observed

that various structures (or cues) occur too infrequently to provide a basis for learning a linguistic generalization, leading to the inference that innate knowledge is required. There is a further intuition that the existence of multiple unreliable cues merely worsens the learning problem. However, such intuitions do not take into account the possibility of nonlinear combinations of constraints, which connectionist models illustrate.[3]

The other major concept *is division of labor* (Seidenberg, 1992; Plaut et al., 1996; Harm & Seidenberg, 2004). The idea is that learning to perform a complex task involves finding an efficient partitioning of the problem among interacting subsystems. The load that one subsystem carries depends on what is happening in other systems that are also contributing to performance.   Division of labor (DOL) is the opposite of modularity, the idea of independent, isolable subsystems whose functions are determined in advance (e.g., by evolution). The DOL idea has been worked out in greatest detail with respect to reading, but represents a more general ideal about how the brain solves complex problems (see Gordon & Dell, 2003, for another application of the concept).

What is clear from this partial survey is that connectionist models have mainly contributed to understanding developmental phenomena by explaining how they arise from what Rumelhart and McClellland called the microstructure of cognition. That is, the models (and the broader theoretical framework they instantiate) explain how prominent aspects of behavior emerge from a particular type of computational system. Although this type of computation was said to be "neurally inspired" and broadly consistent with basic characteristics of brain function, the models that most directly bear on cognitive and linguistic development have been only weakly constrained by facts about neurobiology. Of course, many people are actively exploring how the same computational principles can be used to understand neurobiology (e.g., brain bases of learning and memory), but our focus is on models that account for behavior and its development. Looking at the literature one is struck by an apparent tradeoff between biological fidelity and level of phenomenon: The further the phenomena from higher level cognition, the more biologically realistic the model.

Thus a questions arises: what is the fate of connectionist models in light of advances in and a growing emphasis on understanding the brain bases of behavior, at levels ranging from the neural circuits revealed by neuroimaging to cellular neurobiology? One possibility is that current models of language and cognition are an intermediate step: the limiting factor is our lack of knowledge about aspects of neurobiology that are relevant to cognition and its development. When such information becomes available, models will be developed that are based on it. Current-generation cognitive models will be replaced by neurobiological ones. Some enthusiastic neuroimagers go further, asserting that knowledge about brain circuitry will obviate computational theories of the behavior entirely.

An alternative view is that the natural progression is one in which computational models are not just an interim stage but necessary for achieving the goal of linking brain and behavior. The main constraint is not lack of knowledge of the brain bases of cognition (although what is known is indeed limited). Rather, in many areas of cognition we have

not achieved the understanding of basic *computational* mechanisms that are a prerequisite for understanding their brain bases. On this view, computation modeling complements other methodologies: Behavioral experiments can tell us what the effects of stimulus and task manipulations are on overt responses. Imaging can tell us what brain regions and circuits are involved in processing. There is a further need for computational models that explain how brain mechanisms give rise to behavior. Otherwise the behavioral work is isolated from the brain and the neuroimaging has an atheoretical, descriptive character.

One can also argue, as we will here, that computational models serve a further function: they have explanatory value insofar as they provide an appropriate level for understanding basic characteristics of complex cognitive and developmental phenomena. This function does not merely reflect lack of knowledge about brain mechanisms, but rather reflects a deeper issue about the levels of analysis that are relevant to explaining behavior.

This last claim may seem a throwback to an earlier era in which much less was known about the brain (Seidenberg, 1994). It may merely be the case that we still know too little about the brain bases of cognition to build biologically realistic models that account for detailed aspects of cognitive functions and how they develop. Or perhaps modelers merely need to work harder on their neurobiology. Surely neuroimaging techniques are on an upward trajectory toward identifying neural circuits underlying many aspects of cognition. Even neural development is within our grasp: we can image the brains of 3 month olds (Dehaene-Lambertz et al., 2002).   Perhaps we will all be neurobiologists in the end.

These issues matter insofar as they reflect different views about where cognitive neuroscience should be heading. But really, who knows what will happen and who's to say what should? Fields make these decisions, not individuals. Still, we might remove this discussion from the realm of pure speculation and gain some insight by pursuing the following strategy: let us consider some aspect of cognition for which a considerable amount is already known about its neurobiological and computational bases, and then examine how they contribute to understanding the phenomena. With that in mind, we turn to critical periods.

3. Critical Periods In Language Acquisition And Other Domains

The term *critical period* has been used with reference to several types of developmental phenomena, including neurobiological (e.g., critical periods for gene expression), sensory-motor (e.g., development of the visual system), and cognitive (e.g., language).   Our focus is on the sense of critical period (dating from Lenneberg, 1967) referring to a time window during which species-typical communicative behaviors (e.g., song in zebra finches; language in humans) are learned (Doupe & Kuhl, 2000). This rapid and efficient early learning contrasts with subsequent restrictions on learning capacity (often termed a loss of plasticity). Evidence for a critical period in language acquisition (CPLA) derives from studies of diverse aspects of normal and atypical language acquisition, including observations such as:

1. Languages are learned on a consistent maturational timetable despite input that is impoverished (relative to what is learned) and highly variable (across languages, individuals and cultures). This consistency extends across modalities, insofar as signed and spoken languages are learned on much the same timetable (Petitto, 1999).

2. Learning a second language later in life is difficult even with extended experience (Johnson & Newport, 1989; Flege, Liu, & Yeni-Komshian, 1999; Birdsong & Molis, 2002), in contrast to learning two languages from birth;

3. Late learning of a first language (e.g., by deaf individuals whose exposure to sign language is delayed) results in incomplete mastery (Newport, 1990; Neville & Bavelier, 2000; Mayberry, Locke, & Kazmi, 2002);

4. The timing rather than amount of exposure to a language apparently determines level of mastery (Johnson & Newport, 1989);

5. The high degree of plasticity during the critical period affords reorganization in the presence of atypical input (e.g., recruitment of "auditory" language areas given signed input; Neville et al., 1998) and following early brain injury or neuropathology (Dennis, 2000; Vargha-Khadem et al., 1992). Thus these conditions do not preclude acquiring normal language. In contrast, recovery of linguistic functions following brain injury in adulthood is typically limited (Holland, 1989), paralleling the limits on ability to learn a first or second language post critical period.

The standard view is that these facts can be explained in terms of a "language faculty" that evolved in humans (Chomsky, 1986). The realization of the language capacity in an individual requires exposure to a native language coinciding with the maturation or expression of this faculty under genetic control. This theory is thought to simultaneously explain many other facts as well, such as the universal properties of language; the creation of language without a model (Senghas, 2003); creolization and other cases where the child's language is more systematic the parents' (Singleton & Newport, 2004); and the species-specificity of language. The fact that language is learned during a critical period can be taken as independent evidence for the biological basis of the language faculty.

Within this standard approach, the brain bases of language and the critical period are often ignored because they are thought to shed little light on the nature of language qua language. [4]   When they are discussed, it is usually by analogy to the development of the visual system, often ocular dominance columns in cat. For example, as noted by Stromswold (1995), "Although the ability to learn language appears to be innate, exposure to language during childhood is necessary for normal language development, just as the ability to see is innate yet visual stimulation is necessary for normal visual development (Hubel and Wiesel, 1970). The hypothesis that exposure to language must occur by a certain age in order for language to be acquired normally is called the critical period hypothesis." The factors that close the critical period for language are not known but are discussed by reference to aspects of neurobiological maturation such as synaptic pruning,

the development of inhibitory synaptic connections, and other events that have been identified with the loss of plasticity in other species (Hensch, 2003).

Three broad issues arise in connection with this account.

1. There are controversies about each of the types of behavioral evidence seen as evidence for a CPLA. For example, the limits on language learning following the close of the putative critical period are not absolute; there are well-attested cases of late acquisition of native-like competence (e.g., Birdsong & Molis, 2002). The idea of a CPLA seems somewhat Anglocentric, of less concern perhaps in cultures where it is common to achieve high competence in a second or third language at different points in life. The conditions under which second languages are learned differ from those governing L1 acquisition and they have a substantial impact on L2 outcomes (e.g., Flege et al., 1999; Bialystok & Hakuta, 1999). Data concerning the effects of age vs. amount of exposure to L2 (Johnson & Newport, 1989) have been strongly contested (e.g., Bialystok & Hakuta, 1994). Studies of late learners of a first language raise questions about what their communicative experience was like prior to exposure to a natural language and how it affected subsequent learning. Nonlinguistic aspects of these individuals' experiences may also have differed greatly from those that prevail in the usual L1 learning situation. Evidence that language can be acquired despite severe injury to (indeed, complete removal of) left hemisphere tissue that normally supports language raises questions the locus and role of the putative language organ. Languages created without a model do not appear to exhibit the deep, abstract properties said to be characteristic of grammar, and the studies leave open whether development reflects strictly grammatical or more general cognitive processes. Thus, the standard account relies heavily on data that are at best open to other interpretations.[5]

2. The analogy to critical periods in visual development does not advance the issues very far. Analogies are not facts, no matter how forcefully asserted and repeated. Many interesting issues arise in comparing language and vision that are not captured by statements such as "the child's language 'grows in the mind' as the visual system develops the capacity for binocular vision." (Chomsky, 1993, p. 29). Language involves multiple types of sensory information (audition, vision), motoric output systems, a cognitive system that thinks and reasons, and so on. These components do not all develop exclusively in the service of language and thus have their own developmental trajectories. Language also consists of multiple subsystems, which represent different types of information, have different properties and exploit different capacities in different ways. Phonology, for example, which is tied to hearing and speaking, is different from word learning, which involves conceptual knowledge. There is a grain mismatch in comparing the development of this system of interlocking skills and parts with the development of a component of early vision. Even at the level of perceptual processing, the ability to interpret rapidly varying acoustic signals in speech is clearly different in kind and complexity from the segregation of inputs from different sensory organs at the periphery. [6]

3. The standard story assumes that language is essentially unlearned (and unlearnable) except for minor language-specific elements (Pinker, 1994). Given this view it

is plausible to equate language with other capacities that are fully realized within a limited window in early development. However, other recent theories of language and how it is acquired view the language acquisition problem very differently. One such alternative account emphasizes the role of statistical learning and the structure of the input to the child in converging on knowledge that supports the use of language (Saffran et al., 1996; Seidenberg, 1997; Gomez, this volume). This approach provides a different way of thinking about critical period phenomena, which we explore below.[7]

In summary, the standard evidence and arguments for critical periods attempt to establish the plausibility of the idea but leave open basic questions of fact. Still, acquiring language early in life seems patently easier than learning it later. We do not want to conclude from uncertainties about particular findings that the emperor has no clothes; perhaps merely a wardrobe malfunction. Moreover, learning language appears both easier than learning many other things and more age-dependent. Apparently, there is something to be explained here but the characterization of what is to be explained may need modification, which will in turn affect how it is explained.

4. **Biological Bases Of Critical/Sensitive Periods**

Plausibility arguments and analogies aside, is there direct biological evidence for a critical period in acquiring language? The loose coupling of the critical period concept to neurobiology is reflected in the fact that even the date at which it is said to close varies greatly, from 5-6 years (Pinker, 1994) to puberty (Lenneberg, 1967) and points beyond (Hakuta & Bialystock, 1994). There is, however, an extensive literature on the biological bases of critical/sensitive periods as studied in other species. In brief, there are many candidates for intrinsic changes to neural networks that may limit plasticity, including synaptic pruning, changes in the number and distribution of neurotransmitter receptors, and the maturation of inhibition. These developments can be experience-independent. For example, the development of inhibitory circuitry can be controlled by genetic (Hensch et al., 1998; Huang et al., 1998; Fagiolini et al., 2004) and pharmacological (Fagiolini & Hensch, 2000) factors that can be decoupled from activity-dependent changes in plasticity evident during the critical period. For example, whatever the environment in which the animal is raised, interventions that induce the maturation of inhibitory circuitry also induce the onset and closure of a critical period for plasticity.

However, there is also strong evidence that learning itself also plays a role in limiting plasticity (Hensch, 2003; McClelland et al., 1999; Zevin & Seidenberg, 2002). Song learning in zebra finches provides an example that turns out to be relevant to language learning (more so than ocular dominance columns). Zebra finches typically learn song during a sensitive period that closes in early adulthood, after which new song elements are not added, and existing elements are not lost or rearranged. There is variability across bouts of song (e.g., with respect to number of repetitions of a song element) but it is highly constrained. In isolation this fact could be taken as evidence for a classical biologically-determined critical period, the closing of which shuts down the organism's capacity to

learn. However, this view is contradicted by a wealth of recent data implicating learning and environment in the critical period and its aftermath.

Although song is learned within a typical time frame, zebra finches exhibit plasticity well beyond the closing of this period. For example, the critical period can be extended by altering the bird's experience: adult birds raised in isolation can learn new song material from a tutor (Jones, Slater, & ten Cate, 1996). Songs learned under these conditions are slightly abnormal, either because of the experience-independent developments summarized above or because of learning from the bird's own song that occurs in isolation and interferes with subsequent learning. Plasticity can also be reinstated in adults via methods that disrupt motor output (Mehta & Williams, 1997) or auditory feedback (Brainerd & Doupe, 1999; Leonardo & Konishi, 1999). These findings sparked interest in the hypothesis that some form of maintenance learning continues to occur throughout adulthood. When normal feedback is available, the bird's output matches his target and reinforces the existing song representation. Disruptions of feedback lead to maladaptive adjustments of song via this same mechanism. Indeed, lesions of brain regions known to be critical for initial song acquisition prevent changes of song in deafened birds (Brainerd & Doupe, 2000), providing strong support for this view.

Zevin, Bottjer and Seidenberg (2004) explored this further using white noise to prevent adult birds from being able to hear their own songs without damaging their hearing. Under these conditions the birds' songs degrade to varying degrees. This suggests that continued exposure to the bird's own song is necessary for maintenance, i.e., keeping the song network tuned. When noise was removed, the subjects showed no evidence of learning from exposure to a tutor, suggesting a limit on adult plasticity and learning capacity. [8] However, song did change over time with new access to auditory feedback, suggesting a continued capacity to learn.

Finally, results from a large-scale cross-sectional study suggest that plasticity continues to decline gradually late into adult life (Lombardino & Nottebohm, 2000). Variability in adult song, as characterized by the frequency and consistency with which particular song elements are produced declines throughout adulthood, suggesting a progressive entrenchment of the representations that underlie song production. Critically, the response to deafening was also shown to differ as a function of age, such that the effects of deafening were progressively slower and less severe with age.

What conclusions follow from this research? First, there is evidence for reductions in plasticity after the sensitive period, possibly related to non-experience dependent neurobiological developments that extend into adulthood. Second, song birds nonetheless retain plasticity beyond the critical period; continued exposure to song is required for maintenance, and song degrades when it is not available. Finally, learning itself plays a role in limiting plasticity. Degradation of feedback induces song change in a manner that depends on the integrity of nuclei involved in initial song acquisition. Susceptibility to disruption of auditory feedback declines steadily with age throughout the life span, in a manner that is highly correlated with behavioral stereotypy. These observations suggest

that the acquisition and gradual entrenchment of representations that support stereotyped song become increasingly resistant to change as a result of the process of learning itself.

Of course birds are not people and songs are not language. Nonetheless the zebra finch research suggests a picture that may plausibly extend to language learning, with a combination of independent neurobiological events and learning contributing to rapid initial learning followed by a gradual decline in plasticity. Close causal connections between developmental neurobiological events and language learning have not been established, however, and it is unclear whether they are the proximal causes of the major phenomena seen as evidence for a critical period in language acquisition. Of particular interest is the extent to which the landmark phenomena are attributable to learning itself. To investigate this question, we need a computational model.

5. **A Computational View Of Critical Periods**

What we seek is a computational-level theory of the basic CPLA phenomena, including the early rapid learning of L1, the gradual decrease in plasticity/language learning ability, and other facts such as differences in plasticity associated with different components of language and the effects of both the timing of exposure to a second language and structural relations between first and second languages. What follows are some steps toward such an account. PDP networks have not as yet incorporated facts about neurobiological development; the network's architecture, processing dynamics, and capacity to learn are typically fixed (leaving aside the cascade correlation approach; Shultz, this volume). Changes in network behavior are therefore due to how learning occurs given these initial constraints. Thus such networks can be used as a tool to gain evidence as to whether learning itself determines critical period phenomena. In effect, the networks can be used to decouple the confounds between experience and brain development that complicate the interpretation of many other types of data. Because the networks are pitched at this computational rather than neurodevelopmental level, we know in advance that there are some boundary phenomena that they will not capture correctly. For example, there are neurobiological developments associated with normal aging that affect plasticity, which this type of account ignores. The idea, however, is that an account pitched at a computational level intermediate between brain and behavior can provide important insight about –indeed a different view of – critical periods. Such a theory, taken with other facts about neurodevelopment that influence plasticity, should account for the range of observed behavioral outcomes.

5.1 PDP Networks and the Paradox of Success

Learning in PDP networks involves changes to weights on connections between units that are intended to capture the behavior of large ensembles of neurons (see O'Reilly & Munakata, 2000, for further detail). Models are typically initialized with small random weights, consistent with the idea that although early neurobiological development establishes preconditions for learning, humans are not born with highly structured knowledge. We know that early language learning is rapid, given appropriate experiential input. In PDP networks this occurs for specifiable technical reasons related to several

aspects of the architecture (see also Munakata & McClelland, 2003). Early exposure to target patterns results in large weight changes. The system is highly plastic at this point because unit outputs tend to fall toward the middle of the sigmoidal activation function, i.e., on the steep linear portion; as a result, the behavior of the system can change significantly as the weights change, supporting rapid learning. Early learning is mainly sensitive to pattern frequency. As the model learns from patterns sampled from a structured domain such as language, units begin to approach extremal values (e.g., 1 or -1) and several important effects are observed. First, some weights fall to small values, contribute little to accurate performance or additional learning, and are effectively pruned (as seen in neurobiological development). Second, the model becomes less sensitive to pattern frequency; once a pattern is learned, the error signal it produces is very small. Additional exposures have little effect, creating a discrepancy between the statistics of the input (i.e., actual frequency) and network performance. Thus learning in such systems involves transforming rather than merely matching the input statistics (see Singleton & Newport, 2004). Third, the model becomes more sensitive to similarities across patterns; new learning is easiest for patterns that overlap with previously-trained ones. This property also supports generalization: correct performance on novel stimuli in virtue of their similarity to trained patterns.

 Learning to perform a task correctly, then, means that the weights have assumed values that push activations close to desired values. Exposure to additional patterns from the same target set (e.g., language) produces small error signals. Except perhaps under extreme conditions in which the input to the network is radically changed (McClelland et al., 2002), the weights are difficult to adjust further because of their contributions to successful performance. The net result is that the model becomes less plastic in the sense of no longer allowing large weight changes with few exposures. This effect has been termed "entrenchment" (Ellis & Lambon Ralph, 2000; Zevin & Seidenberg, 2002; Elman et al., 1996).

 These network properties establish a basis for early rapid learning followed by a decline in plasticity, the classic "critical period" pattern. The "closing" of the critical period, then, is the loss of plasticity associated with success in learning a task, i.e., entrenchment. A problem arises because weights that are highly favorable for skilled performance of one task (e.g., using a first language) are unfavorable for other tasks (e.g., learning a second language). We call this the *Paradox of Success*: learning to perform a task with a high degree of proficiency may create conditions that interfere with further learning.

 In the case of language, this loss of plasticity is not a wholly negative thing. The knowledge that is acquired is systematic and represented in a way that supports generalization, a defining characteristic of language. The network is trained on examples but generalizes to novel patterns via similarity and recombination of existing elements. In effect the network assimilates new utterances to the existing structure. However, the network's capacity to generalize also gives it a strong tendency to assimilate L2 utterances as well. Running French utterances through a highly trained English network is not an efficient way to learn French. The obvious extreme cases are ones in which novel

phonological contrasts in L2 are not perceived because they are assimilated to existing categories in L1 (e.g., Best et al., 2001). Interference due to increasing entrenchment of L1 provides a basis for the decline in plasticity associated with the "closing" of the critical period.

The situation described above contrasts with the learning of unsystematic facts, such as names for things and individuals, which afford little generalization but do not show critical period effects or a decline in plasticity over time (modulo independent developments such as effects of aging on hippocampal function). In fact the learning of arbitrary facts exhibits tiny entrenchment ("age of acquisition") effects. In brief, there is empirical and computational modeling evidence that arbitrary associations learned earlier in life show a long-lasting advantage over associations learned later (Ellis & Lambon Ralph, 1999; Zevin & Seidenberg, 2002). In practice it takes less time to name an object or person if it was learned early in life, other factors aside. Note, however, that AoA effects differ from critical periods in two important respects. First, AoA effects concern particular items rather than systematic aspects of knowledge. Second, conditions that give rise to AoA effects do not typically result in a failure to acquire novel items. Rather, performance is merely poorer. In contrast, sensitive period effects result in the inability to acquire generalizable, systematic knowledge, for example the ability to categorize novel speech sounds.

Thus, the loss of plasticity associated with the CPLA seems to be specifically related to the capacity to generalize. If this is correct it suggests that there is no critical period-like loss of plasticity in cognitive domains that do not afford generalization. This would be a step toward explaining why only some types of knowledge show the critical period type of developmental profile.

6. Applications to "Late" Language Learning

The above analysis emphasizes computational factors thought to underlie the rapid learning followed by decline in plasticity characteristic of first language learning. With minor extensions we can apply this account to phenomena concerning "late" language learning, two types of which have been studied:

1. Learning a second language, which can occur at varying time lags relative to a first language.

2. Late learning of a first language. This condition has mainly been studied with regard to deaf children of hearing parents who are not exposed to a signed natural language (e.g., ASL) until relatively late (e.g., the onset of formal schooling; Mayberry et al., 2002; Newport, 1990; Singleton & Newport, 2004).   Late exposure to ASL is typically preceded by the development of home sign communicative systems that exhibit some important properties of natural language (Goldin-Meadow & Mylander, 1990).

In both cases, the phenomena concern language learning by individuals who have already had experience with another communicative system (L1 in the case of L2 learners; home sign in the case of late ASL learners).   For brevity we will refer to both cases as L1

followed by L2, although ASL is referred to as L1 in the late acquisition cases, in part because home sign, though rich, may not exhibit the full range of properties of natural languages. The main point is that experience with one communication system precedes experience with another.

We can now identify several computational and experiential factors that interact to determine L2 learning outcomes.

One factor is the structure of the learner's experience. There has been considerable research about how the conditions that govern language learning vary. For example, a child learns a first language under different conditions than an adult learning a second language (e.g., as an immigrant; in a classroom). The late L2 learner already knows a language, the L1 learner does not. Motivational and social factors, including the range of conditions under which L2 is used, account for considerable variance in L2 learning outcomes (Flege et al., 1991), much more so than in L1 learning. Similarly, home sign is learned (created, really) under conditions that differ from those in standard L1 learning. These differences among the conditions under which language is learned are important and contribute to observed outcomes. However here our concern is with other major aspects of the learning environment, suggested by the computational framework, which relate to the timing and consequences of learning.

L1 learning in the monolingual context is *blocked*: the child has extensive exposure to a single language. Language is learned rapidly and as noted above there is an increase in the capacity to generalize accompanied by a decrease in plasticity. It is then difficult (though not impossible of course) to assimilate L2 to this functioning language system. Other social and motivational factors aside, an important characteristic of late exposure to a second language is L1 and L2 experiences are *interleaved*: the individual continues to use L1 but is also exposed to L2. This condition is Blocked Early-Interleaved Late. The complement of this situation is a bilingual environment from birth, in which the child is exposed to both L1 and L2 (Interleaved Early). These conditions (summarized in Table 1) represent extreme points on a continuum determined by how well-learned L1 is at the time of exposure to L2. Similar considerations apply to the home sign vs. L1 situation, keeping in mind the likelihood that home sign differs in some respects from natural languages.

Two computational factors are operating here. One is the degree to which one system has been learned prior to exposure to the second system. L1 monolingual learning is easy: there is no prior entrenched knowledge to interfere with learning, and plasticity is high. Learning both L1 and L2 in the Interleaved Early condition is also relatively easy: no entrenchment here either, and plasticity is again high; the task is harder than learning a single language but ultimately learnable. L2 learning in the Interleaved Late condition is hard: L1 is entrenched and plasticity is low; thus knowledge of L1 interferes with learning L2. Note that this analysis does not make reference any changes over time in the substrates that support learning or the capacity to learn except those that are a consequence of learning itself. [9]

A second factor is whether the experiences with the two systems are interleaved or not. In the Early (bilingual) case, interleaving is good: neither language is entrenched and they both can be learned. This contrasts with the Blocked Early-Interleaved Late situation typical of late L2 learning. Experience with L2 is interleaved with L1. It is the continued experience with L1 that keeps the language network entrenched, interfering with the attempt to accommodate L2. Thus interleaved language experience has radically different effects at different points in time.

6.1 Catastrophic Interference, L1 Loss and L2 Interference

The interfering effect of L1 on L2 turns on the maintenance of L1 weights. As we have noted, zebra finches actively maintain song through continued feedback from the bird's own song. Disrupting this feedback causes the song to drift. (This drift is similar to that which occurs in the speech of individuals with adult hearing loss.) The role of continued experience with L1 in interfering with L2 learning can be understood with reference to the phenomenon of catastrophic interference in PDP networks. McCloskey and Cohen (1989) trained a simple feedforward network on a set of arithmetic problems (e.g., the "ones" problems: 1+0, 1+1, … 1+9). Once the model learned these problems, training switched to the twos problems (2+0, 2+1, etc.). McCloskey and Cohen observed that learning the twos problems resulted in loss of knowledge about the ones. They termed this retroactive interference effect in feedforward networks "catastrophic interference." It was thought to be a problem for such networks because learning new information typically does not cause unlearning of old information. However, the effect is not very general. It depends on strict blocking of the problem sets (Hetherington & Seidenberg, 1989); thus the ones are maintained if there are even a few additional exposures to them during training on the twos. The conditions that give rise to catastrophic interference rarely occur in human learning, outside the context of a verbal learning experiment. In real life, experiences of different things are interleaved, thankfully.

The relevance of these effects to the language learning case should be clear. Learning a second language does not result in unlearning of a first language because experience with L1 does not cease. Catastrophic interference is not a problem for human learners because experience is rarely blocked to the extent the effect demands. However, the down side is that maintenance of L1 interferes with learning L2; L1 remains entrenched for reasons discussed above. When it comes to language learning, then, proactive interference from L1 is more a problem than retroactive interference from L2.

It is interesting to note that catastrophic interference—unlearning of L1 via learning of L2—does occur on rare occasions. Pallier et al. (2003) studied adults who were born in Korea but adopted by French families in childhood (between the ages of 3 and 8 years old). These individuals had been exposed to Korean during the "critical period" but then became French speakers. fMRI data suggested that knowledge of Korean had not been preserved into adulthood; the activation patterns elicited by Korean stimuli were similar for these subjects and speakers who had not been exposed to Korean. Pallier et al.'s interpretation is that L2 (French) replaced L1 (Korean), a naturalistic example of catastrophic interference.

Experience with the two languages was strictly blocked; without any additional L1 "trials," knowledge of the language was lost.

This analysis is supported by the results of two other studies that used designs similar to Pallier et al.'s but yielded somewhat different results. Tees and Werker (1984) found that English-speaking adults who had been exposed to Hindi during infancy could discriminate Hindi speech sounds better than subjects without prior Hindi experience. Oh et al. (2003) found similar results for adults who had been exposed to Korean "prior to age 5 and very little afterward." (p. B54). They observed residual effects of early exposure among college students enrolled in Korean courses. Perhaps the crucial difference between the conditions in these studies compared to Pallier et al.'s is whether there was continued exposure to L1 or not. In the Pallier case, there was essentially no additional exposure to L1 following adoption and relocation. In the Tees and Werker case, the subjects stopped using Hindi as their main language, but they were raised in a Hindi-Canadian community and would have continued to hear Hindi. Subjects in the Oh et al. study also experienced more exposure to the abandoned L1 than in Pallier et al.'s experiment (the subjects were either born in Korea and immigrated to the US with their Korean families, or were US born offspring of Korean speakers).   This set of results is consistent with the strict blocking interpretation of catastrophic interference: L1 is unlearned when experience with the two languages is strictly blocked; however even intermittent re-exposure to L1 is sufficient to keep it maintained at level that facilitated adults' phonological perception.

This analysis makes a further prediction. Hetherington and Seidenberg (1989) showed that even in the highly artificial blocked condition that produced "catastrophic" interference, the model that had "unlearned" the ones problems could rapidly relearn them with additional training. This savings in relearning indicated that although the model produced incorrect responses, the early learning had not been completely erased. This finding suggests that it would be valuable to look at relearning among individuals such as the ones studied by Pallier et al. Although residual knowledge of Korean was not detectable at the resolution provided by fMRI, it might be observed in relearning, indicating that the original language was not completely replaced.

6.2 Interleaving Effects in L2 Learning

The more common circumstance is the one in which continued experience with L1 interferes with learning of L2. These effects are predicted to depend on the degree of entrenchment of the first language. The clearest cases are ones in which a high degree of L1 entrenchment interferes with aspects of late L2 learning. This phenomenon is likely to be relevant the well-studied problem of teaching adult Japanese speakers the r-l discrimination in English (Bradlow et al., 1997; McClelland, Fiez, & McCandliss, 2002). The English phonemic categories /r/ and /l/ are mapped onto a single category in Japanese (an alveolar tap closer to English /l/ than /r/). Adult Japanese speakers have difficulty discriminating and producing /r/ and /l/, even with extensive exposure to English. Several methods have been used to train this discrimination in adult Japanese speakers, with varying degrees of success (see references above). Whatever the upper limits on how well

the discrimination can be learned and maintained, it is clear that the problem is a difficult one. This difficulty could be due in part to continued use of Japanese during training. That is, regardless of how intensive or extended the training regime, or how it is structured, the Japanese subject continues to use Japanese, including the phoneme category to which English /r/ and /l/ are typically assimilated.   No language training regime can enforce cessation of L1 usage (although minimizing L1 usage is an apparent goal of "immersion" programs). Because the Japanese category is so well learned, even a small number of additional trials with it are sufficient to maintain the category and interfere with learning the remapping required for /r/-/l/ discrimination in English.[10]

Above we discussed cases in which L2 overwrote L1 in differing degrees. The Pallier et al., Oh et al., and Tees and Werker studies all emphasized the extent to which L1 knowledge had been lost as function of L2. It would be interesting to know whether the converse also occurred: initial exposure to L1 interfering with mastery of L2, which became the primary language. This would be consistent with evidence that late learners do not fully master subtle aspects of L2 grammar. These effects would be modulated by how much L1 learning occurred prior to exposure to L2 and the degree to which L1 continued to be experienced. Greater experience with the "abandoned" L1 would be expected to create greater interference with L2. This type of investigation would shed light on how much L1 entrenchment has to occur to create interference with L2.

6.3 Late Learners of ASL

The situation in which a sign language such as ASL is learned late is a particularly interesting one.   The literature emphasizes the fact that late learners do not achieve full command of the language's grammar, despite many years of experience with it (e.g., Newport, 1990).   This data seems to parallel similar limitations on L2 mastery (Johnson & Newport, 1989): in both cases the age at which the individual is exposed to the language has an impact on outcome; this seems a more powerful factor than the amount of experience with the language (a factor that Johnson & Newport attempted to control).   We have already suggested that in assessing outcomes associated with L2 learning, it is necessary to consider the degree of language learning that preceded onset of L2 experience. In addition, many researchers have systematically characterized a multitude of factors (e.g., range of conditions under which language is used; motivation and environmental support; frequency of language use and range of communication functions engaged) that differ in L1 vs. L2 learning (e.g., Flege, 1999; Flege et al., 1999; Bialystok & Hakuta, 1999). Both of these considerations come into play in interpreting late ASL learning outcomes.

The late ASL learner will typically have used a home sign communicative system. Such children are not communicatively bereft; they are merely receiving highly degraded natural language input. The important finding from studies of such invented communicative systems (e.g., Goldin-Meadow & Feldman, 1975; Goldin-Meadow & Mylander, 1990) is that they exhibit important properties of natural languages. Thus the relationship of this experience to late ASL learning is similar to the role of L1 in learning L2. There are some potential differences in these situations that are difficult to assess;

home sign may not exhibit all properties of natural languages, and it may not be as entrenched as a typical L1, because of the more limited conditions under which it is used. Still, the late ASL learner is not tabula rasa with respect to communicative ability, and the impact of this knowledge on later ASL acquisition needs to be considered. Note also that the late learner of ASL will typically continue to interleave home sign with ASL. Thus the situation has many of the characteristics of Late-Interleaved learning. Finally, the conditions that exist when the late learner is finally exposed to ASL differ from those in L1 learning.   The child may be learning from peers or teachers whose own early language experiences (e.g., home sign, ASL, signed English) and level of ASL mastery are highly variable.   These factors cannot be overlooked in considering why late learners fail to asymptote at the same level of grammatical competence as a child who learned ASL as a first language from deaf signing parents.   Thus, although many aspects of the late ASL learner's experience differ from the other language learning conditions we have discussed, the same principles and factors may govern outcomes in all these cases.

6.4 Similarity Effects

Finally, we must mention another major factor that contributes to L2 learning outcomes: the degree of similarity or overlap between the two languages. We have emphasized the negative impact of L1 learning on L2 acquisition. This emphasis was due to the fact that such data are typically seen as providing evidence for a critical period in language acquisition. However, the effects of L1 on L2 are not uniformly negative. Effects of L1 on L2 are governed by a complex equation whose variables are not fully understood. There is typological variation in phonology, morphology, syntax, organization of the lexicon, and other components of language. Similarity between L1 and L2 could facilitate learning: the fact that one language has inflectional morphology might make it easier to learn a second inflectional language. Similarity could interfere with learning, as in the r/l case, where sounds that are contrastive in English get absorbed into an existing attractor (perceptual category) for the Japanese. Dissimilarity could be helpful: the distinctiveness of Zulu clicks might make them easy for an English speaker to learn. Dissimilarity could also create difficulty: Chinese lacks determiners but that apparently does not make them easy to learn in L2 English.

It is clear that similarity relations between languages affect L2 outcomes (Bialystock & Hakuta, 1994), possibly by modulating effects of language entrenchment, the timing of exposure to a second language, and the interleaving of language experiences. Although progress has been made in understanding these effects, particularly in phonological learning (e.g., Best, McRoberts & Goodell, 2001), models of language learning are not advanced enough to make strong, detailed predictions about relative ease of learning across components of different languages. Here we can only reiterate the important point that language is not homogeneous; it consists of multiple components involving different types of knowledge, which exhibit different information structures. These differences among types of linguistic knowledge will affect L2 learning. For example, lexical learning has a high degree of arbitrariness (i.e., in the mapping between form and meaning), which makes it susceptible to modest entrenchment (AoA) effects but creates little interference with later

learning. In contrast, other domains of language (e.g., speech perception) require a high degree of generalization, which can give rise to effects of pernicious overassimilation.

7. Critical Periods: Complementary Roles of Biological and Computational Accounts

We have presented a summary of some of the biological factors that affect plasticity in nonhuman species, but noted that their connection to the loss of plasticity associated with the classic CPLA hypothesis was conjectural at best. We then sketched a computational learning theory of some of the major phenomena related to L1 and L2 acquisition ascribed to the CPLA.

Although a great deal remains to be learned about early neurobiological development, the findings and observations we have summarized suggest the tentative conclusion that there is no classical CPLA; at best the term describes what normally happens (language is learned rapidly with a gradual loss of capacity to learn other languages); however, there is little evidence for further claim that it is tied to biological developments on a maturational timetable. To the contrary, the major phenomena can be explained in terms of a theory of learning that specifies what is learned, how well it is learned, when it is learned, and how learning one system (e.g., L1) affects subsequent learning (e.g., L2). The crucial explanatory concept is entrenchment, which emerged from considering the phenomena computationally rather than biologically. This factor interacts with the systematic, generative aspect of language to create interference with later language learning.   On our account, learning creates neurobiological changes that reduce plasticity; this is the opposite of the standard theory, in which intrinsic neuromaturational changes limit learning. In short, it is viable working hypothesis that there simply *is no* critical period for language acquisition in any theoretically interesting sense, merely learning phenomena that create the Paradox of Success.

Returning to the broader question about the role of connectionist-computational models in the era of the brain with which we began this chapter, the investigation of critical period phenomena suggests that the models continue to play an important role in clarifying phenomena and why they occur. Surprisingly, some relatively simple principles about learning in PDP networks can explain many of the major critical period phenomena, including ones often thought to implicate biological rather than experiential factors. The account presented above is only schematic; much of the simulation research that is needed to assess this analysis has not been completed (although see Zevin & Seidenberg, 2002, and Ellis & Lambon Ralph, 2000). However, we have built this account on a much more extensive body of PDP modeling and it turns on some rather generic properties of such networks rather than specialized machinery.

What about the neurobiological evidence concerning the brain bases of changes in plasticity? We see this evidence as complementing the computational analysis in several ways. First, neurobiology establishes the preconditions and substrate for the learning phenomena we have described. There is no learning without the relevant neurobiological substrate and the learning effects we have modeled obviously have a neurobiological basis. The critical distinction is between neurobiological events that change the substrate of

learning (e.g., proliferation and pruning) and changes in what has been learned within this substrate (e.g., a song or language). Second, some of the neurobiological facts lend corroborative support to the computational analysis. For example, the modeling predicts that if the relevant experience were withheld (without introducing other pathologies as in the case of Genie), language could still be learned outside the notional critical period. These conditions cannot be created with human subjects, but the predicted effect is seen in zebra finches.   As we have discussed, other details about song learning in this species are consistent with a learning-based view of the critical period for language.[11]   Third, neurobiology is relevant to accounting for detailed facts about observed outcomes. There are changes in neurobiology over the lifespan that affect the capacity to learn. They are not the kinds of changes that create the decline in plasticity characteristic of language learning, which is better explained by the computational learning theory. However, they play a role in explaining other facts, such as age-related changes in language, memory, and learning, and indeed set the boundary conditions on language learning outcomes.   Thus at least for the behavioral phenomena we have discussed, the theoretical equation that provides the most explanatory leverage is one in which a computational theory of learning + neurodevelopmental boundary conditions yield a range of observed behavioral outcomes.

In summary, examining one phenomenon in detail has led to the conclusion that computational and biological accounts play complementary roles in understanding at least some major cognitive phenomena. It seems likely that what is true of critical periods in language acquisition will carry over to many other phenomena for which our current level of understanding is similarly limited. In many areas of cognition, the limiting factor on understanding their brain bases is not a lack of neurobiological knowledge (although what we know is indeed limited). Rather it a lack of understanding of the phenomena at the computational level represented by neural networks.   The computational level theory provides insight about many phenomena, and suggests directions for additional behavioral and neurobiological research. This seems to be a vivid realization of the vision of the originators of the PDP approach (Rumelhart & McClelland, 1986).

Footnotes

1. For brevity we will use the term "connectionist models" in reference to this class of connectionist models, acknowledging that there are other types, e.g., interactive activation models incorporating localist representations. We exclude these models because they have not been applied as widely to neural or behavioral development, probably because they are fundamentally less well-suited to these topics.

2. One could say the same about current theories of language evolution, which are more focused on explaining the fact of language evolution rather than the existence of particular aspects of grammar. Such theories do not explain any of the aspects of language that are putatively highly abstract and unmarked in the overt structure of utterances.

3. People are obviously capable of combining cues nonlinearly. Solving crossword puzzles provides an everyday example. Considered independently, the across and down clues to particular words may only weakly suggest possible answers, but their probabilities greatly increase if both have the same letter at the position where they intersect.

4. We could not find a reference for this assertion but it is an attitude we have encountered frequently among generative grammarians.

5. For linguists such as Jackendoff (2003), the definitive evidence concerning the innateness of grammar comes from recent studies of the emergence of sign language without a formal model in Nicaragua (Senghas, 2003). These findings are fascinating but the view that they are the pure expression of innate universal grammar is questionable. The extent to which the emerging communicative system will come to exhibit various properties of grammar, including the abstract ones thought to differentiate language from other types of cognition, is not yet clear. Utterances in the early stages of the development of the system were closely tied to immediate actions and individuals. The language is changing over time, but this may be due to perceptual and memory forces that are general rather than specific to language. In short, the study may provide elegant evidence for how language emerges from communicative needs and general cognitive factors.

6. Chomsky has variously compared the innateness of language to many aspects of vision, including perception of 3-dimensional space, color, object perception, attention to visual primitives, and binocular vision, going so far as to claim there is an innate "grammar of vision" (Werry, 2002). These analogies are not a substitute for actual visual science. For example, the common comparison between language and the development of ocular dominance columns echoed in the Stromswold quote is vexed given recent evidence that the latter is even more hard-wired than previously thought. Using modern visualization methods, Crair et al. (2001) demonstrated that the earliest inputs from thalamus to primary visual cortex in cat are already segregated by eye. In ferrets, which are highly altricial, complete removal of visual experience does not interfere with the development of ODCs (Issa et al., 1999; Crowley & Katz, 1999). This is quite unlike language which definitely requires substantial experiential input.

7. Insofar as statistical learning plays a central role in both language and vision (compare, e.g., Saffran et al., 1996, and Kirkham et al., 2003), the two may indeed turn out to be similar but not for the reasons Chomsky suggested.

8. There are several reasons why the subjects did not learn from tutors. Zebra finches exhibit other sensitivities to learning conditions; for example, they learn from a live tutor but not an auditory recording. The atypical conditions that existed when the subject was removed from noise may have been sufficient to disrupt attention to the tutor. A more interesting possibility is that the subject can attend to the tutor but does not learn because of the degraded condition of the bird's own song. The subject is not tabular rasa at this point but rather possesses an entrenched deviant song. This existing knowledge may interfere with subsequent learning.

9. Bever (1981) proposed that critical periods in language learning might be a consequence of "normal acquisition processes," which is what we are claiming here. However, Bever viewed normal acquisition differently than we and his "psychogrammar" theory involves a mechanism unlike the ones we have discussed. Bever's concept was that the mental representation of grammar emerged to mediated the competing demands of comprehension and production. Once this alignment of the two language functions was achieved in L1, psychogrammar's raison d'etre was lost, it became disconnected from language use, and it was no longer available to mediate L2 learning. Although these specific concepts do not play a role in our account, Bever's core idea that language learning itself causes the "closing of the critical period" was prescient.

10. McClelland et al. (2002; this volume) have explored the hypothesis that the difficulty that Japanese speakers have in learning the r/l distinction is an unwanted byproduct of a Hebbian learning mechanism. In brief, with Hebbian learning the speaker's own incorrect perception of /r/ and /l/ as alveolar flaps will tend to strengthen the existing phonemic structure rather than promote learning the English phonemic categories. This account is similar to ours in that it is the speaker's continued experience with and maintenance of L1 phonology that interferes with additional learning. McClelland et al. place this effect in the learning algorithm itself; we place it in the interleaving of English and Japanese experiences. The two positions might be different faces of the same coin if the brain utilizes a learning mechanism that has both Hebbian and supervised aspects, as some believe (see O'Reilly & Munakata, 2000, for discussion).

11. Of course, comparisons between the species break down at some point given that their various capacities (perceptual, learning, cognitive) are vastly different. Thus their behavior also deviates from ours in important ways. For example, the loss of plasticity in zebra finches is not related to an ability to generalize, contrary to our conjecture about humans. One song may simply exhaust the finch's capacity to learn, whereas learning one language does not exhaust ours. The severe loss of plasticity associated with learning a language but not many other types of knowledge implicates additional factors that are not relevant to zebra finch behavior.

Acknowledgements

References

Bates, E. (1994). Modularity, domain specificity and the development of language. *Discussions in Neuroscience*, *10*, 136-149.

Best, C.T., McRoberts, G.W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109,* 775-794.

Bever, T.G. (1981). Normal acquisition processes explain the critical period for language learning. In K.C. Diller, ed. *Individual differences & universals in language learning aptitude*. Newbury House Publishers, Inc., Rowley, MA.

Bialystok, E., & Hakuta, K. (1994). *In other words: The science and psychology of second language acquisition.* Basic Books, New York.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences in second language acquisition. In D. Birdsong, ed. *Second language acquisition and the critical period hypothesis.* Erlbaum, Mahwah, NJ.

Birdsong, D., & Molis, M. (2002). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*, 215-249.

Bradlow, A.R., Pisoni, D.B., Yamada, R.A. and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299-2310.

Brainerd, M., & Doupe, A. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature*, *404*, 762-766.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger, New York.

Chomsky, N. (1993). *Language and Thought.* Moyer Bell, Wakefield, R.I.

Crair, M. C., Horton, J. C., Antonini, A., & Stryker, M. P. (2001). Emergence of ocular dominance columns in cat visual cortex by 2 weeks of age. *Journal of Cognitive Neurology*, *430,* 235-249.

Christiansen, M.H., Chater, N. & Seidenberg, M.S. (Eds.) (1999). Connectionist models of human language processing: Progress and prospects. Special issue of *Cognitive Science*, 23, 415-634.

Crowley, J. C., & Katz, L. C. (1999). Development of ocular dominance columns in the absence of retinal input. *Nature Neuroscience*, *2*, 1125-1130.

Dehaene-Lambertz, G., Dehaene, S.,. & Hertz-Pannier, L. (2002). Functional

neurogimaging of speech perception in infants. *Science, 298*, 2013-2015.

Dennis, M. (2000). Developmental plasticity in children: The role of biological risk, development, time, and reserve. *J. Communication Disorders, 33*, 321-332.

Ehri, L. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18,* 116-125.

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

Elman, J., Bates, E., et al. (1996). *Rethinking Innateness*. MIT Press, Cambridge MA.

Fagiolini, M., & Hensch, T.K. (2000). Inhibitory threshold for critical period activation in primary visual cortex. *Nature, 404*, 183-186.

Fahlman, S.E., & Lebiere, C. (1991). The cascade-correlation learning architecture. Technical Report CMU-CS-90-100, School of Computer Science, Carnegie Mellon University, August.

Flege, J.E. (1999). Age of learning and second-language speech. In D. P. Birdsong , ed. *New Perspectives on the Critical Period Hypothesis for Second Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum. pp. 101-132

Flege, J.E., Yeni-Komshian, G. & Liu, S. (1999) Age constraints on second language acquisition. *Journal of Memory and Language, 41*, 78-104.

Goldin-Meadow, S. & Mylander, C. Beyond the input given: The child's role in the acquisition of language. *Language*, 1990, *66* (2), 323-355.

Goldin-Meadow, S., & Feldman, H. (1975). The creation of a communication system: A study of deaf children of hearing parents. *Sign Language Studies, 8,* 226-236.

Gordon, J., & Dell, G. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science, 27*, 1-40.

Gordon, P. (1985). Level-ordering in lexical development. *Cognition, 21*, 73–93.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading, and dyslexia: Insights from connectionist models. *Psychological Review, 163*, 491-528.

Harm, M.W., & Seidenberg, M.S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*, 662-720.

Haskell, T.R., MacDonald, M.C., & Seidenberg, M.S. (2003). Language learning and innateness: Some implications of *compounds research. Cognitive Psychology, 47*, 119-163.

Hensch, T.K. (2003). Controlling the critical period. *Neuroscience Research, 47*, 17-22.

Hensch, T., Fagiolini, M., Mataga, N., Stryker, M., Baekkeskov, S., & Kash, S. F. (1998). Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science*, *282*, 1504-1508.

Holland, A. (1989) Recovery in aphasia. In F. Boller & J. Grafman, eds. *Handbook of Neuropsychology* vol. 2, pp. 83-90. Elsevier: North Holland.

Huang, Z., Kirkwood, A., Pizzorusso, T., Porciatti, V., Morales, B., Bear, M., Maffei, L., & Tonegawa, S. (1999). BDNF regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell*, *98*, 739-755.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *Journal of Physiology*, *206*, 419-436.

Issa, N., Trachtenberg, J., Chapman, B., Zahs, K., & Stryker, M. (1999). The critical period for ocular dominance plasticity in the ferret's visual cortex. *Journal of Neuroscience*, *19*, 6965-6978.

Jackendoff, R. (2003). Precis of Foundations of Langauge: Brain, Meaning, Grammar, Evolution. *Behavioral and Brain Sciences*, *26*, 651-665.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences*, *96*, 7592-7597.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60-99.

Jones, A. E., ten Cate, C., & Slater, P. J. B. (1996). Early experience and plasticity of song in adult male zebra finches (*taeniopygia guttata*). *Journal of Comparative Psychology*, *110*, 154-169.

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst, & N. Smith, eds. *The structure of phonological representations*. Foris, Dordrecht, The Netherlands.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition, 83*, B35-B42.

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

Leonardo, A., & Konishi, M. (1999). Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature*, *399*, 466-470.

Lewis, J.D., & Elman, J.L. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisted. *Proceedings of the 26th Annual Boston University Conference on Language Development*.

Lombardino, A. J., & Nottebohm, F. (2000). Age at deafening affects the stability of learned song in adult male zebra finches. *Journal of Neuroscience*, *404*, 5054-5064.

Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research*, *36*, 1258-1270.

Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature*, *417*, 38.

McClelland, J. L. (1994). The interaction of nature and nurture in development: A parallel distributed processing perspective. In P. Bertelson, P. Eelen, & G. d'Ydewalle, eds. *International perspectives on psychological science, Volume 1: Leading themes,* pp. 57-88. Erlbaum, Hillsdale NJ.

McClelland, J.L. (1989). Parallel distributed processing: Implications for cognition and development. In R.G.M. Morris, ed. *Parallel distributed processing: Implications for psychology and neurobiology*, pp. 8-45. Oxford University Press.

Munakata, Y. & McClelland, J.L. (2003). Connectionist models of development. *Developmental Science, 6*, 413-429.

Neville, H. J., & Bavelier, D. (2000). Specificity and plasticity in neurocognitive development in humans. In M. S. Gazzaniga, ed. *The new cognitive neurosciences,* pp. 83-98. MIT Press, Cambridge, MA.

Neville, H.J., Bavelier, D., Corina, D., Rauschecker, J., Karni, Al, Lalwani, A., Brain, A., Clark, V., Pezzard, P., & Turner, R. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *PNAS, 95*, 922-929.

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

Oh, J.S., Jun, S., Knightly, L., & Au, T. K.-F. (2003). Holding on to childhood language memory. *Cognition, 86,* 53-64.

Pallier, C. Dehaene, S., Poline, J-B., LeBihan, D., Argenti, A.-M., Dupoux, E. & Mehler, J. (2003) Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral Cortex*, *13,* 155—161.

Patterson, K., Lambon Ralph, M.A., Jefferies, E., Jones, R., Hodges, J.R., & Rogers, T.T. (submitted). 'Pre-semantic' cognition in semantic dementia: Six deficits in search of an explanation.

Petitto, L. A. (1999). The acquisition of natural signed languages. In C. Chamberlain, J. Morford & R. Mayberry, eds. *Language acquisition by eye*, pp. 41-50. Erlbaum: Mahwah NJ.

Piatelli-Palmarini, M. (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky.* Harvard University Press.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure.* MIT Press, Cambridge MA.

Pinker, S. (1999). *Words and rules.* Basic Books, New York.

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

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*(5), 377-500.

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

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 5294.

Seidenberg, M.S. (1986). Jacket, hard cover editions of D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing*, vols. I and II. MIT Press.

Seidenberg, M.S. (1992). Beyond orthographic depth: Equitable division of labor. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning*. Amsterdam: North-Holland.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, *275*, 1599-1603.

Seidenberg, M.S., & MacDonald, M.C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*. *23*, 569-588.

Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cognitive Development, 18,* 511-531.

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

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology, 49, 370-407.*

Spencer, A. (1991). *Morphological theory: An introduction to word structure in generative grammar*. Cambrige, MA: Blackwell.

Stromswold, K. (1995) The cognitive and neural bases of language acquisition. In M. Gazzaniga, ed. *The cognitive neurosciences*, pp. 855-870. MIT Press, Cambridge MA.

Tees, R.C., & Werker, J.F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds*. Canadian J. Psychology, 38*, 579-590.

Vargha-Khadem F., Polkey C.E. (1992). A review of cognitive outcome after hemidecortication in humans. In F.D. Rose & D.A. Johnson, eds. *Recovery from*

brain damage: reflections and directions, pp. 137–151. Plenum Press, New York.

Werry, C. (2002). *Rhetoric and reflexivity in Chomskyan and cognitive linguistics.* Unpublished doctoral dissertation, Carnegie Mellon University.

Williams, H., & Mehta, H. (1999). Changes in adult zebra finch song require a forebrain nucleus that is not necessary for song production. *Journal of Neurobiology*, *39*, 14-28.

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

Zevin, J. D., Seidenberg, M. S., & Bottjer, S. W. (2004). Limits on reacquisition of song in adult zebra finches exposed to white noise. *Journal of Neuroscience, 24*, 5849-5862.

Table 1

Language Learning Conditions

_____
_____

<div align="center">

TIMING

_____

</div>

Learning Condition (L2 exposure)	Early	Late
Blocked interferes with L1 (CI)	Monolingual	L2
Interleaved L2	Bilingual	L1 interferes with

_____
_____

Note: "Early" refers to the child's language experience from birth. "Late" exposure assumes that L1 has become entrenched. CI = catastrophic interference.