Psychological Review
1999, Vol. 106, No. 3, 491-528

Phonology, Reading Acquisition, and Dyslexia: Insights from Connectionist Models

Michael W. Harm
Mark S. Seidenberg
University of Southern California

The development of reading skill and bases of developmental dyslexia were explored using connectionist models. Four issues were examined: the acquisition of phonological knowledge prior to reading, how this knowledge facilitates learning to read, phonological and non phonological bases of dyslexia, and effects of literacy on phonological representation. Compared with simple feedforward networks, representing phonological knowledge in an attractor network yielded improved learning and generalization. Phonological and surface forms of developmental dyslexia, which are usually attributed to impairments in distinct lexical and nonlexical processing "routes," were derived from different types of damage to the network. The results provide a computationally explicit account of many aspects of reading acquisition using connectionist principles.

Phonological information plays a central role in learning to read and in skilled reading. Several converging sources of evidence indicate that learning to relate the spoken and written forms of language is a critical step in learning to read (see Adams, 1990, for an extensive review). Children's knowledge of the phonological structure of language is a good predictor of early reading ability (Bradley & Bryant, 1983; Tunmer & Nesdale, 1985; Mann, 1984; Olson, Wise, Conners, Rack, & Fulker, 1989; Shankweiler & Liberman, 1989) and impairments in the representation or processing of phonological information are implicated in at least some forms of developmental dyslexia (Manis, Seidenberg, Doi, McBride-Chang, & Peterson, 1996; Stanovich, Siegel, & Gottardo, 1997). Use of phonological information is not limited to beginning readers; skilled readers also rely on this information in identifying words (Van Orden, Pennington, & Stone, 1990; Lukatela & Turvey, 1994; Seidenberg, 1985; Jared & Seidenberg, 1991; Perfetti & Bell, 1991; Perfetti, Bell, & Delaney, 1988) and integrating words with sentence contexts (Pollatsek, Lesch, Morris, & Rayner, 1992). Phonology plays an important role in working memory (Gathercole & Baddeley, 1993) and may be particularly relevant to retaining information about the literal forms of sentences while ambiguities are resolved. A principal goal in developing models of word recognition is to explain how phonological information is represented in lexical memory and used in reading and how anomalies related to the representation or use of phonology give rise to specific patterns of reading impairment.

The present research investigated the role of phonological information in early reading and dyslexia. Our focus was on using the theoretical framework developed by Seidenberg and McClelland (1989, hereafter SM89) and Plaut, McClelland, Seidenberg, and Patterson (1996, hereafter PMSP) to understand normal and impaired reading acquisition. The initial applications of this framework were to phenomena related to skilled reading (SM89). We then showed how it could account for forms of dyslexia observed in adults following brain injury (Plaut et al., 1996). The present paper represents a further extension of this framework to encompass developmental forms of dyslexia. We present new simulations addressing normal and disordered developmental phenomena. Our research focuses on four issues:

1. Phonological representation. In order to address developmental issues we needed to devise an approach to phonological representation that was an advance over the representational schemes used in previous models of reading.

Michael W. Harm, Computer Science Department, University of Southern California; Mark S. Seidenberg, Neuroscience Program and Department of Psychology, University of Southern California. Michael W. Harm is now at the Center for the Neural Basis of Cognition, Carnegie Mellon University.

This work was supported by National Institute of Mental Health (NIMH) grant 47566 and National Institute of Childhood Health and Development (NICHD) grant 29891 and by a Research Scientist Development Award (NIMH 01188).

We thank Lori Altmann for assistance in developing the phonological representation, and Joseph Allen, Joseph Devlin, Laura Gonnerman, Marc Joanisse, Pat Keating, Maryellen MacDonald, David Plaut, and Robert Thornton for many helpful discussions. We are especially grateful to Franklin Manis for his insights and support.

Correspondence concerning this article should be addressed to Michael W. Harm, Center for the Neural Basis of Cognition, 115 Mellon Institute, 4400 Fifth Avenue, Pittsburgh, Pennsylvania, 15213. Electronic mail may be sent to mharm@cnbc.cmu.edu.

Whereas previous computational models largely focused on adult performance, the current work focuses on how phonological representations develop, and how properties of phonological representation affect learning to read. This emphasis required developing a more sophisticated method of representing phonological information than earlier models had demanded given the kinds of questions they addressed. The simulations described below explore the use of a phonological representation analogous to the attractor networks that have been used in models of semantic representation (e.g. Plaut & Shallice, 1993). Although not a fully general account of phonological structure, this representation incorporates some important representational principles and it allowed us to address developmental issues in considerable detail. It also provides the beginnings of a computational account of a variety of phonological phenomena such as categorical perception of phonemes, although this is not the primary focus of the research.

2. Role of prior phonological knowledge in learning to read. Children bring to the reading acquisition task considerable knowledge of phonological structure derived from experience with spoken language. This is an important aspect of the child's experience that previous models have ignored. For example, the architecture of the Seidenberg and McClelland model included a set of phonological units that would allow the network to represent the pronunciations of words, but this representation did not itself encode very much information about the structure of English phonology. Similarly, the Coltheart, Curtis, Atkins, and Haller (1993) model is endowed with a way of deriving rules governing the correspondences between graphemes and phonemes but this process is not constrained by facts about the phonological structure of the language; hence the model can learn rules for phonological systems that could not occur in human languages. Both models were in effect learning about phonological structure at the same time they learned to map between orthography and phonology. The child, in contrast, already knows a great deal about phonology and mainly has to learn how orthographic representations map onto it. In the simulations presented below, we addressed how the existence of prior knowledge of phonological structure–and differences in the quality of this knowledge–affected learning to read.

3. Bases of developmental dyslexia. The third issue we address concerns the bases of developmental dyslexia. The goal is to be able to explain impairments in learning to read in terms of anomalies in the normal system. There is now good evidence that developmental dyslexia occurs in at least two forms (Manis et al., 1996; Castles & Coltheart, 1993; Murphy & Pollatsek, 1994; Stanovich et al., 1997). These forms are analogous to the surface and phonological subtypes of acquired dyslexia (Patterson, Marshall, & Coltheart, 1985; Beauvois & Derouesné, 1979). The signature deficit of the surface subtype is impaired reading of words with atypical spelling-sound correspondences ("exceptions" such as PINT and HAVE), whereas the signature deficit of the phonological subtype is impaired generalization (i.e., pronunciation of nonwords such as MAVE and GLORP). There is considerable controversy about the bases of these deficits, however. The standard interpretation is that these patterns reflect impairments to separate processing routines (the "routes" in the dual-route model, as in Castles & Coltheart, 1993). The "nonlexical" pronunciation mechanism uses rules to translate from spelling to sound. The "lexical" mechanism involves accessing a word's entry in an orthographic lexicon and using that to access its entry in a phonological lexicon. The dual-route theory holds that exception words can only be read by the lexical route, whereas nonwords can only be read by the rules. The two subtypes of developmental dyslexia (and their analogues in acquired dyslexia) are seen as deriving from selective damage to one or the other pronunciation mechanism: surface dyslexia involves an impairment to the lexical route, and phonological dyslexia the nonlexical route.

Our approach is different. We do not model word recognition and pronunciation in terms of different types of processing mechanisms that apply to different types of stimuli. Rather, our theory is stated in terms of computations involving different types of information (orthography, phonology, semantics). In this approach, all types of words and nonwords are processed in the same way: the presentation of an orthographic pattern as input initiates the spread of activation via weighted connections throughout the network. Below we show that, rather than deriving from damage to different types of naming mechanisms, the two subtypes of developmental dyslexia can be explained in terms of different *types of damage* to the lexical network. This account also explains additional facts about these patterns of developmental dyslexia, including the predominance of "mixed" cases in which both exception words and nonwords are affected. These simulations show how specific patterns of impaired reading can arise from specific types of phonological and non-phonological anomalies.

4. Effects of literacy on phonological representation. In the final simulations we address how the representation of phonological information may itself be affected by learning to read. Several studies have provided evidence that representations of phonology are altered by knowledge of alphabetic orthographies (e.g. Morais, Cary, Alegria, & Bertelson, 1979; Read, Yun-Fei, Hong-Yin, & Bao-Qing, 1987; Morais, Bertelson, Cary, & Alegria, 1986). The surprising implication of this work is that literate and illiterate individuals have somewhat different representations of the structure of spoken language. The models that we employed in our simulations allowed us to examine this issue because the weights on connections encoding phonological information were themselves allowed to change in the course of learning to read. The effects of literacy on phonological representation could then be assessed by comparing the representations before and after training on the reading task.

1. Acquiring Phonological Knowledge

Our first goal was to try to approximate the child's acquisition of phonological knowledge prior to learning to read. We constructed a model that, like the child, was exposed to phonological word forms and learned to represent them in memory. The manner in which the network represented this information allowed it to extract generalizations about the phonological structure of English; in particular, it learned about the structure of phonemic segments (i.e., that they consist of clusters of phonetic features) and about constraints on the sequences of phonemes (i.e., phonotactics). We present tests of the model that assessed what it had encoded about phonological structure. This phonological representation was taken to approximate a beginning reader's knowledge of phonology and was used in subsequent models that learned to map from orthography to phonology.

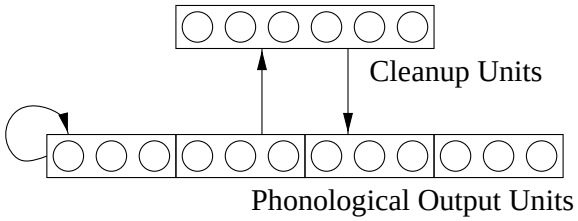

Figure 1. Architecture of the phonological attractor network.

Phonological Representation

The phonological representation scheme that we employed has two principal design features. First, it employs a distributed representation of phonemes in which units correspond to phonetic features. Second, this representation formed part of the larger network illustrated in Figure 1, in which all of the phonetic feature units were connected to each other and to a set of phonological cleanup units, analogous to the semantic cleanup units employed by Hinton and Shallice (1991) and Plaut and Shallice (1993). The representation is slot-based in the sense that the input vector corresponds to a sequence of phonemes, and so it inherits some of the known limitations of slot-based approaches (Plaut et al., 1996). For example, the /b/ phoneme in the initial consonant slot is represented separately from the /b/ in the final consonant slot, and therefore what is learned about the phoneme in one position does not automatically carry over to the same phoneme in another position. Plaut et al. (1996) termed this the *dispersion* problem. It makes the task of learning phonological representations more complicated, insofar as the model has to learn to represent several /b/'s rather than just one. However, this representation has other properties that are more important for our purposes. First, the interconnections between input units and the cleanup apparatus allow the network to encode dependencies across slots. The slot problem is more serious in a simple feedforward network in which these kinds of dependencies cannot be represented at all. Second, this representation allows the model to capture the fact that phonemes in different positions sometimes differ phonetically. The fact that word initial voiceless plosives (e.g., /p/,/t/) in English are generally aspirated provides a well-known example.[1] Finally, phonological codes were centered on the vowel. Vowels are the primary source of variability in phonology and in orthographic-phonological correspondences; thus, centering on the vowel minimized the dispersion problem for those aspects of the representation for which it is most salient. In summary, the present representation uses phonetic features in slots corresponding to phonemes, but minimized the dispersion problem by incorporating direct connections between input units, a set of cleanup units, and vowel-centering. Phonemes are represented by phonetic features but in contrast to standard distinctive feature matrices (such as Chomsky & Halle, 1968), the network can encode dependencies across features and segments.

Phonemes were represented using a vector of 11 real-valued units, each of which corresponded to a phonetic feature. The set of features and representations for individual phonemes were drawn from recent theorizing in phonetics and phonology (Gorecka, 1992; Steriade, 1993). Units could express values ranging between -1 to 1. Some features, such as *labial* and *pharyngeal* were binary, taking on values of -1 and 1. Others, such as *voice* were ternary, allowing values of -1, 0 and 1. The *sonorant* feature took values along a continuous gradient, representing an encoding of the sonority hierarchy. A monosyllable was represented as 6 of these phoneme slots, in CCVVCC formation. The monosyllable was vowel centered, with diphthongs occupying the middle two vowel slots. The second of the two vowel slots was coded as an empty phoneme (all features having a value of -1) if the vowel was not a diphthong. For example, the phonological form of the word BAT would be /_bæ_t_/ while the form of the word BLADE would be /blejd_/. Tables 1 and 2 summarize the phonemes and features used. The total representation for the phonological form of a monosyllable was 66 units, with 6 slots of 11 units defining a phoneme. Silent phonemes were coded by setting all features to -1.

We found 95 uninflected monosyllabic English words that could not be represented in this template, specifically those requiring 3 consonant phonemes before or after the vowel (e.g. STREET, WHILST). These words were excluded for purely pragmatic reasons. The addition of leading and trailing consonant slots would raise the number of phonological units from 66 to 88. This would increase the size of the phonological component from 6,996 weights to 11,264 weights, and significantly increase network training time. Given the large number of simulations described below, it

[1] Such allophonic variation was not utilized in the current study, but is something to be investigated in future research.

Table 1
Phonological Feature Representation: Consonants

Symbol	Sonorant	Consonantal	Voice	Nasal	Degree	Labial	Palatal	Pharyngeal	Round	Tongue	Radical	Example
/p/	-1	1	-1	-1	1	1	0	-1	1	0	0	*p*at
/b/	-1	1	0	-1	1	1	0	-1	1	0	0	*b*at
/t/	-1	1	-1	-1	1	-1	1	-1	-1	1	0	*t*op
/d/	-1	1	0	-1	1	-1	1	-1	-1	1	0	*d*og
/k/	-1	1	-1	-1	1	-1	-1	-1	-1	-1	0	*k*ite
/g/	-1	1	0	-1	1	-1	-1	-1	-1	-1	0	*g*ive
/f/	-0.5	1	-1	-1	0	-1	1	-1	1	0	0	*f*it
/v/	-0.5	1	0	-1	0	-1	1	-1	1	0	0	*v*ine
/θ/	-0.5	1	-1	-1	0	-1	1	-1	-1	0	0	wi*th*
/ð/	-0.5	1	0	-1	0	-1	1	-1	-1	0	0	*th*e
/s/	-0.5	1	-1	-1	0	-1	1	-1	-1	1	0	*s*it
/z/	-0.5	1	0	-1	0	-1	1	-1	-1	1	0	ja*zz*
/h/	-0.5	1	0	-1	0	-1	-1	1	-1	-1	-1	*h*at
/ʃ/	-0.5	1	-1	-1	0	-1	0	-1	-1	0	0	*sh*ot
/ʒ/	-0.5	1	0	-1	0	-1	0	-1	-1	0	0	bei*ge*
/ʧ/	-0.8	1	-1	-1	1	-1	0	-1	-1	0	0	ca*tch*
/ʤ/	-0.8	1	0	-1	1	-1	0	-1	-1	0	0	*g*in
/m/	0	0	1	1	1	1	0	-1	1	0	0	*m*op
/n/	0	0	1	1	1	-1	1	-1	-1	1	0	*n*ot
/ŋ/	0	0	1	1	1	-1	-1	-1	-1	-1	0	si*ng*
/r/	0.5	0	1	0	-1	-1	-1	1	1	-1	-1	*r*at
/l/	0.5	0	1	0	-1	-1	1	-1	-1	1	0	*l*oop
/w/	0.8	0	1	0	0	1	-1	-1	1	-1	0	*w*in
/j/	0.8	0	1	0	0	-1	0	-1	-1	0	1	*y*es

was felt that the cost in training time did not justify the minor benefit that representing the additional 95 words would yield. Below we describe a simulation using a much larger corpus of words which shows that adding these words does not create any other problems for our approach.

The phonological attractor network was created by connecting all feature units to each other and to a set of cleanup units (in effect a set of hidden units mediating the computation from the phoneme representation to itself). Including these connections allows the behavior of units to change over time; the phonological component becomes a dynamical system whose state can change itself. When trained appropriately, such systems can develop attractor states, or *basins of attraction* (Hinton & Shallice, 1991; Plaut & Shallice, 1993). Such basins can be thought of as a surface in state-space, such that states near a fixed attractor will be drawn into that attractor state. Ideally, the entire 66 dimensional state-space would be characterized by a landscape of overlapping attractor basins, such that any of the infinite number of states the network can find itself in will resolve to a phonemically and phonotactially legal end state over time.

The direct connections between phonological units allowed the encoding of some simple types of dependencies between phonetic features. For example, a given phoneme cannot be both consonantal and sonorant (see Tables 1 and 2); if consonantal is positive, sonorant must be negative, and vice versa. This constraint can be encoded by a negative weight from the consonantal feature within a phoneme to the sonorant feature, forcing them to have opposite signs if they are both nonzero. However, English phonology also exhibits more complex dependencies that cannot be represented by simple direct connections. For example, consider the relationships between the *degree* feature on the first two phonemes of a syllable. As summarized in Table 3, there is a constraint against both degree features being set to 1, although they can both be -1, or they can have opposite signs. This contingency cannot be encoded by direct connections between the two units and is the kind of phenomenon that motivates the use of networks with a layer of so-called *hidden units* (Rumelhart, Hinton, & Williams, 1986). When hidden units are utilized in an *auto-attractor*, meaning a set of units that map their activation from themselves onto themselves over time, they are called *cleanup units*, because they assist the units in "cleaning up" the output activation values (Plaut & Shallice, 1993), that is, coercing the patterns into a legal configuration.

In the phonological attractor network depicted in Figure 1, each unit's dynamics can be described as a nonlinear squashed sum of its input. The hyperbolic tangent activa-

Table 2
Phonological Feature Representation: Vowels

Symbol	Sonorant	Consonantal	Voice	Nasal	Degree	Labial	Palatal	Pharyngeal	Round	Tongue	Radical	Example
/i/	1	-1	1	0	0	-1	0	-1	-1	0	1	h*ee*d
/ɪ/	1	-1	1	0	0	-1	0	-1	-1	0	-1	h*i*d
/e/	1	-1	1	0	-1	-1	0	-1	-1	-1	1	p*ai*d
/ɛ/	1	-1	1	0	-1	-1	0	-1	-1	-1	-1	h*ea*d
/æ/	1	-1	1	0	-1	-1	0	1	-1	-1	1	h*a*t
/a/	1	-1	1	0	-1	-1	-1	1	-1	-1	-1	h*o*t
/ɔ/	1	-1	1	0	-1	-1	-1	-1	1	-1	-1	p*aw*
/o/	1	-1	1	0	-1	-1	-1	-1	1	-1	1	t*oa*d
/ʊ/	1	-1	1	0	0	-1	-1	-1	1	0	-1	t*oo*k
/u/	1	-1	1	0	0	-1	-1	-1	1	0	1	b*oo*t
/ʌ/	1	-1	1	0	-1	-1	-1	-1	-1	-1	-1	h*u*t

Table 3
Distributions of Degree Features in Consonant Slots Preceding the Vowel

Degree C1	Degree C2	Example	Legal
-1	-1	_RAT	Yes
-1	1	_PAT	Yes
1	-1	BRAT	Yes
1	1	BPAT	No

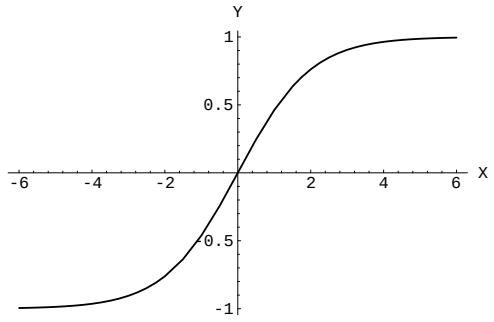

Figure 2. Activation curve for hyperbolic tangent function used in simulations: $y = \tanh \frac{x}{2}$.

tion function (hereafter 'tanh') was chosen, rather than the more traditional logistic function, because it has a number of properties that make it attractive for this application. The hyperbolic tangent (see Figure 2) has the familiar s-shape, but its input/output curve passes through the point (0,0). The more traditional logistic activation function passes through (0,0.5), meaning an input of 0 to a unit results in an output of 0.5. With the tanh function, in the absence of any input the units produce a zero output, allowing for ambient, inactive states. Further, the tanh activation function preserves the sign of the input: negative input means the output will be negative, and positive input means the output will be positive. This makes it easier to read weights as correlations between units. Each phonological unit has an auto-connection: a weight set to 0.75 and frozen to that value. The activation of the phonological units (positive or negative) tends to drop off towards zero over time, in the absence of any external driving input.

Training Corpus

A set of 3,123 monosyllabic words was chosen from various sources, including lists used in previous research and an online dictionary. Proper nouns and morphological variations on words (such as plurals, past tense, etc.) were excluded in order to keep the size of the training corpus manageable. As with the complex onset words described previously, the decision to exclude the inflected words was purely a pragmatic one related to computing time. On p. 14 we describe the results of an additional simulation demonstrating that the model can be trained on almost 8,000 monosyllabic words without creating any additional problems.

Each word was assigned a frequency derived from its frequency of presentation in the Wall Street Journal Corpus (Marcus, Santorini, & Marcinkiewicz, 1993). This corpus is much larger than the more commonly used Brown corpus and provides more robust frequency estimates. This frequency was transformed into a probability of presentation by a logarithmic transformation (see Plaut et al., 1996, for a discussion of log frequency compression):

$$p_i = \frac{\log((f_i/100)+1)}{\log m/100} \quad (1)$$

Here f_i is the frequency of word i, m is the frequency of the most frequent word in the corpus (THE, frequency 2.7 million occurrences). Words with a probability p_i less than

0.05 were set to 0.05. Log frequency was used to facilitate training; a much larger amount of training would be needed to give a reasonable coverage of all words. For example, in the Wall Street Journal corpus, the word THE occurs approximately 50 thousand times more often than the word ISLE, and 16 thousand times more often than the word CZAR. Using probabilities of presentation that are a linear function of word frequency would be computationally intractable with online learning. The total sum of the WSJ frequencies of the words used in the training corpus is approximately 21 million. If we wish an item with a count of 1, such as FILCH, to be 90% likely to appear at least once, we would need to sample approximately 50 million words. Using log compressed probabilities, the number of necessary samples drops to about 22 thousand.

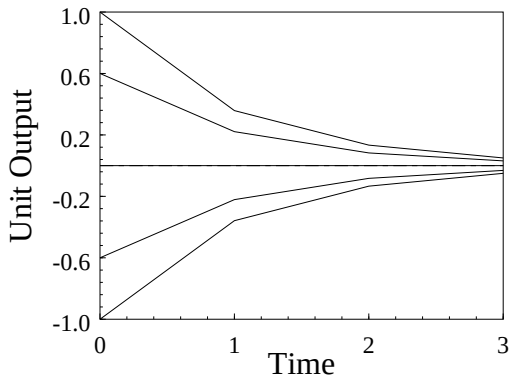

Figure 3. Dynamics of a unit with auto-decay. Positive and negative values decay to zero over time.

Training Method

We then trained this network on the phonological codes for the words in the corpus. The goal of training was for the network to develop a representation of the phonological structure of English monosyllables. In reality, children develop such representations in the course of learning to comprehend and produce spoken language, tasks we were not prepared to simulate. We therefore used a simplified procedure in which the model merely had to learn to represent and retain phonological codes over time. On each trial, the phonological representation of the word was clamped onto the phonological units. These units were given a hardwired tendency to decay their activation values over time. The network's task was to retain the input pattern despite the tendency for unit activations to decay. This task pressures the network to form dynamical attractors that embody the statistical regularities in the training set. Weight adjustment was performed on the difference between network output and the phonological form of the target word.

Initially, all weights in the phonological attractor network were assigned small random values between -0.1 and 0.1. The exceptions to this were the 66 connections from each of the phonological units to themselves, which were frozen at 0.75. Units therefore had a tendency to retain a fraction of their previous activation level, and to experience a gradual rather than immediate decline in their output level. Figure 3 shows the output of units over time with initial values of 1, 0.6, -0.6 and -1, assuming no other input to the units. The units' activation eventually drops off to zero.

The network was trained using the backpropagation through time training algorithm (Williams & Peng, 1990). The output of each unit at a given time is a function of the sum of its aggregate activation, according to the following formulas.

$$o_i^t = f(x_i^t) \quad (2)$$

$$x_i^t = \sum_{j \in U} w_{i,j} o_j^{t-1} \quad (3)$$

Here o_i^t denotes the output of unit i at time t, x_i^t refers to the input to that unit at time t, f is the activation ("squashing") function which maps the input of a unit to its output, U is the set of all units, and $w_{i,j}$ denotes the weight from unit j to i. Each unit, then, takes the weighted sum (weighted by w) of the outputs of other units on the previous time tick $t-1$, and this becomes the input x to that unit for tick t. The output of that unit, o is the result of applying the activation function f to the input value x.

In training, an error measure E was defined as in Equation 4, equal to the sum of squared differences between the output vector o and the target vector d, summed over all time ticks T and all units I according to the formula $E = \sum_i^I \sum_t^T (o_i^t - d_i^t)^2$. Minimizing the error E essentially means minimizing the *distance* between vector o and d, since E is the square of the euclidean distance between o and d.

The network was run for a preset number of time ticks, with each unit updating its activation at each time slice according to the weights and activations of all other units for the previous time slice according to Equations 2 and 3. Then the derivative of the error E with respect to each weight w in the network was computed, as per the standard backpropagation equations (Rumelhart et al., 1986). Each weight was then adjusted by changing its value according to the negative of the error derivative, multiplied by a small constant called a learning rate (denoted μ). Early piloting revealed that a learning rate $\mu = 0.001$ was appropriate, and this value was used throughout training of the phonological network.

The network was trained using online learning. During training, a "zero error radius" of 0.1 was used, meaning that errors less than 0.1 were counted as zero. The activation functions typically employed in connectionist networks cannot reach their extremal values except in the limit of an infinitely large input. The "zero error radius" is used to avoid overtraining the network, which can never exactly obtain the extremal values.

Words were sampled probabilistically from the training set according to their probability value p, as computed by Equation 1. On average, a word with $p = 0.5$ is selected by the network about twice as often as a word with $p = 0.25$.

To train the network, the following algorithm was used:

1. A word was sampled randomly from the training corpus according to the frequency distribution (Equation 1).

2. For time tick 0, the 66 phonological units were clamped with the appropriate values for that word.

3. The network was run for 4 ticks, with units unclamped.

4. For ticks 2-4, the output of each phonological unit was compared with the actual value of the word, and the difference was propagated backwards through the network, generating error gradients for each weight.

5. The weights were updated according to their error gradient.

6. Continue with step 1.

Because each unit had a positive auto connection weight, it tended to retain the sign of its initial value. Because the auto connection weights were frozen at a low enough value that the units' activation would drop off, each unit needed increased input activation from other units and from the cleanup units in order to reach the target output. Training was halted after a million trials, when it was observed that the sum squared error was not decreasing.

Results

Several tests were devised to assess the nature of the phonological representations that the model developed. The general strategy was to quantify the ability of the model to retain and repair degraded or incomplete phonological representations, and to characterize the attractor dynamics the model had formed. In later sections we will relate the quality of these attractors to specific phenomena observed in studies of speech perception, such as phonemic restoration effects and categorical perception of consonants. Because our focus is on reading, we have not exhaustively examined the model's capacity to simulate such speech perception effects or attempted to simulate a broad range of data. We can show, however, that the network encodes sufficient phonological information to produce several of the effects that have been observed in humans, subject to implementational limitations such as the restriction to monosyllables. These results suggest that it would be fruitful to further explore the relevance of this kind of architecture to speech perception phenomena.

Pattern Retention. The first method of assessing the phonological attractors involved observing how well the model performed on the task on which it was trained: retaining phonological patterns over time. A *nearest-neighbor measure* was used to assess the correctness of the phonological output. For each of the 6 output phoneme slots, the 11 features were matched against the set of existing phonemes. The phoneme that was closest in euclidean distance to the actual output was considered the output phoneme for that position. A word was scored as correct if all of the output phonemes were the correct ones.

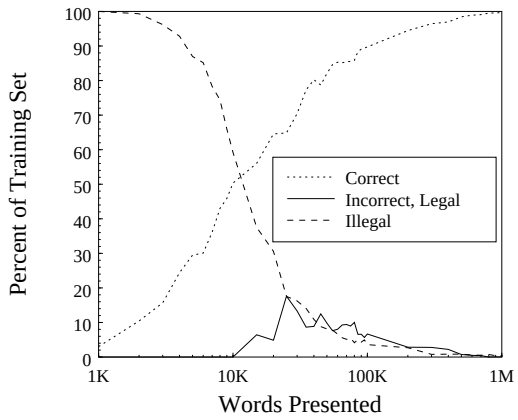

Figure 4. Pattern retention accuracy over the course of phonological training.

A second, more stringent measure was also used to identify *illegal* phonemes. The featural output of the model can be correct by the nearest neighbor measure but still not correspond to a legal combination of features (for example, an output that is featurally closest to /k/, but is not consonantal). An output was therefore scored as *legal* if and only if there exists a phoneme in which all 11 output features are within 0.5 of that phoneme's representation.

Figure 4 summarizes the evaluation of the training set over the course of training. At asymptote, just 11 items of the original 3123 were incorrect, and the number of illegal phonemes dropped to 10.

Pattern Completion. We next assessed the network using a pattern completion task. Information was deleted from an input pattern and we observed the extent to which the model could fill this gap given what it encoded about phonological structure. This test was conducted for all items in the training set. Each of the 66 features was taken in succession, and its value was left unspecified. All other features for the given word were clamped to their appropriate values. The network was allowed to run for 4 ticks, as it was during training. Then the difference between the unclamped feature's value and its target was measured and recorded. This was done for all 66 features over all 3123 words. This test assessed the capacity of the network to coerce an incomplete phonetic pattern into one that is phonemically legal within the target language.

The average magnitude of the error e_j for each feature j was computed over all 3,123 words in the training set: $e_j = \sum_{i=1}^{3123} |o(i)_j - d(i)_j| / 3123$. The results are shown in Figure 5 (top). In general, the error is quite low for most

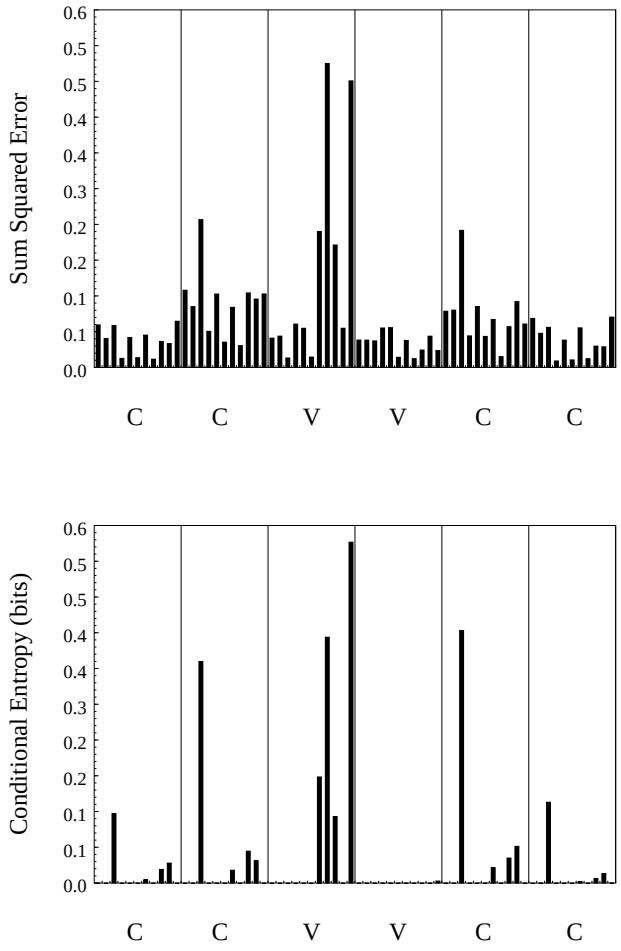

Figure 5. Top figure depicts the sum squared error on pattern completion task, by feature. Features in each phoneme slot, in order, are: sonorant, consonantal, voiced, nasal, degree, labial, palatal, pharyngeal, round, tongue, and radical. Bottom figure shows the conditional entropy of these features.

features, averaging < 0.1 for most items. Recall that during training a zero error radius of 0.1 was used, so it is not surprising that many values hover near 0.1; the network considers the effective, trainable error in such a case to be zero. Since the activations (and targets) for each unit range from -1 to 1, the range of possible squared errors is 0 to 4, so an error of 0.1 is quite low. Thus, for a majority of features, the network fills in the correct value based on the neighboring features to a high degree of accuracy.

A number of features do have high error values, however. The voiced features in the 2nd and 5th slot are high, as are a cluster of vowel features related to place and manner of articulation. The explanation for these effects can be seen by examining the featural representations of all the phonemes. The voiced feature is the only one that distinguishes /p/ from /b/; also /t/ from /d/, /k/ from /g/, and several other minimal pairs. Given a word form with the voiced bit unspecified, the network has no way to know what the correct value is. In short, the voiced feature is relatively unconstrained by its neighboring features. In contrast, the consonantal feature is totally constrained by the other features in a segment. It is therefore not surprising that errors are higher on the less constrained items.

To demonstrate this more rigorously, the *conditional entropy* for each feature was computed. The entropy $H(\mathcal{X})$ of a distribution $\mathcal{X}$ is defined as:

$$H(\mathcal{X}) = \sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{4}$$

See Cover and Thomas (1991) for derivation and discussion. The conditional entropy, or conditional uncertainty of distribution $\mathcal{X}$, given the environment $\mathcal{Y}$, is defined as

$$H(\mathcal{X}|\mathcal{Y}) = -\sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log_2 p(x|y) \tag{5}$$

For each feature in the phonological output array, the conditional entropy of the feature was computed relative to the values of the 10 other features in its phoneme slot. The

entropy was computed over all words in the training set, weighted by the frequency of each word.

Figure 5 (bottom) shows a plot of the conditional entropy, or uncertainty, of each unit in the phonological output in units of bits. As expected, the voiced features for the consonant slots show high uncertainty. The consonants in slot 1 and 6 show less uncertainty because they are empty more often than the inner consonants, and the features of empty slots are easily predictable from its environment.

Visually, the match between the mean sum squared error and the conditional entropy is quite good. A Pearson correlation over the two sets of numbers revealed a good match: $r = 0.88$, $t(64) = 15$, $p < 0.0001$. Thus, errors in the network approximate the conditional uncertainty in the training set. This result is not surprising given that the residual error of a network trained using the sum squared error measure approximates the conditional variance of the training set (Bishop, 1995).

We next examined the kinds of errors that were made. Given that the model's inaccuracy is a function of the uncertainty of a feature, is it the case that a different, but legal phoneme is always produced? The feature completion task was repeated as above, but the output was assessed according to whether the output phoneme was correct or incorrect. Incorrect patterns were further classified as either legal or illegal, using the method described in the preceding section.

Production of incorrect phonemes was infrequent, averaging 8.4% over all features and training set items. The production of an illegal phoneme occurred for 1.13% of the features and training set items, so 13.5% (1.13%/8.4%) of the incorrect items resulted in an illegal phoneme. The feature that produced the most illegal phonemes when left unspecified (14%) was the pharyngeal feature on the vowel phoneme. The conditional entropy of features was also correlated with the tendency for those features to produce illegal patterns when unspecified: $r = 0.72$, $t(64) = 8.3$, $p < 0.0001$.

Even in ambiguous environments (see Figure 5, bottom) the network was more likely than not to produce the correct word; the error rate never exceeded 50%. When the correct word was not produced, the next most likely outcome was creating a novel, legal word (e.g. PONE, when the target was BONE). The production of illegal patterns was largely limited to the place and manner of articulation of the vowel. These features are the ones that are the most underdetermined by their environments (Figure 5, bottom).

Attractor Basins. The phonological network has formed what are called attractor basins. The main idea is that if one considers the 66 phonological features as forming a high dimensional space, with each feature representing one dimension, the trained network consists of a set of attractor states in this space. A point in this space corresponding to the phonological form of a word will be subject to the network dynamics, and pulled in various directions within the space. Those forces have the effect of limiting the regions in which the network can settle, such that noisy or degraded input is very likely to be coerced into a coherent pattern.

To visualize the 66 dimensional space is of course impossible. However, it is possible to provide a small example that can be visualized. Suppose one only considers two features of the 66; the *round* and *radical* features, which will be manipulated with respect to the vowel. The word /kʌt/ provides an environment in which these two features can be manipulated. In the context of the word /kʌt/, the four extremal value combinations of these two features define four phonemes. Of those four, only 3 are legal phonemes in the /k_t/ context. A combination of -1 and -1 gives /kʌt/ (CUT), 1 and 1 gives /kot/ (COAT), 1 and -1 gives /kɔt/ (CAUGHT), and -1,1 is illegal.

To explore the attractors within the phonological space, all features for the word /kʌt/ were clamped to their actual values except the round and radical features of the vowel. These features were systematically set to values ranging from -1 to 1 in increments of 0.1, giving 400 initial states for a word. For each of these states, the network was run, and the two features in question were allowed to drift. For each of the 400 combinations of values, the states of the two features were sampled after running for two ticks. The direction in which each feature moved was recorded. This created a two dimensional vector field, as depicted in Figure 6 (arrow magnitudes are scaled by a constant for readability). The figure illustrates the direction that the phonological state moves when the rest of the form of the word is held constant. Figure 6 shows that the network pulls away from the illegal state of -1, -1 and toward a nearby legal configuration. The infinite number (subject to machine precision) of initial states defined by the plane in Figure 6 becomes coerced into only 3 final states: the three legal corners of the plane.

The attractor basins can be shown in three dimensions by computing for each point on the two dimensional lattice of Figure 6, the distance a point moves over 4 time ticks (by which time it has settled). Figure 7 depicts the attractor basins for the vector plot of Figure 6. The metaphor being used here is that a point in the dynamic space is like a ball that rolls along the surfaces formed by the attractors. Points initially at stable attractors (the three corners) do not change their state, and are hence shown at $z = 0$. Hilltops show the divisions between basins, or regions the points roll into: points near the COAT corner roll into that corner; points along the middle trough roll into the CAUGHT corner, while points near CUT roll there.

To examine the sensitivity of these attractors to the local environment, the experiment was repeated with the word /kæt/ used as the initial state. This is identical to the previous trial, except that the palatal feature of the vowel is set to 0 instead of -1. The effect of the different value of palatal is that suddenly rounding is illegal; in English, front vowels

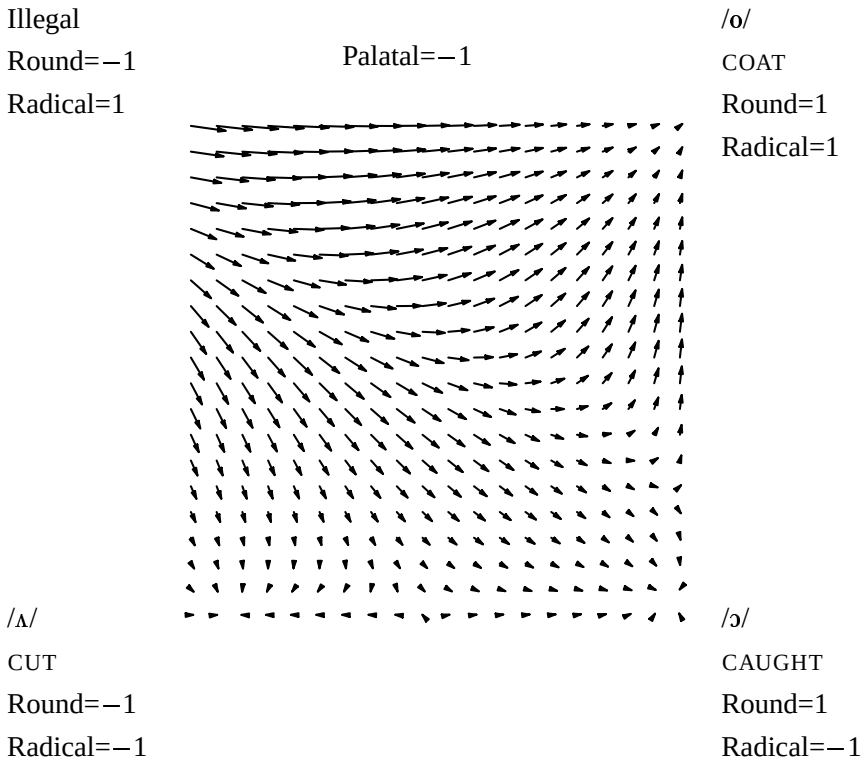

Figure 6. Phonological attractor, depicting 3 legal phonemes derived from combinations of two features: round and radical.

cannot be round (although there are languages in which this is allowed, such as German). When palatal is 0, rounding is -1 and radical is -1 the resulting phoneme is /æ/ (rhymes with CAT). For a configuration of 0, -1, 1 the result is /ɛ/ (rhymes with PET).

Figure 8 depicts the differing attractors for the same features (contrast with Figure 6). In this case, the network shuns any positive value on the round feature, correctly capturing the generalization that palatal=0 prohibits round=1.

The inputs to the round feature were examined to see how the network was able to capture this generalization. It turns out that the input to the rounding feature, averaged over all words in the corpus, is negative. This is largely due to the fact that rounding is off more often than on; its mean value over all words is -0.4, and the median is -1. In the environment of the word /kæt/, the input from all other phonological units and the cleanup units is -2. The weight from the palatal feature to the rounding unit is -2, such that when palatal is -1, it cancels the ambient disposition to turn off this feature. The network has thus implemented the following "rule": if palatal is 0, rounding must be -1. If palatal is -1, rounding can be either on or off. Importantly, the network has done so using "soft" attractors, so that intermediate values which the network never experiences in training still tend to get pulled into a legal state.

In summary, the network represents phonological knowledge in terms of attractors in state space. We have shown how this knowledge can be used to complete patterns that are missing a feature to repair or complete partial or noisy representations. We provide additional analyses of the model's phonological representations below. We now turn to a reading model that learns to map orthographic representations onto this structured phonological knowledge.

2. Learning to Read

The question addressed in this section concerns the role of prior phonological knowledge in learning to read aloud. Given the extensive behavioral evidence linking phonological representation, reading acquisition, and dyslexia, we expected that having this knowledge in place would facilitate learning the mapping between spelling and pronunciation. These simulations also provide a closer analogue to the child's experience than had earlier word reading models, permitting more valid comparisons between model and

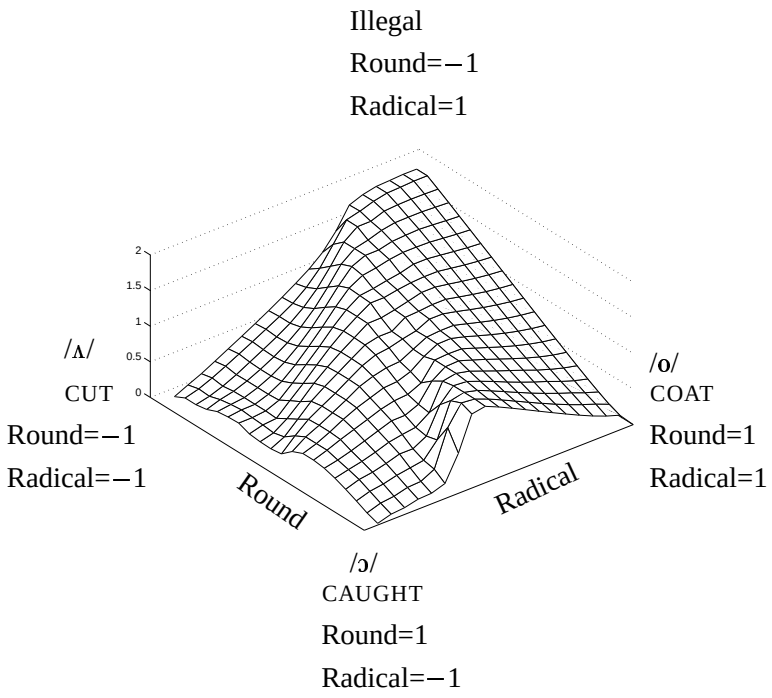

Figure 7. Attractor basins plotted spatially (see Figure 6). A point landing on this space would "roll" into one of the three legal attractor states. The 'z' axis measures distance a point travels in the dynamic space.

child performance.

The simulations reported below involve comparisons among three conditions. In the Trained Attractor condition, the weights that resulted from the pretraining procedure provided the initial state of the reading model. In the Untrained Attractor condition, the same network architecture and task were used, but the phonological attractor part of the network was initialized with small random weights. This model had the capacity to encode higher-order dependencies among features but unlike the Trained Attractor model, it did not have this knowledge in place at the start of learning to read. The third, Feedforward condition utilized a simple feedforward network; the connections between phonetic feature units and the cleanup apparatus were eliminated, leaving only connections from orthography to the hidden units, and from the hidden units to the phonological units. This model had a more limited capacity to represent dependencies among features. These conditions allowed us to examine the relative importance of having phonological knowledge in place prior to learning to read compared to simply having the capacity to learn and represent such knowledge in the course of learning to read. Based on previous findings we expected the feedforward network to perform more poorly, particularly on nonword generalization, because of its restricted capacity to represent phonological structure.

Architecture

The architecture used in the Trained and Untrained Attractor conditions is illustrated in Figure 9. The input layer was a set of 208 units representing the spellings of words. These were fully connected to an intermediate level of 100 hidden units, which in turn were fully connected to output representation, which was the phonological attractor net described above. In the Feedforward condition, the connections between phonological units and the cleanup units were eliminated. In each case, the model was trained to map the spelling of a word onto its pronunciation.

Eight slots of 26 units each were used to represent words up to 8 letters long, with each slot corresponding to a letter position and each unit representing one letter. Words were vowel-centered, with the first vowel of a word represented in slot 4, and any consonants growing outward from the vowel. The presence of a given letter within a slot was indicated by setting that unit to the value of 1 and all others to 0.

There is considerable inefficiency in this orthographic representation. For example, there are vowel units in the onset and coda positions that are always set to 0 because vow-

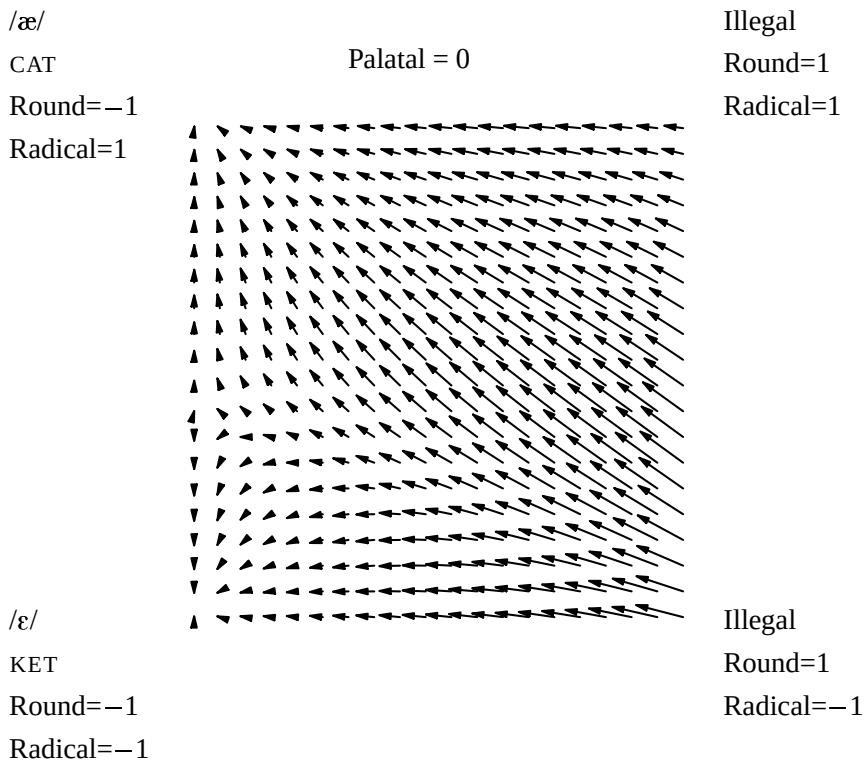

Figure 8. Phonological attractor field, in different phonological environment. Palatal is 0, yielding different dynamics than in Figure 6.

els cannot occur in these positions. Conversely, the consonant units in the vowel positions are always off. The representation was chosen to be as simple as possible, building in very little of the structure of English orthography beyond what was imposed by the basic architecture.

Training Method

The backpropagation through time algorithm was again used, with online learning and words selected for training using the same procedure as described in Section 1. The same training corpus used in the phonological training was used for reading training. After a word was chosen, the orthographic units were clamped with the pattern corresponding to the spelling of the word for 6 time ticks. Unit activations were updated for 6 time ticks. On the final tick the output of the phonological units was compared to the word's phonological target. As is standard for the BPTT algorithm, the discrepancy between the output and the target over each output unit is injected into the output units of the network. This error for each output unit is then used to compute the error for all units which are connected to the output unit: the error for a hidden unit, example, is a function of its contribution to the error for the output units. The error that each hidden unit accumulates is then used in the same way to determine the error on all input units. Similarly, the cleanup units in the phonological attractor network accumulate error based on the error of the output units they are connected to. In this sense, "blame" for the overall error is propagated backward through the network from the output units. Once "blame" is assigned for each unit, weights can be updated by changing them slightly in the direction that would reduce the error. The errors are then discarded, and the cycle repeats with the selection of a new word. The effect of this training procedure is that the weights come to take on values that minimize the error for each word in the training set. Regular, rule governed items exert a similar effect on the weights to the extent that their targets and inputs are similar, while exceptions pull the weights in a different direction. For example, the weights from the orthographic rime of words like GAVE, BRAVE and SAVE all have a similar phonological rime target, so their influence on the values of the weights is similar. The patterns of activity over the hidden units that these words create will have some similarities, due to their overlapping spellings, and some differ-

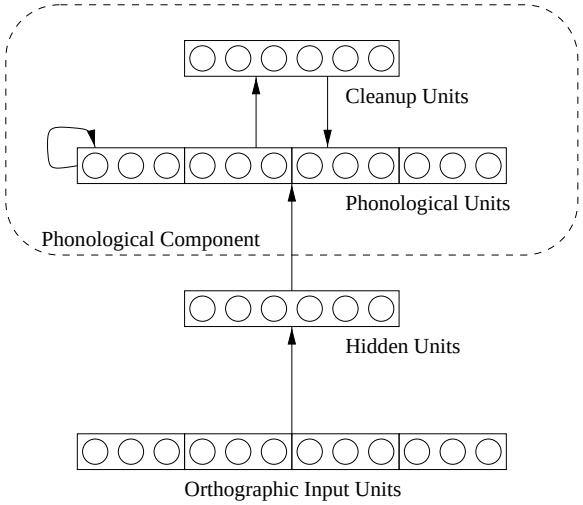

Figure 9. The architecture of the reading model.

ences, due to their differing onsets. The similarities in the hidden unit activities, coupled with similarities in the output targets, produce rule-like behavior. For the exception word HAVE, in contrast, the network must acquire sensitivity to the presence of the H in the environment of AVE in order to override the default behavior created by the regular neighborhood.

Piloting revealed that a learning rate μ of 0.005 was appropriate for the reading components of the model (the connections from the orthographic to hidden units, and from the hidden units onto the phonological representation). Lower values resulted in much longer training times, and higher values led to instabilities in the network. This value is higher than the value that was used in the phonological attractor (0.001). Initial studies used a learning rate of 0.001 throughout the model, and while taking longer to train (on the order of several days per run), these simulations exhibited qualitatively similar performance to ones with a higher learning rate.

One other important feature of the training procedure was the interleaving of two types of training trials. In the phonological acquisition phase described above, the model was trained on a phonological retention task. We now want the model to learn a second task, mapping from orthography to phonology. However, we also want the model to retain its knowledge of phonological structure. Training the model exclusively on the reading task will result in weights that are optimal for this task but not necessarily for the phonological retention task. Blocked training on different tasks is the condition that gives rise to what has been termed "catastrophic interference" (McCloskey & Cohen, 1989). Under these conditions, training on the second task results in a failure to retain all of what was learned in connection with the first task. The solution to this problem is simply to interleave training on the two tasks (see Hetherington & Seidenberg, 1989). Forgetting on the first task is avoided if there are a few additional trials of this type during training on the second task. This interleaving of tasks is also more realistic with respect to the child's experience, which is not strictly blocked by task. The child acquires extensive phonological knowledge through exposure to spoken language; however, their experience with speech does not end once reading instruction begins.

Thus, training on the reading task was interleaved with additional trials on the phonological retention task. We will refer to the latter task as the "listening" task because it involves encoding and retaining a phonological pattern. On each training trial, a random number was computed. Based on this random number, that training cycle was either a reading or listening cycle. On 80% of the trials, the model was trained on the reading task; on 20% the listening task.[2] On reading trials, the network was trained as described above. On listening trials, the model was trained as in Section 1. Hence, the phonological weights had to assume values that would facilitate performance on both tasks.

Length of Training. Eight simulation runs, representing different subjects, were conducted for each of the three conditions. On each run the weights from orthographic to hidden units and from hidden units to phonological units were randomized to values between -0.1 and 0.1. For the Trained Attractor condition, the phonological network was trained as in Section 2 for each of the eight runs. The Untrained Attractor condition had random weights assigned to the attractor network, while the Feedforward network had no attractor network. In all networks, the initial weights and the exact order of presentations of the words were determined by the initial random number seed. Within each training condition, all of the eight runs were identical to each other except for the initial seed. Results presented below, unless otherwise noted, represent the mean performance of the 8 networks within each condition.

For each run, the Trained Attractor network was presented with 1 million words during the phonological training phase, as in Section 1. It was then trained on 10 million words during the reading phase. Since 80% of the trials during the reading phase were reading examples and 20% were interleaved listening examples, each run exposed the network to a total of 3 million listening trials and 8 million reading trials. The Untrained Attractor network was not

[2]This ratio of reading to listening trials is higher than would be experienced by a child. We ended up with this ratio after explorations of a variety of ratios yielded two findings: first, using 20% listening trials was sufficient to prevent significant unlearning of phonology; second, using a larger percentage of listening trials had little effect on mastering either the listening or reading tasks. We therefore used the 80-20 ratio in order to keep the training times relatively low. Essentially the same results obtain if the proportion of listening trials is increased, but the network takes longer to learn.

subjected to phonological pretraining; it simply was trained on 10 million words in the reading phase, with the above distributions of reading and listening trials. The Feedforward network was trained for 10 million reading trials.

We used a large number of training trials during each phase (listening and reading) in order to be able to examine asymptotic levels of performance. As will be seen below, the resulting learning curves were highly nonlinear, with rapid learning during the first million or so word presentations and slower learning thereafter.

Results

We first describe the performance of the Trained Attractor nets, and then provide comparisons to the Untrained Attractor and Feedforward conditions.

Word Performance. The output that each model produced for words in the training set was evaluated using two criteria. First, each phoneme in the word was assessed using the nearest neighbor test described previously. The phoneme in the training inventory that was closest, by euclidean distance, to the output of the network was determined for each position. These phonemes were then compared to the target phonemes. Phonemes were judged correct if they were identical to the target or if they were members of predefined equivalence classes of phonemes. The rationale underlying the equivalence classes is that there are some variations in the production of certain phonemes that participants could produce but would not be detectable. These classes were /ɔ/ and /a/ (e.g. COT and CAUGHT), /ow/ and /o_/ (e.g. the difference in some dialects of English between the vowels of DOE, which can have a trailing /w/ sound, and DOME, which does not), and /ej/ and /e_/ (the later being a shorter, more clipped version of /ej/). All phonemes in a word (or nonword) had to be closest to the target ones for the item to be scored as correct.

In addition, the features for each output phoneme had to correspond to a legal phoneme in the training set; that is, there had to be a phoneme in the training set in which each output feature was within 0.5 of the value for that phoneme. The nearest neighbor criterion affords the possibility that the computed output might be closer to the correct phoneme than any other even though the particular combination of features does not correspond to any phoneme. Imposing the second scoring criterion ensured that such trials would be scored as incorrect.

Figure 10 (left) shows the mean performance of each of the three networks on the training set items over time. There is virtually no difference between the Trained and Untrained Attractor networks. The Feedforward network performs a bit more poorly, reaching the same asymptotic level of performance as the attractor networks, but more slowly. At asymptote, the trained attractor networks averaged 98% correct on the training set. The majority of errors were on low frequency exception words (e.g., CHOIR) or orthographically unusual low frequency words (e.g., MYRRH). These results are in accord with the results presented by Seidenberg and McClelland (1989) and the behavioral data reported by Waters and Seidenberg (1985) and others.

Nonword Generalization. The models' capacities to generalize to novel items were assessed using a set of 364 nonwords. The 86 pseudowords from Glushko (1979), Experiment 1, and 156 of the 160 items from McCann and Besner (1987)[3] were combined to form 239 items (three less than the sum of the two sets, because three items were duplicated in the Glushko and McCann & Besner studies). An additional 125 items were generated by taking existing word bodies and replacing the onset to form a nonword (as in Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994).[4] For most items, there was only one correct pronunciation allowed when scoring the network's output. For some items, if the network produced one of two possible outputs it was scored as correct (e.g. the nonword DOMB could be pronounced either /dam/ as in BOMB or /dom/ as in COMB). This scoring is consistent with behavioral data reported by Seidenberg et al. (1994), who found that the two most common pronunciations of the more than 500 nonwords in their study accounted for over 90% of participants' responses. The development of nonword generalization performance is summarized in Figure 10 (right). At asymptote, the Trained Attractor networks scored an average of 79% of the nonword set correct, as measured by the stringent criterion and 88% correct by the more lax nearest-neighbor measure. Performance of the Trained and Untrained Attractor networks was again quite similar, with only slightly poorer performance on the Untrained Attractor networks early in training. However, the Feedforward networks exhibited much worse performance on the nonword set throughout training.

Replication with a Larger Corpus. As noted earlier, the training corpus used in these simulations included most but not all monosyllabic words in English. Two classes of items were excluded: ones that did not fit the CCVVCC phonological template (e.g., STRING) and inflected words (plurals such as BOOKS and tensed verbs such as BAKES and

[3]Four items (BINJE, FAIJE, JINJE, WAIJE) were excluded because they contained the letter J two positions after the vowel. This letter in this slot is not seen in the training set. As such, the model could never get these items correct. This problem reflects an inherent limitation of the slot based representation scheme (see PMSP for discussion). This problem would not arise if the model had been trained on polysyllabic words which provide coverage of this gap (e.g., BANJO, CONJURE). These words were the only ones from the McCann and Besner and Glushko studies that exhibited this problem.

[4]The entire set of nonwords is available at http://siva.usc.edu/~mharm/papers /dyslex.psyrev /nonword.stim.pdf.

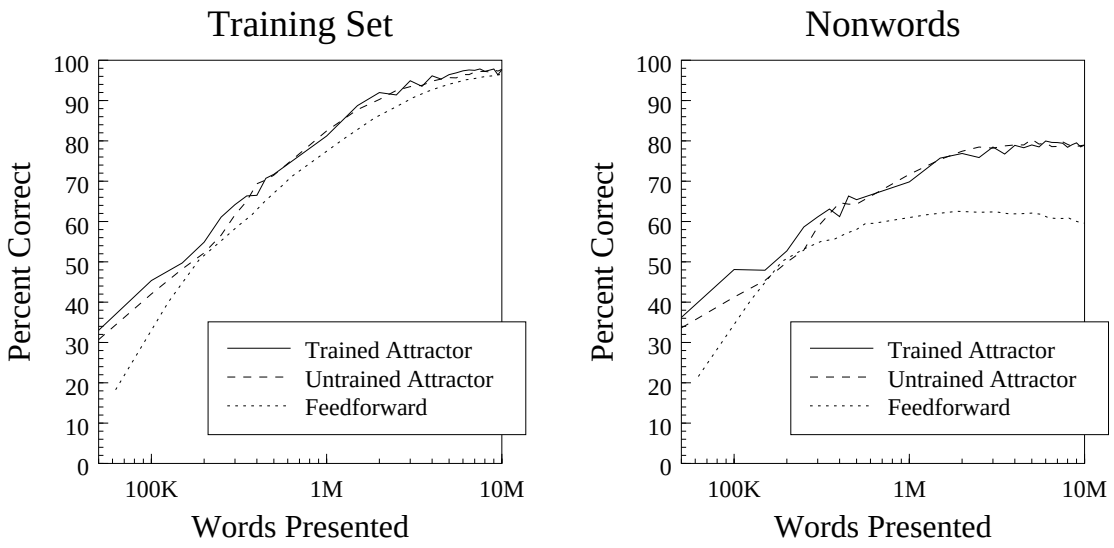

Figure 10. Comparison between the normal reading model, the feedforward model, and the initially untrained attractor model on the training set (left) and nonword generalization (right).

BAKED). These items were excluded for a pragmatic reason, the need to keep network training times within reasonable bounds. However, the exclusion of these words raises questions about the generality of the results we have presented. One question is whether similar levels of performance can be achieved with a larger number of words to be learned. A second question is whether words with complex onsets (such STRING) or codas (such as BURST) present any special challenges. Finally, the properties of English inflectional morphology create complex orthographic-phonological mappings for these words. In both the plural and past tense, the phonological realization of the inflection is conditioned by the previous phoneme. In both BUDS and BOOKS, for example, the plural morpheme is spelled with S. However, whether its pronunciation is voiced (as in BUDS) or unvoiced (as in BOOKS) depends on the voicing of the preceding phoneme which is itself inconsistently cued by the orthography. Thus both MOUTH and TENTH end in TH but differ in voicing; although both form the plural by adding S in the orthography, the inflections are pronounced differently. The mappings between spelling and pronunciation for these words are therefore rather complex. In summary, the words we excluded differ in some respects from the words in the training corpus and raise additional challenges for our approach that need to be addressed.

To explore these questions, we conducted a replication simulation using a much larger corpus. Monosyllables were extracted from the CELEX electronic corpus (Baayen, Piepenbrock, & van Rijn, 1993). All items fitting a CC-CVVCCC template were used, yielding 7,839 words. Most of the additional words are inflected items. The phonological network was expanded from 66 to 88 units to accommodate the larger template, and additional orthographic units were added to fit longer words. Aside from these changes, no other alterations were made to the model's architecture, representations or training regime.

After 10 million trials, the model had correctly learned 99% of the training set, as scored by the strict criterion. Nonword generalization improved: the model correctly pronounced 84% of the benchmark nonword set by the strict criterion, and 90% by the more lax one. The original model had difficulty pronouncing some nonwords that had few neighbors in the original training set; for example, it produced errors on nonwords that look like plurals, such as SNOCKS and PHOCKS. The larger model, which contains many plurals, has no difficulty with these items. These results demonstrate that increasing the size of the training set not only does not create problems for the model, it facilitates performance on nonwords by providing broader coverage of the space of orthographic and phonological patterns.

Performance on words with complex onsets or codas and on inflected words is summarized in Table 4. For comparison we also examined the performance of a strictly feedforward network on these items. Both attractor and feedforward models achieved high levels of performance on these words, with a slight advantage for the former. The models' capacities to generalize were examined by testing them on nonwords with plural or past tense inflections. Here the attractor network performed significantly better than the feedforward network. These results are consistent with the conclusion that the attractor structure is relevant to learning complex spelling-sound mappings; however, the learning of the larger corpus proceeded without complication. The larger model does take significantly longer to

train, however, and so the smaller model and corpus were used in subsequent simulations.

Discussion: The Reading Model

The simulations show that the central findings from SM89 and PMSP replicate using an output representation that is an attractor network employing phonetic features. There were two important other findings. First, the Attractor networks yielded better performance than the Feedforward network, but the advantage was almost entirely specific to nonword generalization. This result is consistent with the earlier finding that the SM89 model performed better on words in the training set than on generalization. If the task is merely to learn the pronunciations of a set of words, a feedforward network is sufficient (cf. SM89). However, being able to pronounce nonwords requires the capacity to combine sublexical orthographic-phonological units in novel ways. Achieving human-like performance on this task requires having a more structured, componential representation of phonological information and the correspondences between orthography and phonology. This additional capacity can be achieved in two ways. One is by building additional structure into the orthographic or phonological representations themselves. That was the path taken by PMSP, whose phonological representation relied on an extrinsically-determined ordering of the phonemes. If this knowledge is not built into the representation, then the network architecture itself must allow it to be encoded. This capacity was provided by the attractor architecture explored here.

The second important finding was that although the attractor architecture was necessary for achieving adequate nonword performance, there was little difference between the Untrained and Pretrained conditions. The Pretrained model learned slightly faster but both models rapidly converged on very similar levels of performance. Thus, it was not necessary to have knowledge of phonological structure in place prior to training on the reading task because this information could be rapidly picked up in the course of training on this task. The model's architecture must allow phonological structure to be represented in a componential way but given the high degree of redundancy exhibited by natural language phonology, this information can be acquired at the same time as knowledge of orthographic-phonological correspondences.

The fact that the Pretrained and Untrained networks yielded similar performance contrasts with results that we presented in Harm, Altmann, and Seidenberg (1994). That study also examined the effects of prior phonological knowledge on acquisition of spelling-sound knowledge and found that pretraining on phonology yielded a significant improvement in performance. However, in that study we did not interleave reading and listening trials. The Pretrained model performed better because there was less unlearning of phonological structure during the reading task. In the simulations reported above, this advantage for the pretrained model was obviated by the inclusion of the interleaved listening trials.

In summary, these simulations suggest that having an architecture that permits the encoding of dependencies among features and phonemes is critical to achieving a high level of reading skill, particularly the capacity to generalize to novel instances. Children normally acquire this knowledge in the course of learning to use spoken language and so it is in place prior to the onset of reading instruction. This is like the Pretrained Attractor condition in our simulations. It is interesting that the Untrained Attractor condition yielded similar performance to the Pretrained condition, insofar as it suggests that having the capacity to represent certain types of knowledge is more important than actually having this knowledge in place prior to reading. However, the Untrained condition does not correspond to anything that could occur in reality; the situation in which the child is not exposed to phonological information until reading instruction begins never occurs. The Untrained condition is informative because it suggests that phonological structure can be rapidly learned from examples but it is not analogous to anything in a child's experience.

It should be apparent that the training procedure that we used only captured some very general characteristics of the child's actual experience in learning to read. These simulations, like earlier ones, used a procedure in which words were probabilistically selected for training based on their frequencies of occurrence in adult English. Reading instruction is quite different: children initially learn to read small vocabularies of words that expand over time. Moreover, children in at least some classrooms (e.g., ones emphasizing "phonics" methods) are provided with additional training that emphasizes similarities between words with respect to subword units such as onsets and rimes. They also receive explicit training in the pronunciation of particular letters and letter combinations. None of this is incorporated in the much simpler method we use to train our models. Two points should be noted. First, there is nothing in our approach that precludes structuring the training procedure in more realistic ways; indeed, the models provide an interesting way to examine whether particular ways of introducing words to children would yield more rapid acquisition. This is an important area for future research, one that could provide insights that would be useful for educators who plan instructional curricula. Second, the method that we used in training the models is probably not the optimal one (for children or for models). Teachers presumably structure reading instruction in specific ways because it facilitates learning and we would expect the same thing to occur in our models. Being more realistic about training would allow more detailed comparisons to children in the earliest stages of learning to read. It would probably allow our models to learn more efficiently as well.

Table 4
Larger Training Corpus Results

				% Correct	
Word Class	Type	Example	N	Attractor	Feed Forward
Plural	/s/	CATS	798	99	95
	/z/	BUGS	1045	98	96
Past Tense	/d/	BUGGED	724	98	90
	/t/	BAKED	599	99	94
Third Person Singular	/s/	BAKES	229	99	99
	/z/	BEGS	307	99	98
Complex	Onset	STREET	211	100	97
	Coda	WORLD	704	98	85
Nonword Plural	/s/	BAIPS	20	95	75
	/z/	GLEWS	20	95	90
Nonword Past Tense	/d/	POVED	20	85	60
	/t/	BAXED	20	90	50

3. Developmental Dyslexia

Developmental dyslexia is the failure to acquire age-appropriate reading skills despite adequate intelligence and opportunity to learn. Whereas acquired forms of dyslexia are observed in premorbidly literate adults following brain injury, developmental dyslexia is observed in children learning to read, apparently as a consequence of congenital anomalies. Our goal is to explain different patterns of dyslexic behavior in terms of different types of impairments to the simulation model that affect the course of acquisition. We also attempt to link these forms of simulated impairment to evidence concerning possible constitutional or experiential factors that limit children's performance.

The causes of developmental dyslexia have been the subject of considerable debate extending over many years. In the recent past, attention has primarily focused on impairments in the representation and use of phonological information as the proximal cause (see Liberman & Shankweiler, 1985; Adams, 1990; Farmer & Klein, 1996, for reviews). Learning to read involves learning how written symbols represent the sounds of language. Children who can sound out words (either overtly or covertly) can then match them to words known from speech, providing a kind of self-teaching mechanism (Jorm & Share, 1983). The training procedure in our networks approximates this self-teaching process: the network generates a phonological code for a letter string and it is then compared to the veridical phonological code that provides the basis for calculating the error used to adjust the weights. On trials when the child has correctly sounded out a word, this teaching feedback is self-generated by comparing the computed code to a word known from speech. Translation from orthography to phonology also permits the child to recognize words that have not been seen before and to learn the pronunciations of new words.

There is strong evidence that individual differences in the representation and use of phonological information are related to level of reading achievement (Bradley & Bryant, 1978, 1983; Mann, 1984; Lundberg, Olofsson, & Wall, 1980; Wagner & Torgesen, 1987). Pre-readers who have developed more segmental representations of phonological structure, as revealed by "phonological awareness" tasks such as repeating a spoken word with a single phoneme deleted, show higher levels of reading ability in later grades (Share, Jorm, Maclean, & Matthews, 1984; Adams, 1990). Impairments in the development of such segmental representations might then be the cause of dyslexia in at least some children.

This account, which is widely accepted among reading researchers, leaves two important issues unresolved. First, what is the nature of the deficit that gives rise to impaired phonemic representations? Many studies have established that dyslexia is associated with poor performance on tasks that require manipulating phonemic representations in working memory. A few studies have attempted to establish causal links between poor phonological representations and impaired reading acquisition (e.g. Bradley & Bryant, 1983). However, the nature of the information processing deficit that gives rise to phonological impairments is not clear. Attention has recently focused on the hypothesis that impairments in phonological representation are secondary to "temporal" processing deficits (Tallal, 1980; Tallal et al., 1996; Merzenich et al., 1996). The processing of speech involves perceiving small differences among rapidly changing signals. Tallal and others have provided evidence that the capacity to process brief and/or rapidly changing acoustic stimuli is impaired in some children. Tallal's hypothesis has generated considerable interest but is also controversial and the focus of ongoing research. One problem is that the exact nature of the "temporal processing deficit" is unclear; many studies providing evidence for such a deficit utilized complex tasks that involved both perceptual and memory components, making it difficult to determine what kind of

impairment led to poor performance. A related question is whether the deficit is specific to speech or reflects a more general problem that also occurs in other modalities (Di Lollo, Hanson, & McIntyre, 1983; Chase & Jenner, 1993; Galaburda & Livingstone, 1993). Finally, the evidence for deficits in speech perception is strongest in children who exhibit broader impairments in the use of spoken language (Joanisse, Manis, Keating, & Seidenberg, 1998).

The second important question is, how does an impairment in phonological representation interfere with learning to read? Phonological dyslexics are not equally impaired in all aspects of word reading. Previous theories have not explained how deficits in phonological representation give rise to the specific patterns of behavioral impairment that are observed in these children. Why are certain aspects of reading affected and not others?

We addressed these issues by introducing impairments in the representation and processing of information in the phonological attractor. The principal result was that the main impact of these impairments was on generating phonological codes for unfamiliar letter strings (nonwords). This is important because impaired nonword reading is the signature deficit in the behavioral pattern termed developmental phonological dyslexia (Temple & Marshall, 1983; Castles & Coltheart, 1993). Thus the model provides a computational link between phonological impairments and specific aspects of dyslexic reading. The simulations also provide some suggestive leads about possible bases for phonological impairments and whether these impairments will also affect speech perception. One puzzle about the phonological deficit hypothesis is that many dyslexics who perform poorly on "phonological awareness" tasks appear to have normal speech perception and production. It is not clear why a phonological impairment would not affect the use of spoken language as well. Our simulations suggest that a phonological impairment that is not severe enough to interfere with basic aspects of speech perception can nonetheless have a significant impact on reading acquisition. With a more severe phonological impairment, performance on both reading and speech perception tasks is affected.

A second type of dyslexia. Although the evidence that phonological information plays important roles in learning to read, skilled reading, and dyslexia is compelling, several recent studies have converged on the conclusion that some reading impairments are not caused by phonological deficits (Castles & Coltheart, 1993; Murphy & Pollatsek, 1994; Manis et al., 1996; Stanovich et al., 1997). These studies identified a subgroup of dyslexics whose word recognition was significantly below age-appropriate levels but whose performance on nonword reading was not. Castles and Coltheart (1993) and Manis et al. (1996) referred to these children as "surface dyslexics." This term was originally applied to cases of acquired dyslexia in which the patient is more impaired on reading exception words than nonwords (e.g. Patterson et al., 1985). The term was extended to the developmental surface dyslexics in the above two studies because they too were more impaired in reading exceptions than nonwords. Thus, phonological and surface dyslexia are complementary patterns in which either exception word or nonword reading is more impaired. This double dissociation is classically interpreted with the dual-route model as arising from separate impairments to the lexical or nonlexical route.

Differences between the subtypes of dyslexic children are illustrated by the summary data from the Manis et al. study presented in Figure 11. Dyslexic participants (mean age 12.4) who were reading at about the 4th grade level were compared to groups of same-aged and younger normal readers. Dyslexics were identified as surface or phonological dyslexic on the basis of discrepancies between exception and nonword reading, using the following procedure. Levels of exception word and nonword reading were observed in samples of same-aged normal readers. Surface dyslexic participants (N = 15) were defined as those children whose exception word reading was lower than expected given their level of nonword reading, based on the regression of nonword scores on exception word scores for the normal readers. Conversely, phonological dyslexic participants (N = 17) were children whose nonword reading was lower than expected given their level of exception word reading. The surface dyslexics' performance closely matched that of younger normal readers in terms of overall level of performance and both groups read exception words more poorly than simple nonwords. Phonological dyslexics' performance was quite different. Their level of exception word reading was like that of younger normals but they were much worse at reading nonwords. Thus, although both surface and phonological dyslexics performed more poorly than same-aged normal readers on both exceptions and nonwords, the surface participants were relatively more impaired on exception words and the phonological dyslexics on nonwords.

The term "surface dyslexia" is not very informative about the nature of these childrens' impairment or its underlying cause. Such children have been labelled surface dyslexic because they are more impaired on exceptions than nonwords. However, this focus on impaired exception word reading overlooks two prominent aspects of their behavior. First, most of these children are impaired on both exceptions and nonwords compared to normal readers of the same age. Although surface dyslexia is often described as a "selective" impairment in exception word reading, these children's impairment is typically not limited to this type of word. Second, the reading performance of the "surface dyslexic" children in both the Manis et al. and Stanovich et al. (1997) studies was indistinguishable from that of younger normal readers, whereas the performance of the phonological dyslexics was quite different. Beginning readers,

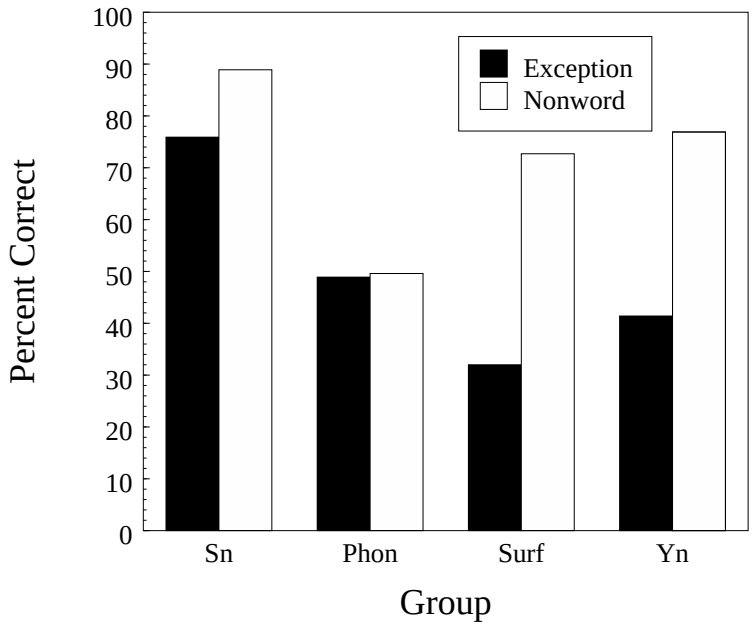

Figure 11. Data from Manis et al. (1996), both dyslexic groups exhibited impairment on both exceptions and nonwords, but phonological dyslexics showed a greater impairment on nonword performance, and the surface dyslexics showed a greater impairment on exception word performance. Phon = phonological dyslexics; Surf = surface dyslexics, Sn = same-aged normals, Yn = younger normals.

like "surface" dyslexics, are poorer at reading exception words than sounding out nonwords. Insofar as their performance on both types of stimuli quantitatively matches that of younger normals, the surface dyslexics can be said to be developmentally delayed. Because they exhibit a general developmental reading delay rather than a specific impairment in reading exceptions, we suggest that the term "reading delayed" is more accurate than "surface dyslexic" and we will use it throughout the remainder of this article except when referring to earlier studies or specific claims of the dual-route model. In contrast, phonological dyslexics exhibit a pattern of performance that is not seen in good readers at any age. In particular, their nonword reading is extremely poor given their level of exception word reading.

Manis et al. (1996) provided additional evidence that these are distinct subtypes of dyslexia with different causes. This evidence derived from performance on two validation tasks, phoneme position analysis and orthographic choice. The former involves repeating a word or nonword and identifying the position of a phoneme (e.g., "what sound comes before the /t/ in /skwʌpt/"). The latter involves identifying the correct spelling of a word, with a pseudohomophone as foil (e.g., RANE vs. RAIN). The phonological and surface dyslexics also exhibited a double dissociation on these tasks: phonological dyslexics were impaired on phoneme position analysis but not orthographic choice and surface dyslexics performed in the opposite way. The studies by Stanovich et al. (1997) and Murphy and Pollatsek (1994) yielded similar results. Taken together, these data strongly suggest that there are two distinct patterns of impaired reading with different causes.

Manis et al. (1996) also examined variability among the individuals within each group in order to determine whether any of the participants exhibited truly selective impairment in the reading of exceptions or nonwords. The dual-route model attributes the surface and phonological subtypes to impaired acquisition of the lexical and nonlexical reading mechanisms, respectively. The model therefore predicts that there could be children who are normal in reading one type of letter string and impaired on the other, that is, "pure" cases with truly selective impairments rather than "mixed" patterns in which both exceptions and nonwords are impaired, but one more than the other.

In general, the participants identified as phonological dyslexics in the Manis et al. study were impaired on both nonword and exception word reading compared to same-aged normal readers. Defining a "pure" case of phonological dyslexia as one in which the participant's performance on exception words was within a standard deviation of normal age matched children, but nonword performance was one standard deviation or more below that of the normal children resulted in the identification of 5 pure phonological dyslexics (out of 17). Their data, along with means for the same aged normal controls, are shown in Table 5. These 5 participants were among the least impaired of the phonological dyslexic participants and scored relatively well on

Table 5
The "Pure" Phonological Dyslexics from the Manis et al. Study

Participant	Exception	Nonword	Phonological Test
122	76	65	83
124	67	72	96
138	69	67	83
147	72	72	92
514	65	54	95
Mean SN	75(12)	89(9)	87(10)
Mean Phon Dys	49(15)	49(10)	63(20)

Note. Values shown are percent correct. Standard deviations are shown in parenthesis. SN = same aged normal participants; Phon Dys = phonological dyslexic participants.

Table 6
The "Pure" Surface Dyslexics from the Manis et al. Study

Participant	Exception	Nonword
101	47	80
139	25	80
148	47	84
149	40	83
1506	47	83
Mean SN	75(12)	89(9)
Mean Surf Dys	32(11.6)	72.7(10.4)

Note. Values shown are percent correct. Standard deviations are shown in parenthesis. Of the 5 pure cases, all but 1 were among the least impaired of the surface dyslexics.

the phoneme deletion test.

Five "pure" surface dyslexics (out of 15) were found in the Manis et al. study; these children were within 1 standard deviation on nonword reading but at least 1 standard devition below normal on exceptions. Their data are summarized in Table 6. All of the "pure" cases were less impaired in nonword reading than the average of the surface dyslexic group, and all but one were less impaired on exception word reading than the group. Thus, in a study of 51 dyslexic children, 10 were identified as "pure" subtype cases; these were mildly impaired children who happened to fall just below the cutoff criterion in reading one type of stimulus but not the other.

To summarize, the Manis et al. study and other recent research suggests that there are at least two forms of developmental dyslexia. In the phonological subtype, performance is below age-expected levels on both exceptions and nonwords, but is worse on nonwords; this impairment appears to be secondary to deficits related to the representation or processing of phonological information. The performance of the children in this group does not resemble that of younger normal readers because their level of nonword reading is so poor given their level of exception word reading. In the reading delay subtype, performance is also below age-expected levels on both exceptions and nonwords, but is relatively worse on exceptions. Their overall performance pattern closely resembles that of younger normal readers. Truly selective impairments in which the child is impaired in reading one type of letter string but normal on the other are rare and tend to be associated with mild deficits. We next consider how current models of word recognition can account for these data.

Theories of Developmental Dyslexia

The dual-route model takes the surface and phonological subtypes as strong evidence for the two naming mechanisms that it entails. The lexical route is the only one that can generate pronunciations of exception words. The nonlexical route is the only one that can generate pronunciations of nonwords. Therefore, an impairment in the lexical route will yield poor exception word performance but leave nonword naming unaffected. An impairment in the nonlexical route will have the opposite effect.

The model that we have been developing since Seidenberg and McClelland (1989) does not explain these patterns in terms of damage to separate subsystems that process exception words and nonwords because there are none: words and nonwords are processed using the same units and weighted connections. Our approach is instead to view these patterns as the result of different types of damage to this system. By hypothesis, one type of damage has a bigger impact on exception word reading and another on nonwords. Below we present simulations exhibiting these effects. This approach to explaining the impairments is consistent with the idea that they arise from different types of neurobiological anomalies, but does not require the auxiliary assumption that there are separate lexical and nonlexical routes.

Our account of these phenomena differs from the dual-route theory in four major respects.

1. The nature of the impairment in phonological dyslexia. Coltheart's view is that it derives from impaired acquisition of the grapheme-phoneme correspondence rules that are the basis of the nonlexical route. Because these rules are not adequately mastered, the child mispronounces nonwords. This account ignores the fact that such children exhibit broader impairments in the representation and use of phonology. They perform poorly on many non-reading tasks that involve the use of phonology, including the phonological awareness tasks that have been widely studied. The dual-route framework treats these impairments as essentially unrelated; children who are impaired in learning grapheme-phoneme correspondences happen to also have additional impairments in the representation and use of phonology. In our approach, the two deficits are causally related: phonological dyslexia derives from an impairment in the representation of phonological information. This impairment affects performance on tasks involving the use of

this information, one of which is reading acquisition, especially nonword reading.

The hypothesis that phonological impairments rather than impaired rule-learning underlie phonological dyslexia derives from two sources. First, as mentioned above, there is a vast developmental literature relating phonological impairments to reading difficulties (e.g., Shankweiler & Liberman, 1989; Olson et al., 1989; Share, 1995; Wagner & Torgesen, 1987; Tunmer & Nesdale, 1985). Second, the hypothesis is consistent with a body of computational modeling work. One important source was Besner, Twilley, McCann, and Seergobin's (1990) observation that because the SM89 model performed relatively poorly on nonwords its performance was like that of a phonological dyslexic. Plaut et al. (1996) demonstrated that improving the model's phonological representation yielded better generalization. Harm and Seidenberg (1996) presented preliminary simulation results showing that degrading a model's capacity to encode phonological regularities led to poor generalization. Similar results were reported by Brown (1997), who compared a small scale model using SM89 representations to one using PMSP representations. Consistent with the earlier work, he found that the network with SM89-style phonological representations performed more poorly on nonwords. These results led Brown to propose that impoverished phonological representations are implicated in developmental dyslexia.

The present study advances this work by showing that these results obtain with more realistic phonological representations, by introducing a phonological attractor in which phonological structure is learned, by providing computational analyses of why phonological representation is related to generalization, by relating the phonological impairments we introduce into the model to various deficits in phonological awareness and speech perception seen in some dyslexic children, and by relating the behavior of the model to behavioral data on children's reading performance.

2. The nature of impairment in the surface (or "reading delay") subtype. Castles and Coltheart's (1993) view emphasizes the impairment in exception word reading seen in these children but this is only one part of their behavior. The broader picture is that they are developmentally delayed with respect to reading, which yields impaired performance on all types of stimuli not just exceptions. Whereas Coltheart and colleagues explain this pattern in terms of impaired use of the lexical route, we view it in terms of factors that cause this type of general developmental delay.

3. Accounts of selective and mixed patterns. We see it as a problem for the Coltheart et al. (1993) theory that some kinds of selective impairments predicted by the dual-route model have not been observed. In principle, the dual-route model affords the possibility that exception word and nonword reading could completely dissociate, with perfect performance on one and nil performance on the other, a pattern that has not yet been observed. In the Castles and Coltheart (1993) and Manis et al. (1996) studies, there were a small number of children categorized as "pure" surface or phonological dyslexics. However, Manis et al. found that there was a strong relationship to degree of reading impairment: the "pure" children were mild cases. Below we show how this pattern arises in our model. The dual-route model however, predicts a much broader range of dissociations than have been observed.[5] The fact that most dyslexics are impaired on both exceptions and nonwords also presents a problem for the dual-route model. These stimuli are handled by separate mechanisms and so the observed pattern can only be explained by assuming that in most cases both routes happened to develop anomalously. Why both routes should routinely be impaired together is unclear and there is certainly no independent evidence (e.g., from neurobiology or neuroimaging) that this is so. Our theory provides a simple account of the mixed cases: because a single mechanism is used to generate pronunciations for all letter strings, a given type of developmental anomaly will tend to affect both exceptions and nonwords, though not necessarily to the same degree. Phonological impairment has a bigger effect on nonwords than exceptions; with only a very mild impairment, nonword performance can fall slightly below age-expected levels while exception word reading is within normal limits. With a more severe impairment, both nonwords and exceptions begin to be affected; the impairment continues to have a bigger impact on nonwords but performance on both types of items falls below age-expected levels. The opposite effects are seen in the surface/delay type of pattern. The types of impairments we explore below have these effects.

4. Relationships to normal reading. Finally, there are differences in how the two subgroups of dyslexics compare to normals. Whereas the surface dyslexics' performance is very much like that of younger normal readers, the phonological dyslexics' performance is not. The dual-mechanism theory offers no explanation for these different patterns. In contrast, we show why one pattern of impairment tends create behavior that looks like younger normal reading whereas the other does not.

---

[5] Castles and Coltheart (1996) present a case study of a 10 year old child (MI) whose pronunciation accuracy was very low on exception words but within normal on regular words and nonwords. However, MI cannot be taken as providing evidence for a normal nonlexical route with a selective impairment in the lexical route because his reading of even the items that he pronounced correctly was highly atypical. His reading was very effortful and he often took several seconds to sound out a word, methodically working through letter by letter before producing the final pronunciation (e.g., "c..o..n..t..ex..t... context!") (Castles, personal communication). Rather than a selective impairment in the lexical route, MI is better characterized as a severe mixed case who utilized an off-line compensatory strategy.

Phonological Dyslexia Simulation

We simulated phonological dyslexia by impairing the representation of phonological information before training the model to read. We then examined how reading acquisition proceeded under these conditions. The capacity of the network to represent phonological structure was impaired in two ways differing in severity. The mild form of damage involved imposing a degree of weight decay on the weights in the phonological attractor. With this method, each weight w in the attractor is reduced in magnitude according to the formula $\Delta w = -w\rho$ where ρ is the decay constant. Weight decay places bounds on the magnitudes that trained weights can acquire, thereby reducing the depth of the phonological attractors. Pilot studies revealed that a slight degree of weight decay (using a decay constant $\rho = 0.00005$) had a small effect on the ability of the network to perform phonological tasks such as pattern completion. This condition will therefore be referred to as the *mild* phonological impairment. A more severe form of damage was also explored, which involved removing the phonological cleanup units, the continued imposition of weight decay, and the severing of a random 50% of the connections between the phonological units. This has a larger effect on the ability of the network to encode phonological dependencies and will be referred to as the *moderate* phonological impairment. After presenting these results, we will describe a third simulation with an even more severe phonological impairment.

The effect of differing levels of phonological impairment on the pattern completion task (see p. 7) was tested. The mean sum squared error over all simulations and all features was measured. The normal model produced mean error of 0.08, the mildly impaired model 0.16, and the more severely impaired model 0.38 ($F(2,21) = 1829$, $p < 0.001$). Thus, as damage increases, the ability of the network to perform the task is gradually degraded. This result indicates that the two types of anomalies did degrade the representation of phonological information. We now consider how reading acquisition proceeds with these types of impairments in place. To compare the mild and moderate phonological impairment conditions to the normal model discussed in the previous section, eight simulations of each condition were run. As with the normal model, a different random number seed was chosen for each simulation run.

Figure 12 shows the results for the mild phonological impairment condition. Whereas performance on exception words is essentially unaffected, nonword performance is significantly impaired. This pattern corresponds to a "pure" phonological dyslexic. The pure cases observed by Manis et al. (1996) were also only mildly impaired. Figure 13 shows the performance of the model with moderate phonological damage. Nonword performance declines further and exceptions also begin to be affected, yielding the "mixed" pattern. The difference in mean exception word performance between the normal and severely impaired models was largest, 19.2%, at 500,000 trials.

More Severe Impairments. The phonological impairments presented above involved reducing the capacity of the phonological attractor in some way, either through weight decay or weight decay conjoined with lesioning. These methods produce the correct patterns of results; small levels of impairment produce a pure case, and greater impairments produce a mixed case. For the mixed case, however, the level exception word impairment was smaller than observed in the Manis et al. study. This is potentially problematic; if phonological damage cannot produce impairments in reading as large as are seen in the behavioral literature, it would undermine the phonological impairment hypothesis.

There is a limit on the degree of impairment that can be produced by simply removing connections in the phonological attractor; in the limit, with all connections severed, performance would be identical to the Feedforward simulation discussed earlier. However, a more serious impairment was created by making the computations within the phonological attractor more noisy. Formally, at each time slice of processing, the effective weight $w'_{i,j}$ was derived from the weight $w_{i,j}$ according to the formula $w'_{i,j} = w_{i,j}(1.0 + \sigma p(t))$, where σ is a free parameter, and $p(t)$ is a gaussian distributed random variable with standard deviation σ that was varied across simulations. $\sigma = 0.1$ yielded performance very similar to the moderate phonological impairment condition discussed above, while $\sigma = 0.2$ resulted in extremely impaired learning (exceptions and nonwords never scoring better than 20% correct).

Figure 14 shows the developmental curves for the normal and severely impaired models. The normal data are the average of 8 simulations and the impaired data the average of simulations using sigmas of .115, .125, and .15. Noise corruption is clearly capable of producing large impairments in both exception and nonword performance. A value of $\sigma = 0.15$ produced performance that is 50% lower than the normal model on exceptions, and 51% lower on nonwords. The simulations illustrate the continuity between degree of phonological impairment and level of reading performance: the larger the noise corruption parameter σ, the worse the exception and nonword reading. As in the mild and moderate simulations, phonological impairment primarily affects nonwords, with exceptions implicated at more severe levels.

Table 7 shows the results of the simulations with mild phonological damage (weight decay), moderate damage (lesioning), and various values of σ. All networks were evaluated at 1.5 million trials. The table also provides data from individual participants in the Manis et al. study. The participants' performance varied considerably; one was classified as normal, two were moderately impaired, and two more severely. Each simulation creates a deficit pattern that is closely matched by a participant in the behavioral study.

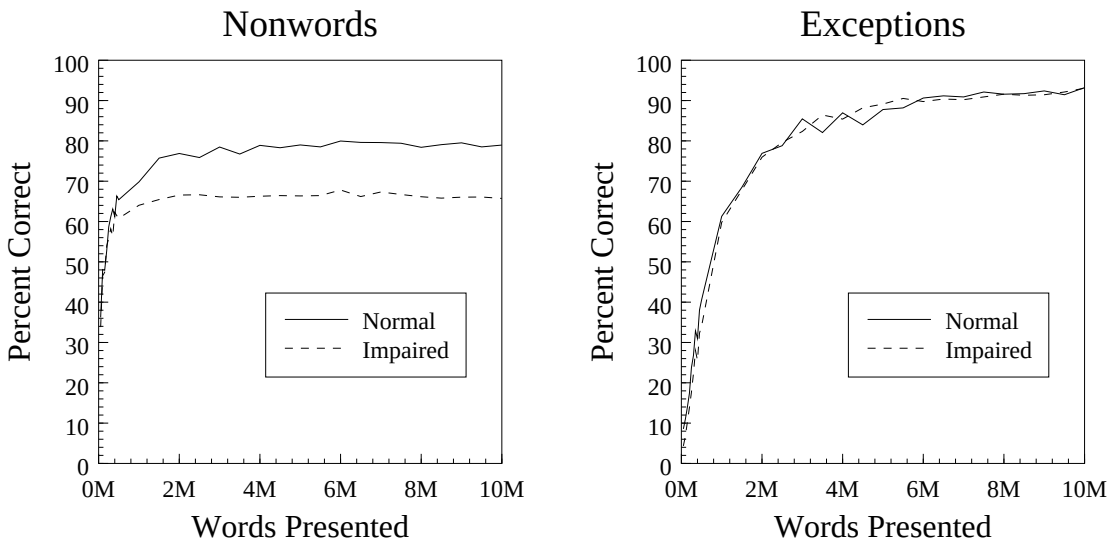

Figure 12. Impairment of nonwords (left) in the presence of mild phonological damage. No effect is seen on exception word reading (right).

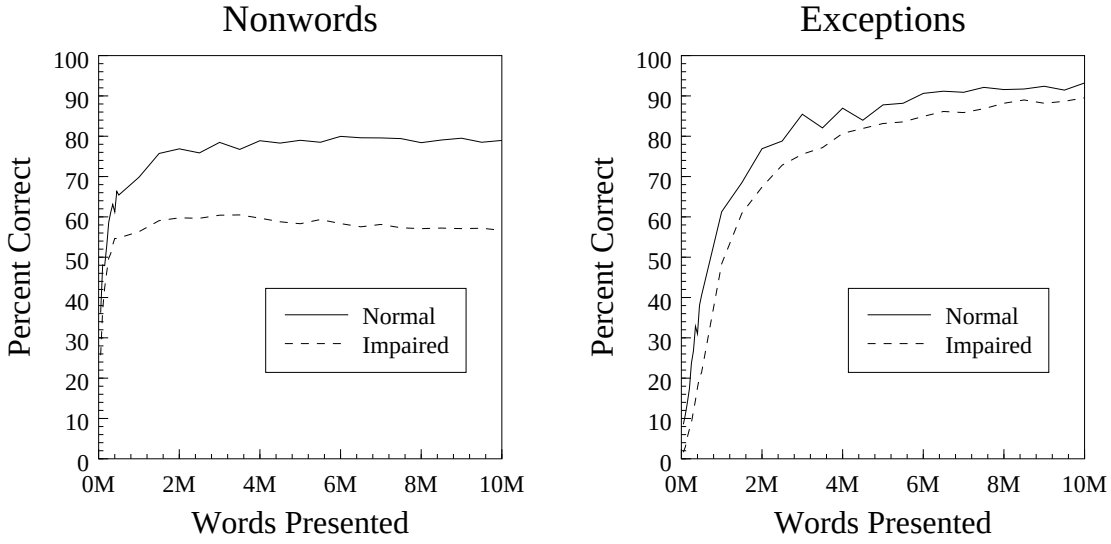

Figure 13. Greater impairment to nonwords resulting from moderate phonological impairment (left). Exception words also begin to be affected (right).

The Computational Basis of Nonword Impairments. Both the behavioral and simulation data suggest that phonological impairments have their main impact on nonword generalization. We now use analyses of the network to get at why this outcome obtains. The basic insight is this: Our model contains a phonological attractor structure whose function is to complete, clean up or repair incomplete or noisy phonological patterns, using the knowledge of phonological structure that is represented in these weights. Having this structure in place in the normal model affects what is learned in the weights mediating the computation from orthography to phonology. Specifically, with the clean-up apparatus in place, the mapping through the hidden units can be relatively imprecise: the output from the hidden units only has to be exact enough for the clean-up apparatus to resolve into the correct pattern. This imprecision turns out to be relevant to nonword generalization. Without the clean-up apparatus, the mapping from orthography to phonology must be more precise; the model can learn the mappings for words in the training set but generalizes poorly. In short,

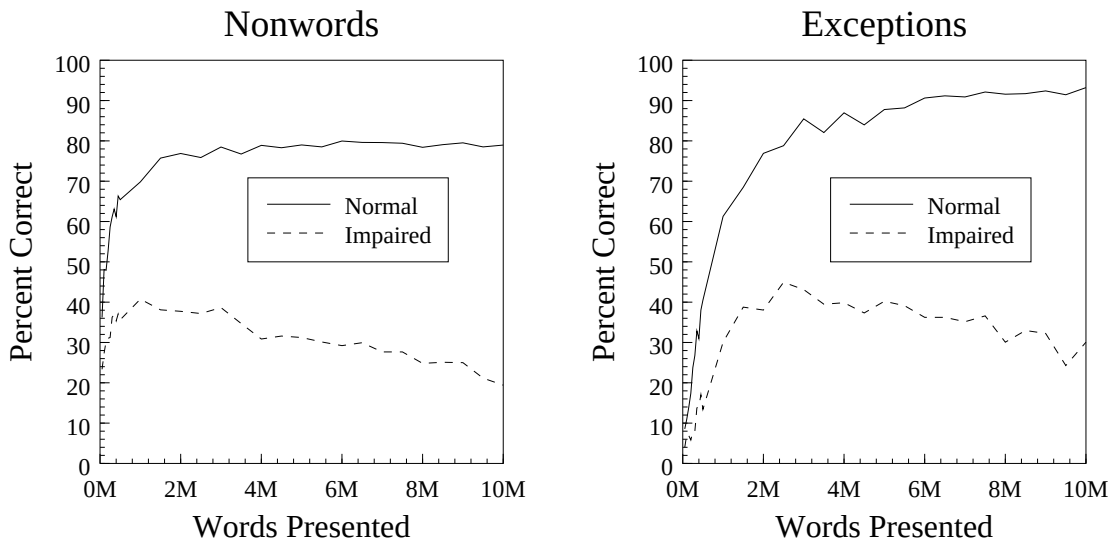

Figure 14. Invasive phonological impairment simulation. Nonwords (left) and exception words (right) are both severely impaired, and cannot recover.

Table 7
Phonological impairment, compared with participants from Manis et al. (1996)

	Score		Level of Impairment	
	Exc	NW	Exc	NW
Participant 76†	76	83	+1	-6
Model, Mild Impair.	68	65	0	-10
Participant 44†	67	73	-8	-16
Model, Moderate Impair.	60	58	-8	-17
Participant 138‡	69	67	-6	-22
Model, $\sigma = 0.115$	42	52	-11	-22
Participant 151‡	54	56	-21	-33
Model, $\sigma = 0.125$	49	41	-19	-34
Participant 319‡	32	37	-43	-52
Model, $\sigma = 0.15$	18	24	-50	-51

Note. †Normal Participant. ‡Phonological Dyslexic Participant. Participant level of impairment is the difference between participant's score and the mean for the same-aged normal controls (exc: 75%, nw: 89%). Model level of impairment is the difference between the model's score and the mean for the normal model (exc: 68%, nw: 75%). All model measurements were taken at 1.5M iterations.

the clean-up apparatus discourages overfitting the training data.

To demonstrate these effects more clearly, we examined the phonological output computed on the basis of input from the hidden units in the absence of any input from the phonological attractor network. This was accomplished by taking trained networks and then performing the test with the phonological weights removed. The output of these networks were compared the computed output for each word to the veridical pattern. This yields an index of the precision of the mapping performed by the hidden units. The normal model's error was 36.7, the mildly impaired model's error was 14.9, and the more severely impaired model's error was 8.4 ($F(2,21) = 1357$, $p < 0.001$). Thus, the output of the hidden units is less precise for the normal attractor model than in the two phonologically impaired conditions.

We have hypothesized that the effect of requiring the hidden units to perform a more exact computation is overfitting of the training data, which interferes with generalization. In order to examine this, we constructed a test that was performed on both the normal and moderately phonologically impaired networks. The inconsistent neighborhood -EAT was examined. This neighborhood has a large number of rule governed pronunciations (e.g., EAT, MEAT, BEAT, SEAT, TREAT) and several exceptions (e.g., GREAT, THREAT, SWEAT). A typical run of the normal model was used, which scored correctly on all words and nonwords containing this word body. We also examined a run of the phonologically impaired simulation that produced correct output for all of the words with this body but an error on a similar nonword GEAT.

One of the features that distinguishes the /i/ phoneme

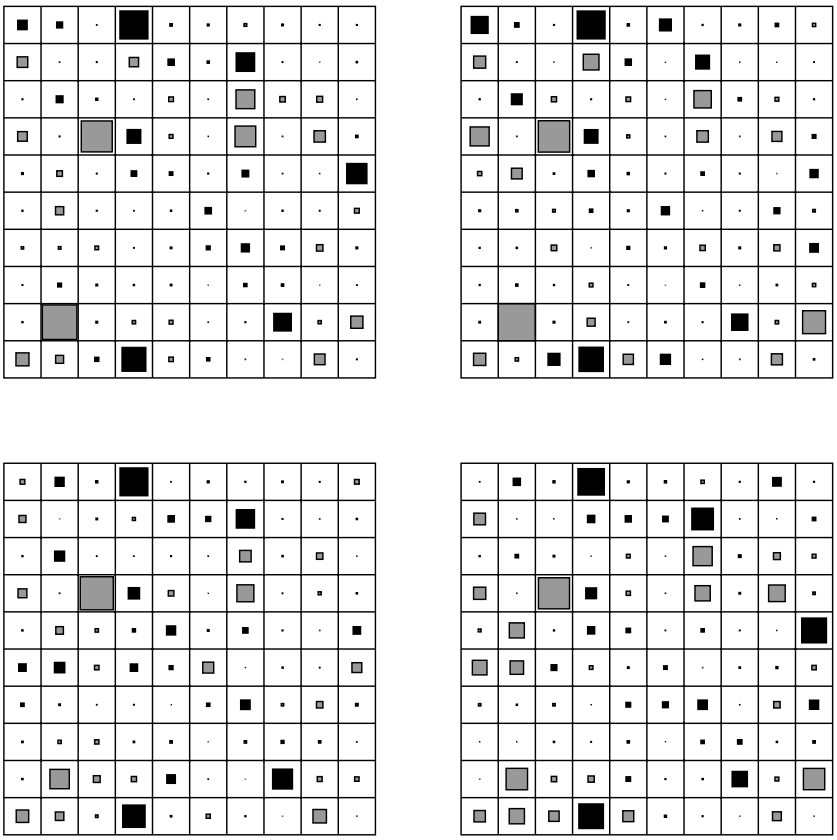

Figure 15. Normal model hidden unit contribution to *tongue* feature for words (top left to bottom right) EAT, MEAT, TREAT and the nonword GEAT. Positive values are shown in black, negative values are grey. Each cell's shaded area depicts its magnitude from 0.0 to 0.5.

from /e/ is *tongue*. By looking at the hidden unit contributions to this feature over a set of informative words, we can begin to see what it is about the normal network that affords generalization, and what about the phonologically impaired network that prevents it. Hinton diagrams were used to visualize varying contributions of the hidden units. Each hidden unit is connected to the *tongue* feature by a variable weight; the activity that each hidden unit contributes to that feature is the product of the hidden unit's output and that weight. In the figures that follow, the product of each hidden unit's activity and the weight connecting that unit to the *tongue* feature is plotted. Figure 15 shows the contributions of the hidden units to the *tongue* feature for the words EAT, MEAT, TREAT and the nonword GEAT. Figure 16 shows the corresponding plots for the network with moderate phonological impairment. A scale of -0.5 to 0.5 was used for all Hinton graphs, with the size of the shaded box representing the ratio of the value's magnitude from 0.0 to ± 0.5. A value > 0.5 or < -0.5 resulted in a solid cell in the plot. Positive values are shown in black, negative values in grey.

Figure 15 shows that the hidden unit activities for the various words are quite similar. The hidden units are all receiving different activation from the orthographic onsets of the words, and yet for the purposes of the *tongue* feature on the vowel the words are behaving essentially alike. Contrast this with Figure 16, activations for the same words in the impaired model. The figure shows that the phonological impairment results in many more units making larger contributions, both positive and negative, to the tongue feature. It also shows that the words are more different from each other and from the nonword GEAT than in the normal condition.

Borrowing a technique from neuroimaging studies, we used a subtractive method to highlight these effects more clearly. The hidden unit activation "images" for the normal model's representation of EAT and MEAT were subtracted, as were EAT and TREAT, and MEAT and TREAT. The average absolute value of these differences differences for the normal and impaired models are plotted in Figure 17. For the normal model, the differences are small because it represents the three words very similarly. In contrast, the differences are larger for the impaired model because there is less overlap in its representations of the three words.

A similar subtractive procedure was performed to assess differences between the models' representations of the nonword GEAT and the three rhyming words. The difference

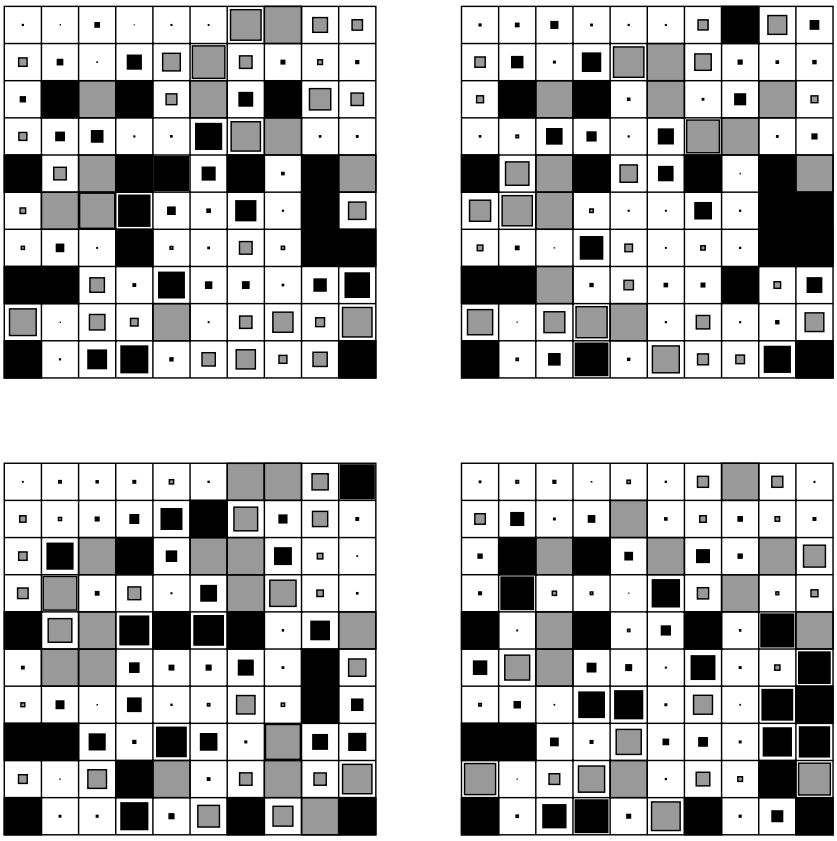

Figure 16. Moderately phonologically impaired model's hidden unit contribution to *tongue* feature for (top left to bottom right) EAT, MEAT, TREAT and GEAT. Each cell's shaded area depicts its magnitude from 0.0 to 0.5.

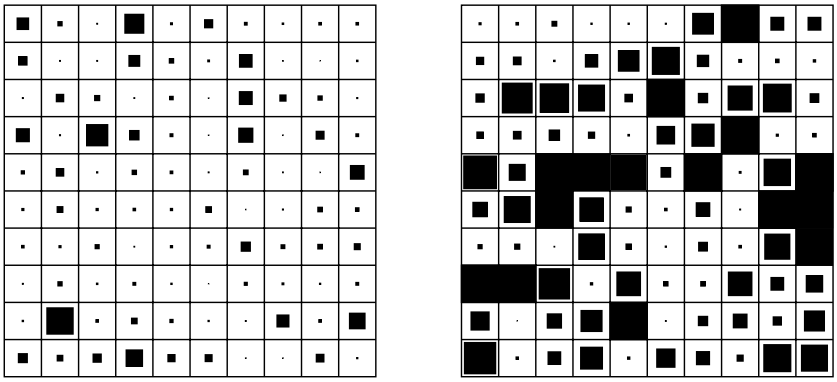

Figure 17. Mean differences between hidden unit activity for words MEAT, TREAT, and EAT for normal (left) and phonologically impaired network (right). The normal model shows small differences compared to the impaired one, indicating that the three words are more similarly represented in hidden unit space. Each cell's shaded area depicts its magnitude from 0.0 to 0.5.

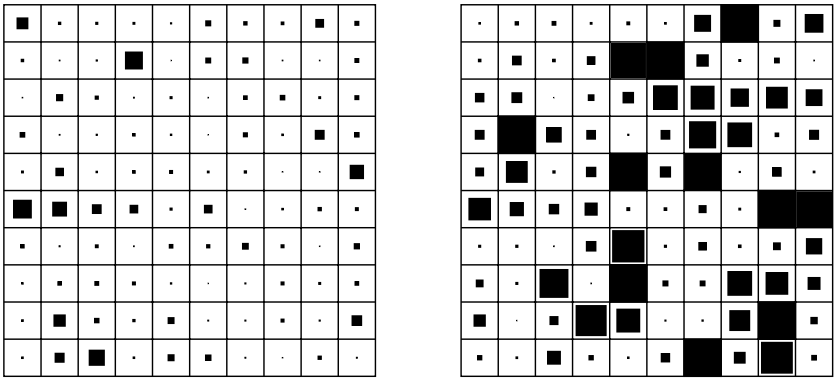

Figure 18. Mean differences between hidden unit activity for the nonword GEAT and EAT, MEAT and TREAT for normal (left) and phonologically impaired network (right). The phonologically impaired network shows greater average differences between the words and the nonword. Each cell's shaded area depicts its magnitude from 0.0 to 0.5.

between EAT and GEAT was computed and stored, as was the difference between MEAT and GEAT, and TREAT and GEAT. The absolute values of these differences were averaged and plotted in Figure 18. The differences are smaller for the normal model.

Table 8 shows these effects numerically. The mean absolute values of the differences of the hidden unit contributions to the tongue feature for EAT, MEAT, TREAT, GEAT are shown, for both the normal and impaired simulation. The aggregate input and output of the tongue feature for each item is shown as well. For the normal network, the hidden unit outputs for the four items are very similar to each other. The effect of this similarity can be seen in the output column, where the outputs are all within threshold of the target value of zero.

In contrast, the impaired network shows large differences in the mean contributions from the hidden units, ranging from 0.14 to 0.22. The aggregate inputs to the tongue feature for the three words are very similar, but the differences between the hidden units are large. Thus, the impaired model is able to pronounce the three words EAT, MEAT, TREAT correctly, but in a qualitatively different way from the normal model. The normal model pronounces EAT, MEAT, TREAT using very similar hidden unit representations, whereas the impaired model produces correct output using different representations of the words.

The impact of representing the words differently from each other is seen in performance on the nonword GEAT. The normal model can pronounce the nonword GEAT correctly, due to the overlap in hidden unit activity among EAT, MEAT, TREAT. The impaired model cannot pronounce GEAT correctly because the representations it has developed for the words do not overlap enough; the nonword is not sufficiently close to any of the word representations to support the correct pronunciation. Put simply, the normal network treats the three words EAT, MEAT, TREAT similarly, and is hence able to pronounce a similar nonword GEAT. The impaired network treats the three words differently from each other, representing them more like unanalyzed, individual wholes with less overlapping structure. Therefore it cannot take advantage of the similarity between them when reading the nonword, even though it can correctly pronounce the words.

Measuring the sensitivity of the hidden unit layer to a particular input feature is also illuminating. By taking the difference in hidden unit activity, as projected onto the *tongue* feature, for the words EAT and MEAT, we can see the overall sensitivity the network has developed to the M orthographic feature. More formally, the sensitivity to the M feature is defined as $s = \sum_{i=0}^{100} |h^i_{eat} - h^i_{meat}|$, where h^i_{eat} is the contribution of the ith hidden unit to the *tongue* feature for the word EAT, while h^i_{meat} is the same measure for the word MEAT. A network which has no sensitivity to M in the context of the word MEAT would measure $s = 0$, while a network which is attending to the M would have a higher s value. The s value is a measure of the degree to which the network has formed word specific representations within the regular pool of words ending in EAT.

The model with moderate phonological impairment shows a higher level of sensitivity to the M letter in MEAT than the normal model does (Figure 19). More importantly, the sensitivity is monotonically increasing with training, whereas the normal model develops a lower level of sensitivity and does not increase. As training progresses, the phonologically impaired model is becoming more sensitive to spurious aspects of the input; the letter M is not necessary for the pronunciation of the vowel in the word MEAT but the impaired model is attending to such information anyway.

The -EAT example suggest that the phonologically impaired networks develop solutions to the orthography to

Table 8
Mean Differences in Hidden Unit Contributions

	Mean Differences				Input From	Feature
	EAT	MEAT	TREAT	GEAT	Hidden Units	Output
Normal Net						
EAT	0	0.04	0.04	0.04	-0.72	-0.01
MEAT		0	0.05	0.05	-0.87	0.03
TREAT			0	0.05	-0.42	-0.02
GEAT				0	-0.91	-0.09
Impaired Net						
EAT	0	0.14	0.17	0.15	1.08	-0.10
MEAT		0	0.24	0.16	1.15	-0.03
TREAT			0	0.22	0.98	-0.14
GEAT				0	2.16	0.59

phonology mapping problem that are more word specific than when the phonological attractor apparatus is functioning. However, we need to consider whether this pattern reflects a general property of the networks rather than idiosyncrasies associated with -EAT. To test this, a set of *neighborhoods* was identified. A neighborhood was defined as a set of words whose orthographic vowel and coda were identical, and also rhymed (for example GAVE, BRAVE, SAVE but not HAVE). A total of 443 neighborhoods were found in the training corpus. The standard deviation of the hidden unit contributions to the vowel's place of articulation was measured for each neighborhood. This number was averaged over all hidden units. The average of these measures over all neighborhoods, and over all normal models was then computed. This procedure was repeated for the moderately phonologically impaired model. The normal models showed a mean standard deviation of 0.1046 while the impaired model showed a mean standard deviation of 0.2818. The mean standard deviation within a neighborhood was also much higher in the impaired networks ($F(1,14) = 225.5$, $p < 0.001$). This indicates that in general the hidden unit representations in the impaired model are much more diverse within neighborhoods than the normal model.

To summarize, the task of the hidden units in the phonologically impaired simulation is more difficult, in that their mapping onto phonology must be *more* accurate than in the normal model. This requirement for greater accuracy causes the model to attend to more word-specific aspects of the input; overfitting the training data is the result. Thus, the model is biased to become a whole-word reader, forming overly divergent representations for words with orthographic and phonological commonalities. Having formed these word specific representations then interferes with computing output for novel items.[6]

This account is consistent with the finding that phonological dyslexics sometimes exhibit greater reliance on orthographic structure than nondyslexics. Rack (1985) found that dyslexics had a greater sensitivity to orthographic cues in a recall task than normals; conversely, they showed less sensitivity to phonological cues. Other studies suggesting that dyslexics show greater sensitivity to orthographic structure have been taken as indicating that they use a more visually-based strategy in reading (see Snowling, 1991, for a review). Our view is that this dependence on orthographic structure is not a strategy but just a consequence of how the mapping between orthography and phonology is learned when the capacity to represent phonological structure is limited.

As we have noted, a mild phonological impairment only affects nonwords whereas a more severe phonological impairment affects exceptions as well. The reasons why exceptions are eventually affected follow from the previous analysis. Phonological impairment requires greater accuracy in the mapping from the hidden units onto phonological output. The greater the impairment, the greater the required accuracy. Achieving higher levels of accuracy requires the recruitment of more hidden units. Comparing Figures 15 and 16, it is apparent that the phonologically impaired simulation used more hidden units in the computation than the normal model. As more phonological damage demands greater hidden unit resources to perform accurate mappings, there are fewer computational resources available to perform more specific tasks, such as exception word decoding. Hence, we start to see an exception word

[6] It is theoretically possible to create the overfitting problem in a model by other means, such as the use of too high a ratio of hidden units to patterns. This could result in poor generalization even in the presence of normal phonology. However, such a developmental condition (entirely normal word reading and phonology but very poor nonword reading) is as yet unattested in the literature. Additionally, the introduction of a training regime that actively discourages the computation of the sound pattern of a word ("instructional" phonological dyslexia; see Joanisse et al., 1998) could result in poor generalization performance in the absence of phonological impairments.

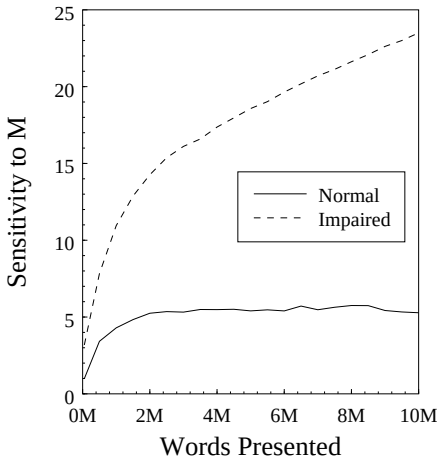

Figure 19. The sensitivity to the letter M in processing the words MEAT versus EAT; normal and moderately phonologically impaired model.

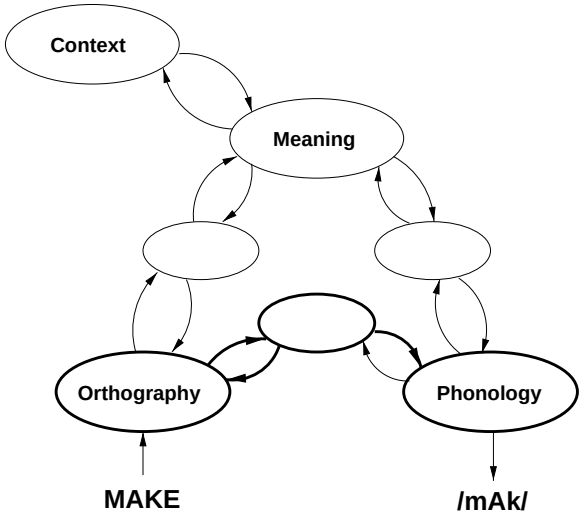

Figure 20. The "triangle" model from Seidenberg and McClelland (1989).

reading deficit at more severe levels of impairment.

This analysis of the basis of exception word deficits in phonological dyslexia is interesting because it is closely related to the account of exception word errors in surface dyslexia presented below. In surface dyslexia, exceptions are more impaired than nonwords. One way to produce this pattern is by reducing the number of hidden units mediating the computation from orthography to phonology. Reducing the computational capacity of the model in this way has a bigger impact on learning exceptions than on learning regular spelling-sound correspondences. Our account of the exception word deficit in severe forms of phonological dyslexia is that the phonological impairment indirectly has the same effect: the capacity of the network is taxed because the orthography→phonology pathway must encode more of the regularities in the system, leaving fewer resources available for exceptions. Thus, in phonological dyslexia, impaired exception word reading arises from lack of computational resources in the orthography→phonology pathway, indirectly caused by the primary phonological deficit. In surface dyslexia, impaired exception word reading arises more directly from reduced computational capacity caused by the lack of hidden units. Thus, we have achieved a unified account of exception word errors in the two cases.

Other "pure" phonological dyslexics. The analysis of the phonologically impaired model predicts that mild levels of impairment yield pure cases, whereas more severe levels of impairment produce mixed cases. As discussed earlier, these results are consistent with the Manis et al. (1996) study. However, there are case studies in the neuropsychological literature of pure yet severe cases of developmental phonological dyslexia. Our view is that these patients' performance reflect other factors outside the scope of our models. One important factor discussed by Manis et al. is the kind of remediation the individual has received. For example, Howard and Best (1996) discussed patient MJ, who was in her eighties at the time of testing but described as a "developmental" phonological dyslexic because she was a lifelong poor reader without any known neuropathology. She exhibited a severe nonword reading impairment with normal word reading. This type of patient can be explained by appeal to the "triangle" model (Seidenberg & McClelland, 1989) shown in Figure 20. Generating the pronunciation of a letter string normally relies on the computation from orthography to phonology. However, if this pathway is completely disabled by brain injury, it is possible to pronounce familiar words by means of the computation from orthography to semantics to phonology. This part of the network does not support the pronunciation of nonwords because they are not represented in semantics. Patient MJ shows extensive evidence of using this semantically-based pronunciation: her reading latencies revealed exaggerated sensitivity to semantic variables such as imageability, and lower sensitivity to spelling→sound properties such as consistency. Patient MJ has had ample opportunity to develop compensatory strategies to deal with her developmental dyslexia. Extensive remediation emphasizing semantic approaches to reading can be expected to result in improved word reading at the expense of nonword reading, and as such can lead to a relatively pure nonword naming deficit.

Speech Impairments and Phonological Dyslexia. A large number of studies have investigated whether the phonological deficits seen in many dyslexics are secondary to more basic impairments in the processing of speech. There is good evidence for this relationship in children with developmental language impairments (sometimes called "specific language impairment"; Bishop, 1992). Many of these children have an impairment in speech perception and pro-

duction, which is thought to underlie their deficits in other aspects of language including phonology, morphology and syntax (Leonard, McGregor, & Allen, 1992; Leonard, Sabbadini, Leonard, & Volterra, 1987; Tallal & Piercy, 1973a, 1973b; Tallal, Stark, & Curtiss, 1976; see Bishop, 1992, 1997 and Joanisse & Seidenberg, 1998 for reviews). By definition phonological dyslexics are poor readers whose language skills are otherwise thought to be normal. If a speech perception deficit were the proximal cause of phonological dyslexia, it would have to be one that has little impact on comprehension or production yet interferes with reading.

Studies of speech perception deficits in dyslexia have yielded mixed results. Most studies have focused on the categorical perception of phonemes such as /b/ and /d/. In several group studies (e.g. Godfrey, Syrdal-Lasky, Millay, & Knox, 1981; Werker & Tees, 1987; Reed, 1989; De Weirdt, 1988; Manis et al., 1997) dyslexic children exhibited less-pronounced categorical perception effects than normals. However, the effects have often been small and not statistically robust. Other studies failed to yield such effects at all in some conditions. Hurford and Sanders (1990), for example, found group differences in phoneme perception in a study of second grade normal and dyslexic children, but failed to find any such group differences with fourth grade children. Moreover, analyses of individual participants in Manis et al. (1997) showed that speech perception deficits were only seen in a subset of dyslexic children. Although additional research needs to be conducted, it appears that many dyslexics who exhibit clear deficits on tests of phonological knowledge perform normally on simple tests of speech perception.

As we will show, the phonological attractor component of the model exhibits categorical perception of consonants. We could therefore examine whether the kinds of phonological impairments that we have introduced in simulating phonological dyslexia create an impairment in this aspect of speech perception.

Categorical perception experiments typically utilize both identification and discrimination tasks. In the identification task, stimuli are constructed with consonants that vary linearly along a continuum between two exemplars, for example from /b/ to /d/, which differ in their second and third formants. Participants are asked to label these tokens as instances of /ba/ or /da/ in a forced choice. Their identification functions are then analyzed with respect to points along the continua. A standard finding is that participants' identification functions tend to be relatively flat and consistent within a category boundary, with a very sharp transition at the boundary point. Although the stimuli vary linearly from one token to another, participants' identification curves are markedly nonlinear.

In the standard discrimination task, participants are given pairs of stimuli and asked to judge whether they are the same or different. The basic phenomenon is that participants are very poor at discriminating stimuli within a category, and much better when the contrasting stimuli span a category boundary. When discrimination scores are plotted against stimuli pairs, a sharp peak is typically found at the category boundary, with much poorer (and flatter) performance within categories (Liberman, Harris, Hoffman, & Griffith, 1957; see also Harnad, 1987). Werker and Tees (1987) and Godfrey et al. (1981) examined the categorical perception of phonemes by normal and reading disabled children. Both studies found group effects on the slopes of the identification functions; the disabled readers' identification curves were slightly less steep than that of control children, although the effect was only marginally significant in Werker and Tees ($p < 0.06$) and in one of two analyses by Godfrey et al. ($p < 0.08$ in one, $p < 0.01$ in another). Werker and Tees and Godfrey et al. also analyzed the children's discrimination scores. Both studies used a formula developed by Pollack and Pisoni (1971) for predicting discrimination performance from a participant's identification curves. When categorical perception is normal, this formula accurately predicts discrimination scores. The match between predicted and obtained discrimination therefore provides an index of deviations from true categorical perception. This procedure provides a more sensitive assessment of the child's phoneme perception than just the identification task. It is possible that a relatively steep identification function might be obtained even if perception of stimuli were more continuous than categorical, through the application of a simple thresholding decision criteria (Massaro, 1987). However, if perception were truly less categorical, the child would then show discrimination performance that is different from that predicted by their identification curve, in contrast to normals.

In both Werker and Tees (1987) and Godfrey et al. (1981), the reading disabled children's discrimination scores were more deviant from their predicted abilities than the normal children's. Thus it was concluded that their perception of the speech tokens was not as strongly categorical. These effects were stronger and more reliable than in the analyses of the slopes of the identification curves.

We replicated the Werker and Tees (1987) and Godfrey et al. (1981) studies, examining the normal and impaired models' processing of speech stimuli like the ones used in categorical perception experiments. Stimuli ranging along a continuum from /b/ to /d/ were created by linearly interpolating feature values. Each ith weighted feature x_i was created from the ith feature of a veridical /b/ (b_i) and the ith feature from a veridical /d/ (d_i) according to the formula $x_i = (1.0 - \alpha)b_i + \alpha d_i$, where α ranged from 0.0 to 1.0. Eleven tokens were created, equally spaced along the continua, by varying α from 0.0 (a pure /b/) to 1.0 (a pure /d/) in steps of 0.1. The α parameter can be thought of as the proportion of /d/ to /b/ in the generated tokens. The vowel /a/ was placed in the vowel slot to create the complete set of tokens ranging from /ba/ to /da/. These stimuli

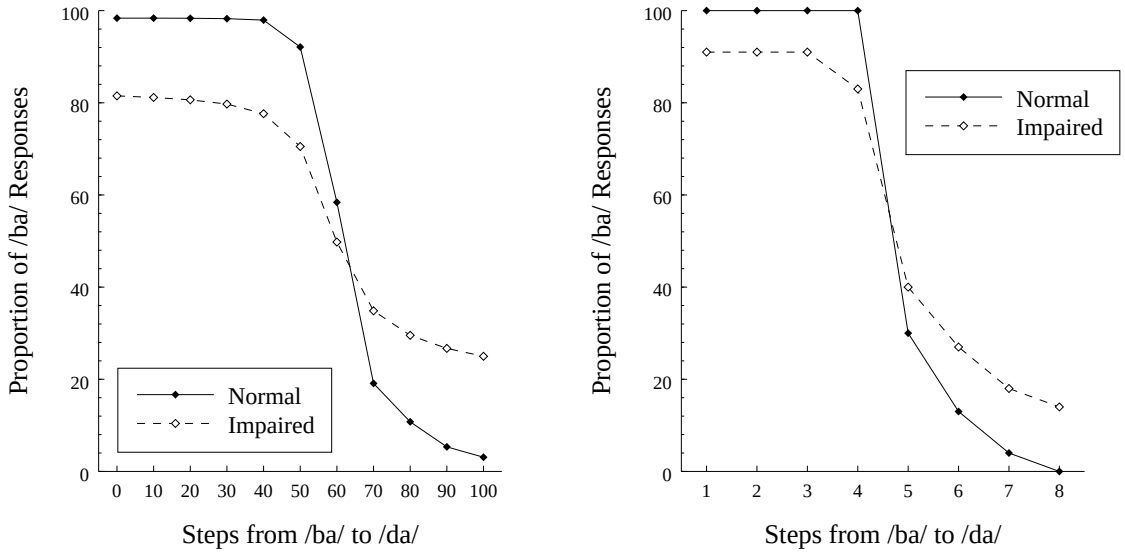

Figure 21. Normal and phonologically impaired models' identification curves (left). Compare with data from Werker and Tees (1987), Figure 2 (right).

were presented to the normal network and to the mild and moderately impaired models. Results for the two impaired models were averaged together, on the view that the participants in the behavioral experiments also varied in terms of the severity of their phonological deficits.

Stimuli were presented as follows. For the first two time ticks, the phonological units of the network were clamped with the representation of the test token. The network was then allowed to run until five ticks, with the values unclamped. The phonological representation at the conclusion of five ticks was then recorded. The euclidean distances between this output phonological representation and veridical /b/ and /d/ were calculated.

To simulate the identification task, the distances from the networks' output to /b/ (Δ_b) and /d/ (Δ_d) were used to generate a probability of labeling the output as either of those stimuli. The probability of an /b/ response, p_b was computed as $p_b = 1.0 - (\Delta_d/(\Delta_d + \Delta_b))$. The probability of an /d/ response was the mirror image: $p_d = 1.0 - (\Delta_b/(\Delta_d + \Delta_b))$ (note that $p_b + p_d = 1.0$). These probabilities were recorded for each network's response to each stimulus token.

Figure 21 depicts the identification curves for the normal and phonologically disabled models, along with data from Werker and Tees (1987). The phonological impairments that we used to simulate phonological dyslexia produce identification curves that qualitatively replicate the Werker and Tees data. The phonologically impaired models' identification curves appear less steep and are less absolute at the endpoints, compared with the normal models. Following Werker and Tees and Godfrey et al., the normal and impaired models' identification curves were submitted to a logistic regression analysis; as in their studies, the impaired simulations' slopes were reliably shallower than the normal models' ($F(1,22) = 8.6$, $p < 0.01$).

To simulate the discrimination task, we examined the processing of tokens that differ by two intervals in Figure 21, as in a standard two-step discrimination task. Each step represents a 10% difference between stimuli; thus all pairs of stimuli differed by 20%. The model was run on each stimulus in a pair using the same procedure as above, and discrimination was modeled by computing the euclidean distance between the computed outputs. This distance was scaled by a constant to yield a probability of correctly discriminating the tokens: the closer the euclidean distance of the outputs, the more difficult the discrimination. The normal model was used as a baseline to establish this constant; it was found that dividing the euclidean distance by eight yielded a good match to the predicted discrimination function.

Figure 22 presents the results. The normal model's discrimination curve closely matches the predicted curve, and shows the expected sharp peak in performance at the category boundary (at about 60%, as in Figure 21), with much worse performance within the category. The impaired models' discrimination scores, in contrast, do not match those predicted by their identification scores. Further, the characteristic sharp peak in performance at the boundary is not seen. As in the Werker and Tees study, we performed an analysis of variance using group (dyslexic or normal) and pairing (0-20, 10-30, etc.) as factors, and the difference between predicted and actual values as the dependent variable. There was a reliable effect of group ($F(1,198) = 191.6$, $p < 0.001$) and pairing ($F(8,198) = 2.3$, $p < 0.02$). These

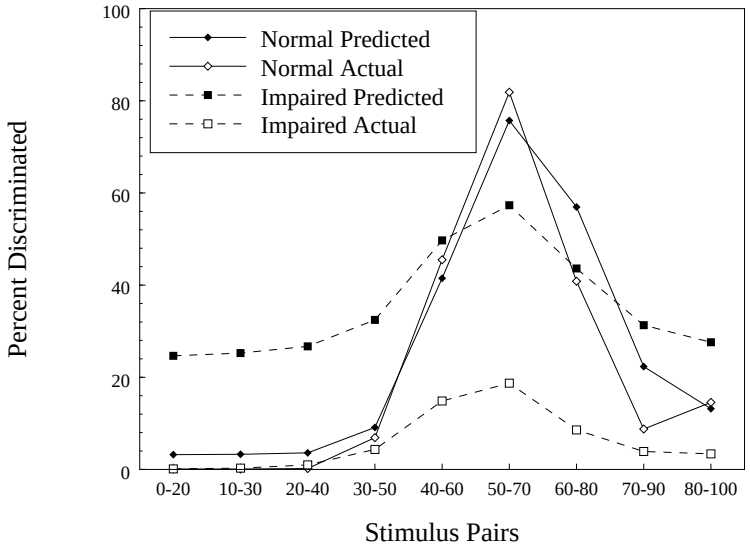

Figure 22. Predicted and observed discrimination values for the normal and phonologically impaired models.

results match those obtained by Werker & Tees's two step condition and demonstrate that the phonologically impaired models exhibited less strongly categorical perception than the normal models.

Because Manis et al. (1997) found that the group effects they observed on the slope of the identification curves were driven by only a subset of the dyslexic children, with many of the phonological dyslexics exhibiting perfectly normal identification curves, we repeated the above experiments using only the normal and mildly impaired models as groups. The slopes of the mildly impaired condition were slightly lower than that of the normal models, but in contrast to the above experiment there was no reliable group difference ($F(1,14) = 1.0$). The discrimination test, however, still yielded a reliable group effect ($F(1,126) = 219$, $p < 0.001$).

In summary, the mild phonological impairment had a significant impact on reading acquisition but not on the phoneme identification task. This result is consistent with a body of findings indicating that dyslexia is not strongly associated with significantly deviant identification performance. As the Werker and Tees and Godfrey et al. study suggested, effects of this mild impairment can be detected using more sensitive measures such as phoneme discrimination. A more severe phonological impairment yields effects on both tasks. This pattern, though not characteristic of most dyslexics, is typical of children with Specific Language Impairment (SLI; see Joanisse & Seidenberg, 1998). These children exhibit impairments in the use of spoken language; their identification functions are markedly deviant; and they are also typically dyslexic. Thus, phonological dyslexia may represent a milder form of the impairment seen in many cases of SLI.

A final issue concerns the relationship between the types of anomalies we have used to create phonological impairments and current hypotheses about dyslexia and language impairment. We created these impairments in two ways: by modifying the architecture or by adding noise to the computations within the phonological system. Both of these manipulations reduce the capacity of the network to encode aspects of phonological structure, which compromises reading in specific ways. The relationship between these types of impairments and the anomalies that underlie phonological dyslexia is unknown. Our models are not closely linked to neurobiology and so these types of impairments cannot be equated with specific brain mechanisms. Given that these impairments give rise to the right sorts of behavioral effects, there is reason to investigate further what their neurobiological correlates might be. At a behavioral rather than neurobiological level of analysis, the simulations can be considered in light of Tallal's hypothesis that dyslexia and language impairment are associated with a temporal processing deficit. Our manipulations involved changing the representational capacity of the system and the efficiency with which patterns were computed. Neither of these involves direct manipulation of the temporal processing dynamics of the model. Rather, the effect of both types of anomalies is to disturb the model's temporal dynamics: the model converges more slowly and less accurately on target patterns. The simulations suggest that it would be useful to think of temporal processing impairments as one of the consequences of more basic underlying deficits.

Reading Delay Dyslexia

Our account of the "surface" form of developmental dyslexia is that it reflects a general delay in the acquisi-

tion of reading skill rather than a selective impairment in reading exception words or in the "lexical route." In the early stages of acquisition, children who are learning to read normally are poorer at reading exceptions than simple nonwords. Hence their performance fits the pattern that in older children has been called "surface dyslexia." In the delay pattern, the dyslexic child's performance is like that of a younger normal reader. This contrasts with the phonological dyslexic pattern, which is not seen in younger normals. The main point of this section is that the delay pattern can be created in the model in several ways and that this fact may be relevant to understanding differences among these children. One way is simply to provide less training for the normal model. Earlier in training the model exhibits poorer performance on exceptions than nonwords, compared to later in training, when performance on exceptions (and regular words) exceeds that on nonwords (see Figure 10). Thus, like the beginning reader, the model early in training exhibits the "surface dyslexic" pattern. This pattern represents a reading delay when it occurs in older children, like the participants in the Manis et al. and Castles and Coltheart studies. The model suggests that this pattern would result if children who have normal capacities (i.e., network architecture and ability to learn) read less often or receive less feedback about their reading so that they benefit less from it. There is good evidence that reading ability is related to amount of reading experience. Stanovich and his colleagues have attempted to assess relative amounts of reading experience ("print exposure") using measures such as the Title Recognition Test (Stanovich & Cunningham, 1992, 1993; see also McBride-Chang, Manis, Seidenberg, Custodio, & Doi, 1993). In these studies, print exposure was correlated with reading skill even after variation in phonological decoding skill was partialled out of the regression equation. Thus, taken with the print exposure results, the model suggests that low levels of performance seen in some dyslexic children may be the result of impoverished reading experience. Our specific prediction is that such children will exhibit the surface/delay pattern rather than the phonological dyslexic pattern.

The extent to which the reading delay pattern can be attributed to lack of reading experience needs to be investigated further. We do not know whether the cases of surface dyslexia identified in previous studies were associated with lack of reading experience. Manis et al. did collect data on Stanovich and Cunningham's (1992) Title Recognition Test, but found no differences between the surface and phonological subgroups on this measure. This finding is ambiguous; it may be that more sensitive measures of reading experience are needed, but it is also possible that the impaired performance of the surface dyslexics in that study was not due to lack of reading experience. As we discuss below, there are other ways to produce the delay pattern in our model. It would also be important to investigate further the nature of this putative impoverished reading experience. It could reflect differences in amount of reading associated with cultural or socio-economic factors but it also might reflect the ineffectiveness for many children of some methods being used to teach reading.

A second way to create a delay is to use the standard architecture and provide the normal amount of training, but use a non-optimal learning rate. This creates a situation in which the model does not obtain the normal benefits from a given amount of experience. We examined this possibility by conducting simulations in which we varied the learning rate parameter of the model. The term *learning rate* is used here for historical reasons; it refers to specific parameter in the learning algorithm, not the overall rate at which a network learns, which is affected by many other factors. The gradient computations from the backpropagation algorithm specify the direction in weight space for the network to move; the learning rate parameter determines how far in that direction the weights should be changed. A learning rate that is too large can cause the network to oscillate or become trapped in local minima. A learning rate that is too small can cause a network to take a very long time to train. While techniques exist for automatically determining an appropriate step size (e.g. Jacobs, 1988), very often the appropriate value is determined by trial and error. The value we arrived at for the normal simulations was $\mu = 0.005$. To create a condition in which the network is not as able to profit from training experience as the normal model, we ran a simulation with a much smaller learning rate $\mu = 0.0001$. All other aspects of training were identical to the normal model. Figure 23 shows the results. With a lower learning rate the network experiences dramatically slowed acquisition of exception words, and a lesser but still pronounced impairment on nonwords.

A third way to produce the delay pattern was explored by Seidenberg (1992), who reported a simulation that examined the effects of degrading the orthographic input to the SM89 model. The purpose of this simulation was to explore how deficits in the encoding of orthographic input would affect learning to read. Such visual-perceptual deficits have long been hypothesized to be a cause of dyslexia. The evidence for such impairments is inconsistent, as might be expected if this type of impairment were relatively rare and not the only basis for reading impairment. Seidenberg (1992) degraded the orthographic representations in the Seidenberg and McClelland (1989) model by ensuring that more orthographic units were active for each word than normal. This decreased the discriminability of individual words. This impairment created a general delay, with poorer performance on regular and exception words and nonwords.

Finally, a fourth way to create a reading delay is to reduce the model's capacity to encode information regarding the mapping from orthography to phonology. As we have observed, the hidden units play an important role in the encoding of orthographic-phonological correspondences. The

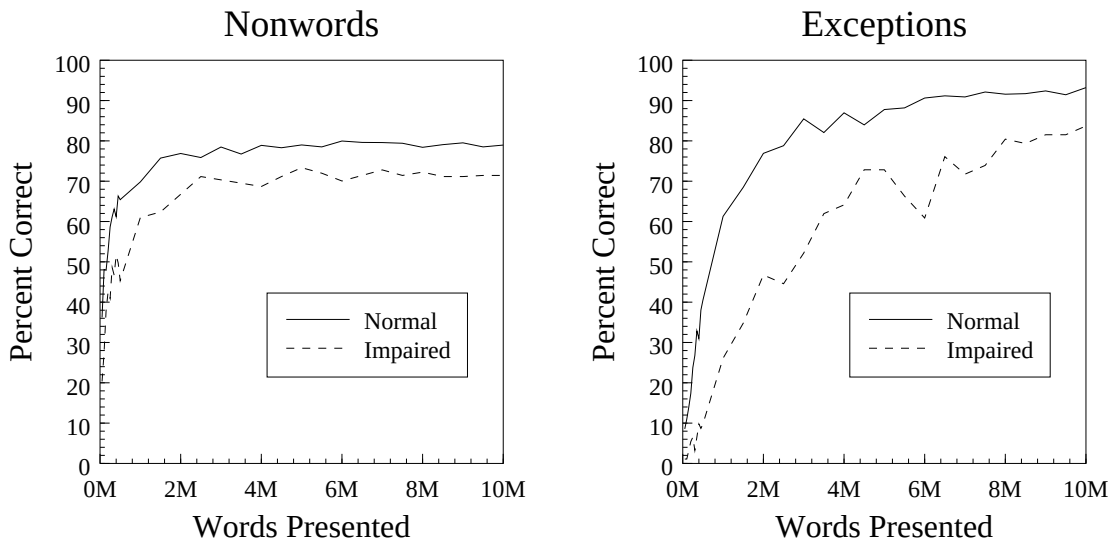

Figure 23. Lower learning rate simulation. Exception word performance (right) is more affected than nonword performance (left).

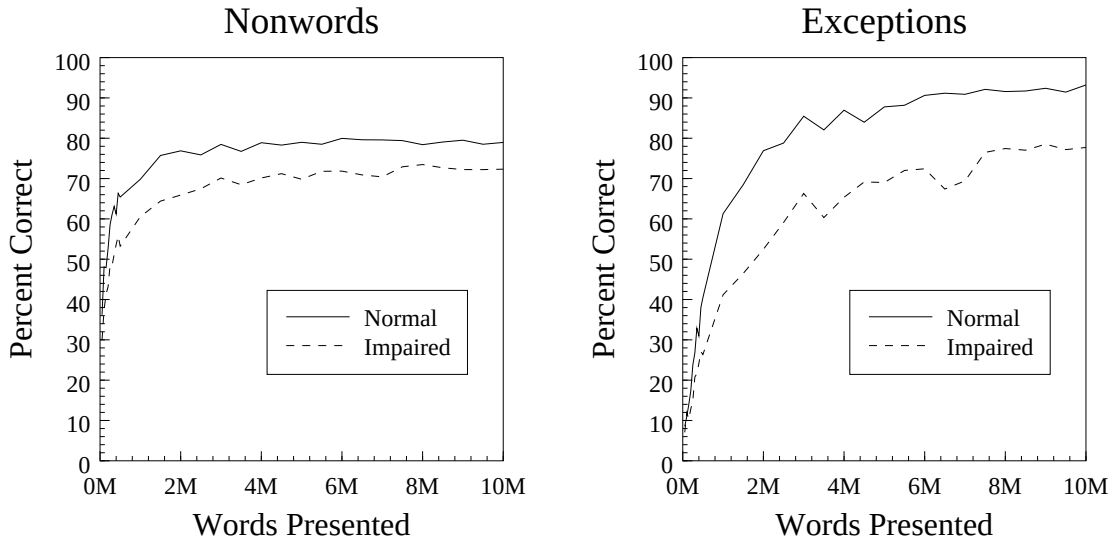

Figure 24. Effect of lower number of hidden units on nonwords and exceptions.

network must have the capacity to encode both systematic aspects of these correspondences and the idiosyncrasies associated with exception words. Although not the focus of their article, Seidenberg and McClelland (1989) reported the results of a simulation in which their model was configured with half the usual number of hidden units mediating the computation from orthography to phonology. This manipulation had a bigger impact on the acquisition of exception words than regulars, but they did not examine its effects on nonword generalization.

We replicated the Seidenberg and McClelland experiment using the attractor network, reducing the number of hidden units from 100 to 20. Twenty hidden units does not allow the network to learn the complete training set and so represents a severe deficit. As before, eight simulations were conducted with different initial random number seeds. The phonological component was pretrained exactly as in the normal model, reflecting the absence of a phonological processing impairment.

The results are given in Figure 24. Both nonwords and exceptions show decrements but, importantly, the impact is greater for the exceptions. The result is a "mixed" surface

dyslexic, showing a primary impairment to exception words and a secondary impairment to nonwords.

Analysis of the Effects of Reducing the Number of Hidden Units. Reading exception words generally requires attention to larger portions of the words than does reading regular words: whereas MINT can be correctly pronounced by looking at the onset M- and rime INT, pronouncing PINT correctly requires using information from the entire word. The reading of exception words can be seen as a series of xor-style problems (Minsky & Papert, 1969), in which one unit's state depends on the states of other units in the environment. Reducing the number of hidden units primarily affects the capacity of the network to encode dependencies that span more letters. Although acquisition is slowed for all types of items, with sufficient training the model can eventually learn the simple and redundant patterns characteristic of regular words. Exceptions, however, continue to be impaired.

The greater demands imposed by exception words can be quantified by developing a measure of computational work. By measuring the entropy of the vowel phoneme (see Equation 4) across all words in the training set, we can derive a measure of how much information is needed per word to communicate that vowel. Comparing this measure to the conditional entropy (Equation 5) of the vowel phoneme with respect to the orthographic vowel, we can see the extent to which the orthographic vowel reduces the uncertainty of the vowel phoneme. The extent to which increasing conjunctions of orthographic information reduce the uncertainty of the vowel phoneme can be measured in this way; first measuring the entropy of the vowel phoneme $H(\mathcal{Y})$, then the conditional entropy of the vowel phoneme with respect to the first orthographic vowel $H(\mathcal{Y}|\mathcal{X}_1)$, then with respect to the first orthographic vowel and the letter that follows it $H(\mathcal{Y}|\mathcal{X}_1,\mathcal{X}_2)$, and so on for the maximum 4 letters comprising the orthographic rime $H(\mathcal{Y}|\mathcal{X}_1,\mathcal{X}_2...\mathcal{X}_4)$

Figure 25 plots the conditional uncertainty of the vowel phoneme over rimes of different length for both the whole training set and a subset of the training set containing only regular items. The uncertainty of the vowel phoneme when a window of 3 letters into the word body is considered is essentially zero for the regulars, but still high for the entire training set. Thus, regulars in general require less orthographic information to disambiguate their pronunciations.

Encoding higher-order dependencies is what the hidden units are for, and as Figure 25 suggests, although all types of words tend to require the capacity to represent such dependencies, exceptions are more likely to require more than 3 letters to be disambiguated. With fewer hidden units, the capacity of the network to encode these dependencies is reduced, which has a larger effect on exceptions. If the network were unable to encode dependencies covering more than 3 letters it would still get most of the regulars correct but the exceptions would be highly impaired. With a more severe deficit, the capacity of the network to learn generalizations covering 3 letters will become impaired, which begins to affect mastery of regulars will suffer as well. Because nonword performance is parasitic on mastery of regular correspondences, the "mixed" cases showing impaired nonword performance naturally fall out of greater degrees of impairment. Of course, a network suffering a gradual reduction in hidden units will not suddenly be unable to combine 3 letters; the network is trying to optimize performance over items by their frequency, and as such will lose the capacity to attend to large word bodies in lower frequency items (e.g., YACHT) before high frequency ones (e.g., THE).

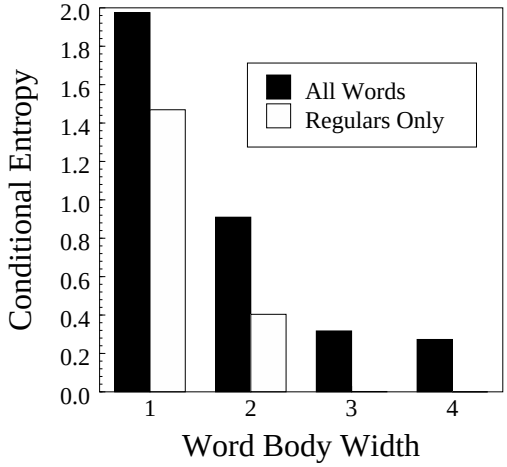

Figure 25. Conditional Entropy of vowel phoneme with respect to 1, 2, 3 and 4 letters in rime.

Comparisons to Behavioral Results

Having described the phonological and reading delayed simulations and shown that they exhibit general features of dyslexic performance, we can now provide closer comparisons to behavioral data. Figure 11 presented summary data from the Manis et al. study. The figure illustrates several findings. Surface/delay dyslexics were impaired on both exceptions and nonwords compared to same-aged normal readers; they were more impaired on exceptions than nonwords; and their performance closely resembled that of younger normal readers. Phonological dyslexics were also impaired on both exceptions and nonwords compared to same-aged normal readers; they performed at the same level on exceptions and nonwords but compared to both normal reader groups they were more impaired on nonwords; their performance was not like younger normal readers.

Figure 26 presents the data from comparable conditions in our simulations. The mean performance of the normal model, the delay dyslexic simulations and the most extreme phonologically impaired simulations were measured at 1.5 million training trials. In addition, the normal model's per-

formance was assessed with fewer training trials (0.5 million), which yields performance similar to the younger normals in the Manis et al. study. These simulation results capture all of the main results seen in Figure 11. The only deviations between Figures 11 and 26 are related to the somewhat lower levels of nonword performance in the model. Thus, the simulations replicate the kinds of dissociations between nonwords and exceptions seen in behavioral studies at approximately the same levels of performance.

A final analysis addressed the different developmental patterns in the phonological and delay cases. Manis et al. (1996) found that whereas surface dyslexics exhibited behavior characteristic of younger normal readers, the phonological dyslexics exhibited an aberrant pattern not seen in normal readers at any age. Figure 26 exhibits this pattern, but we can make the point in a more general way as follows. The behavioral data show that surface dyslexics' performance on exception words and nonwords is quantitatively within the range of younger normal readers, whereas the phonological dyslexics' is not. Figure 27 shows the performance of the normal models and the surface dyslexia simulations on these types of items. It can be seen that over different levels of performance, the normal and surface dyslexic models show similar ratios of word and nonword accuracy. In contrast, the curves for the phonological dyslexic simulations deviate from those in the normal and surface dyslexic models, because of the low levels of nonword performance compared to exceptions. The data are consistent with the conclusion that the surface dyslexics pattern of impaired reading is like that of younger normal readers but the phonological dyslexic pattern is not.

This pattern can also be seen in the results of a regression analysis that used the normal model's performance to predict the nonword performance of the impaired models. A regression was performed to predict the normal model's nonword scores from its performance on exceptions($r^2 = 0.95$, $F(1,31) = 288$, $p < 0.0001$). This regression equation was then used to predict the nonword performance for each of the impaired models in turn, given their performance on exceptions. Figure 28 shows the results of this analysis at different points in training. The normal model does a good job of predicting the surface/delay models' performance; the residuals are small and not much larger than for the normal model itself. In contrast, the phonological dyslexic models yield larger residuals, indicating that nonword performance is not as well predicted by exception performance. These results obtain at all levels of training. Thus, the surface/delay models require more training, but their relative performance on exceptions and nonwords is like the normal models. The phonological dyslexic models are on a different trajectory, because of the more extreme impairment in reading nonwords.

The model provides insight about why the developmental patterns differ for the two subtypes of dyslexia. Consider again the mapping that the hidden units must perform. With an impairment in the phonological attractor's capacity to represent information (phonological dyslexia), the nature of the task the hidden unit layer must solve is changed. Instead of having to map an orthographic form onto an approximate phonological form which is then refined into the correct pronunciation, the output of the hidden unit layer must be relatively exact. In contrast, the input/output task facing the hidden unit layer in the case of reduced hidden units is the same. It is not the *nature* of the task that the hidden units must solve that is changed, but the *capacity* of the hidden units to perform that. Thus, in the phonological dyslexic simulations, changing the nature of the task causes the model to arrive at a solution that is different from normal; in the surface dyslexic/delay simulations, the task remains the same but the model arrives at the solution more slowly, producing a developmental delay.

Summary of Dyslexia Simulations

The behavioral literature suggests two distinct subtypes of developmental dyslexia, one related to a phonological impairment and one reflecting a general delay in the acquisition of reading skill in the absence of a phonological impairment. The modeling work presented here accounted for the phonological subtype in terms of damage to the network's capacity to develop highly structured phonological representations; this in turn has an impact on the pronunciation of nonwords at mild levels of impairment, and exceptions as well at more severe levels. Phonological damage affected not only the rate at which the networks learned and the asymptotic level of performance but also the developmental trajectory, creating a deviant pattern. The pattern that has been called surface dyslexia is created by any type of impairment that slows the acquisition process, yielding a developmental delay. We discussed several ways such a delay could be produced in the model and it remains for further research to determine which of these causes is relevant to particular subgroups of children.

This account is consistent with the results of behavioral genetic studies of the heritability of dyslexia that implicate separate phonological and nonphonological factors (Olson, Forsberg, & Wise, 1994). Olson and his colleagues have provided extensive evidence concerning the heritability of phonological coding skills (Olson et al., 1989). More recently Olson et al. (1994) reported significant heritability for a factor they termed orthographic coding. The data derive from performance on an orthographic choice task in which participants decide which of two stimuli is the correct spelling of a specified word. The alternatives are either a word and matched pseudohomophone (e.g., RAIN-RANE) or two homophones (PAIR-PEAR). This task is one of the few that assesses orthographic knowledge without introducing phonological confounds. Manis et al. (1996) found that their surface dyslexic participants performed more poorly than normal readers on this task, whereas phonological dyslexics did not. It is clear that the kinds of

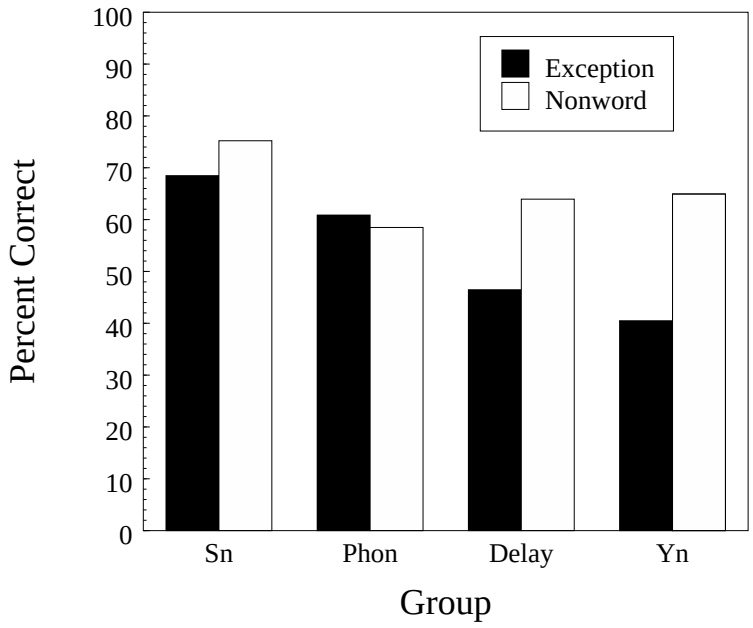

Figure 26. Performance of normal models (Sn), phonologically impaired models (Phon), and delay/surface models (Delay) at 1.5M word presentations, and the normal model at 0.5M word presentations (Yn). Compare with Figure 11.

impairments that we have hypothesized to underlie surface dyslexia could affect performance on the orthographic choice task. The task involves remembering how a particular phonological wordform is spelled. The ability to associate a spoken word with its pronunciation would be expected to depend on factors such as amount of reading experience, ability to learn, and visual encoding. Thus the behavioral genetic data are consistent with the conclusion that there are separate phonological and nonphonological impairments that underlie phonological and surface/delay dyslexia, respectively.

It should be clear that although the several types of impairments related to reading delay produce qualitatively similar effects on exception and nonword reading, they make different predictions about performance on other tasks. For example, it should be possible to obtain independent evidence as to whether some children whose capacities are otherwise normal merely read less often. Such children would be expected to greatly benefit from interventions that simply provide additional experience. Similarly, only the children whose delayed reading is related to a learning problem should exhibit this type of deficit on other tasks, and they would be expected to benefit less from additional experience. We would also expect only some children who exhibit the delay pattern to show deficits on tasks related to perceptual encoding of print. Finally, the hypothesized resource limitation is harder to independently establish, given that it may be specific to reading and therefore leave performance on other tasks unaffected. This type of deficit might be implicated by excluding the other possibilities. A child who has adequate perceptual and learning abilities who receives appropriate training and experience yet exhibits a developmental delay may have a problem of this type.

Our simulations also examined the effects of different degrees of impairment. Factors that affect reading performance, such as the quality of phonological representations or computational capacity, may vary across individuals. The simulations suggest that relatively "pure" cases of phonological or delay dyslexia, in which performance on only one of the two criteria types of stimuli is below normal limits, are associated with relatively mild forms of impairment. With more severe impairments, both types of stimuli are affected, creating a "mixed" pattern that is most common. These predictions are consistent with data from the Manis et al. (1996) and Castles and Coltheart (1993) studies. At present, we envision only one other way of creating a pure pattern: extensive remediation that heavily emphasizes specific decoding strategies. Remediation that focuses on mastering spelling-sound correspondences or developing a sight-word vocabulary may create dissociations between exception word and nonword reading. These and other ways in which children's remediation experiences may mask their underlying deficits are discussed by Manis et al. (1996).

The two types of dyslexia have quite different underlying sources, and their effects are different in specific respects. However, if one merely looks at the most severe surface and phonological dyslexic simulations, both are impaired on reading both nonwords and exception words. This

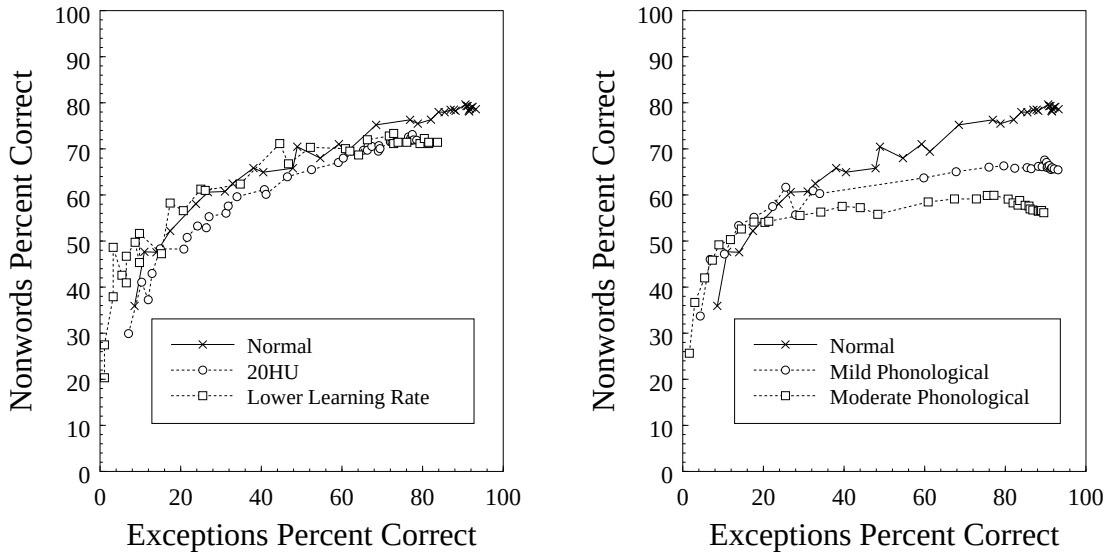

Figure 27. Nonword versus exception word performance measured throughout training, for surface/delay dyslexia simulations (left) and phonological dyslexia simulations (right). The relationship between exception word and nonword accuracy is similar for the surface/delay and normal simulations but not the phonological and normal simulations.

raises a strong cautionary note regarding the clinical diagnosis of developmental dyslexic children. The model predicts that these mixed cases can arise from very different underlying deficits. It will be necessary to obtain data concerning performance on other tasks to differentiate among the many children who exhibit the mixed pattern. This would seem to be a prerequisite to implementing remediation programs that are relevant to the child's underlying problem. Remediation practices need to be sensitive to the differing etiologies among dyslexics who exhibit qualitatively similar levels of performance on simple tasks.

4. Effects of Literacy on Phonological Representation

The final simulations examined how phonological representations are affected by the acquisition of reading skill. We have seen that models that have richer representations of phonological information perform better on the task of learning to map from orthography to phonology. The main impact was on the capacity to generalize, i.e. pronounce unfamiliar letter strings. This capacity plays an important role in becoming a skilled reader and deficits in this capacity are seen in many dyslexics. These results are compatible with evidence that prereaders who have developed more segmental representations of phonology do better at learning to read.

However, other evidence suggests that achieving segmental phonological representations is the outcome of learning to read an alphabetic orthography. The evidence is provided by studies of literate and illiterate participants indicating that only literates have the ability to segment spoken words into component phonemes (Morais et al., 1979; Read et al., 1987; Morais et al., 1986). On this view, "phonological awareness" tasks such as phoneme counting or deletion are highly correlated with reading skill because they require manipulating phonemic representations and the achievement of such representations is one of the results of becoming a skilled reader. There has been considerable controversy as to whether segmental phonological representations are a prerequisite to learning to read or the outcome of achieving literacy (Cossu, Rossini, & Marshall, 1993; Liberman, Shankweiler, Liberman, Fowler, & Fisher, 1977; Morton & Frith, 1993)

An alternative possibility that we can explore using the simulation model is that there is a reciprocal relationship between the development of segmental phonological representations and learning to read (Morais, Alegria, & Content, 1987). In the course of learning a spoken language, children develop representations of higher-order relationships among features that support segmental phenomena such as being able to delete a phoneme from a word. Children who have had more success at developing such representations are better prepared for learning to read. The development of such representations is greatly accelerated, however, by exposure to alphabetic writing systems.

We have already shown (in Section 1) that an attractor network trained to encode phonological representations of words develops knowledge of relationships among features and segments. In Section 2 we trained a model to associate orthographic codes with these phonological representations. The training procedure involved interleaving

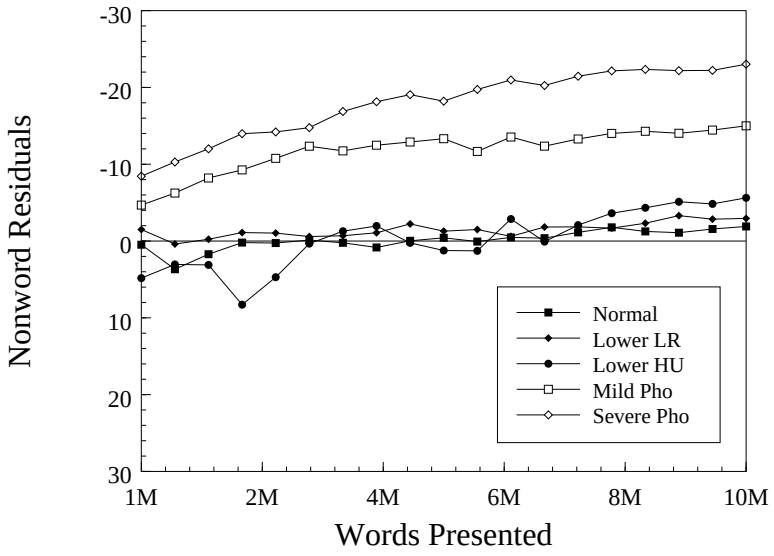

Figure 28. Residuals for nonword performance as predicted by exception word performance. Delay condition models show very small residuals; phonologically impaired models show much larger disparity between predicted and obtained nonword performance.

reading and listening trials that differed in terms of which weights were adjusted. On reading trials, the model was trained to compute a phonological code for an orthographic input and all weights were modified. On listening trials the model was trained on the pattern retention task and only the weights within the phonological attractor were modified. Thus, the weights within the phonological apparatus had to achieve values that allowed the network to perform both tasks. The fact that these weights were affected by their role in the reading task provides the basis for effects of literacy on phonological representation.

Pre- and Post Literate Analysis

The design of these simulations is very simple: we examined the phonological attractor network before and after training on the reading task. The pre-literate model consisted of the attractor network trained as described in Section 1. The post-literate model was the same attractor network after training on the reading task. Differences between the two are effects of literacy on phonological representation. We also included a third condition as a control. Training on the reading task involved additional listening trials that amounted to 10 million extra trials like the ones in the pre-literate phase. In order to assess whether any differences between the pre- and post-literate nets were merely due to the number of phonological training trials, a third condition was included: the overtrained illiterate condition, which was the pre-literate model trained for an additional 10 million listening trials.

Pattern Completion. The pattern completion task (Section 1) was repeated using the the phonological attractor networks associated with the 3 conditions described above. Recall that the procedure on this test involved deleting one feature of each phoneme in every word and examining how the model restored the pattern after several time ticks. As before, eight simulations were used in each condition. The literate network's overall sum squared error (*sse*) in the pattern completion task was lower ($sse = 0.0732$) than either the pre-literate network ($sse = 0.0824$) or the network that only received additional phonological training ($sse = 0.0804$). This condition effect was significant ($F(2, 21) = 18$, $p < 0.0001$). This result indicates that the reading task allowed the network to learn more about the relationships between features than did either phonological pretraining or additional phonological training.

Segment Restoration. The pattern completion test assessed the network's ability to complete an individual feature within an otherwise intact word. We now consider the model's capacity to restore entire segments and show that its performance is greatly affected by training on the reading task. The test was based on the phoneme restoration phenomenon (Warren, 1970). In such studies, the participant typically hears a word or sentence with a phoneme replaced by an extraneous noise (e.g. a cough, buzz, or hiss). The auditory illusion that participants report is that the distorted word was intact; participants often insist that the noise was in addition to or outside of the word (see Warren, 1996, for an overview). Some of these restoration effects involve the top-down use of semantic and pragmatic contextual information, phenomena beyond the scope of the present research. Our test was more narrowly focused on the extent to which the model could fill in segments of isolated words

based only on phonological knowledge. If a phoneme in a word was distorted by noise, how likely was the network to restore a phonotactically legal phoneme, given the constraints of the phonological environment?

Each of the 6 phoneme segments was tested in turn. On each trial, a word was chosen from the training set and the input units were correctly initialized to the word's phonological form, except for the 11 features in the deleted segment. Those features were set to random values in the range $(-0.25, 0.25)$. The values of all correctly-specified features were clamped. The 11 initially random features were left free to be changed by the network. The network was run for the standard 4 ticks, at which time the test segment was evaluated according to the nearest neighbor metric. This test was repeated for all words and all six phoneme positions.

Data were scored in terms of whether the phonological output was phonotactically legal or illegal. The phonotactic legality of the phonemes being restored was defined in terms of onsets, vowels and codas: if the phoneme that was randomized was in the onset, then the resulting onset after processing was tested against all other onsets in the training set. If it resulted in an onset that existed in the training set, then it was scored as a phonotactically legal restoration. For example, if the /r/ in the word /brejd/ was randomized, and the resulting output was /blejd/, the onset /bl/ was compared to all other onsets in the training set. Since /bl/ exists as an onset, that restoration is a legal one. If, in contrast, the network restored a /bkejd/, that would be scored as illegal, since the onset /bk/ does not exist in the training set. Similar tests were used for restored segments in the vowel/diphthong slots, and the coda slots.

Figure 29 depicts the percentage of illegal responses across segment for the phonological attractor network, the overtrained illiterate network and the literate network. For all phoneme slots, all networks were able to produce a legal phoneme from the local environment most of the time. The range of phonemes that constitute a legal blend in English is in fact quite constraining. If a random phoneme is substituted for the network's output, then across all segments the result is phonotactically illegal 72% of the time. The models perform much better than chance, indicating that the phonological attractor has absorbed knowledge of English phonotactic regularities.

The networks that were not trained on the reading task exhibit imperfect knowledge of phonotactics; they produce phoneme blends that no human native speaker of English would make, particularly in the coda. The literate network, however, yielded better performance than either the preliterate or overtrained networks, as is shown in Figure 29. Note especially the improvement in performance on the second vowel position and coda, on which the non-literate networks had performed most poorly. This is an important result because the test specifically assesses effects related to segmental phonological structure. After training on the pattern retention task, the phonological attractor network had a tendency to replace missing segments with other well-formed segments, but the tendency was much stronger after training on the reading task.

Table 9
Mean Squared Difference in Weight Magnitude Between Literate and Overtrained Non-Literate Networks

Connection	Squared difference
within onset	0.067
onset / cleanup	0.055
onset / rime	0.020
rime to rime	0.096
rime / cleanup	0.117

Magnitude and Distribution of Weights

The weights in the networks were examined in order to investigate why performance on the above tasks improved in the literate network. One finding was that the the average magnitude (absolute value) of the weights in the literate network's phonological component (0.250) was significantly higher than either the pre-literate network's (0.168) or the overtrained illiterate network's (0.199) ($F(2, 14790) = 681$, $p < 0.0001$). Figure 30 provides data about differences between the weights in the post-literate and overtrained illiterate conditions. The figure shows the average squared difference between the magnitudes of the weights in the two nets, averaged over phonemes. The six phoneme segments and the cleanup unit group are shown in a matrix, with the "from" connections shown as rows and the "to" connections shown as columns.

Visually, it appears that larger differences are in the rime (vowel and trailing consonants). The projections from the rime to the cleanup units and back underwent particularly large changes. The diagonal of Figure 30 indicates the connection changes within a segment; that is, the weights from a segment to itself. Those sections underwent even greater change overall than the connections from one part of the rime to another.

Table 9 provides summary data concerning different types of weights. The biggest effects were on the weights between rime and cleanup, the weights within the rime, and the self-connections. The effects within the onset and from onset to rime were smaller. These results suggest that the weights within the phonological attractor were affected by literacy in ways that preferentially changed within-segment weights over inter-segmental weights, and within-rime and within-onset connections over those crossing the onset-rime boundary. The results are consistent with evidence suggesting that learning to read results in increased sensitivity to onset and rime units (Treiman, 1992).

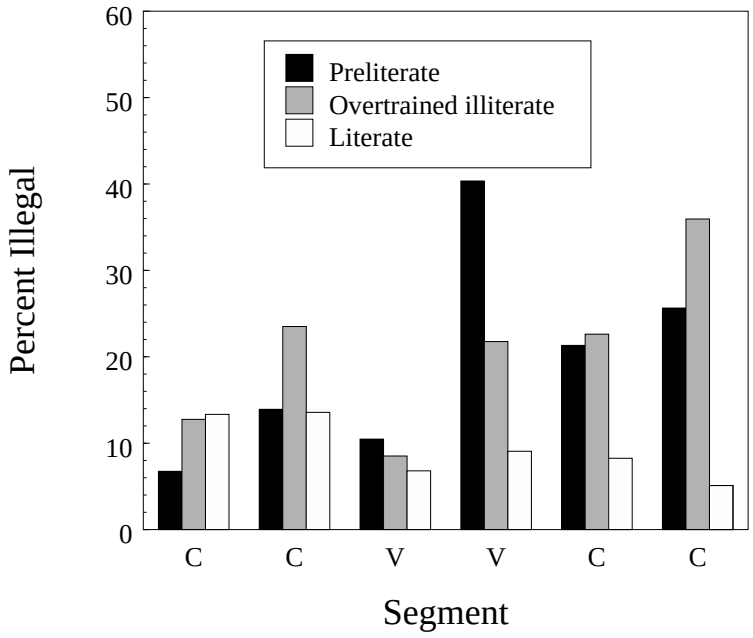

Figure 29. Effect of literacy on segment restoration task. Again, the literate network is better than either pre-literate network or network which received extensive auditory training.

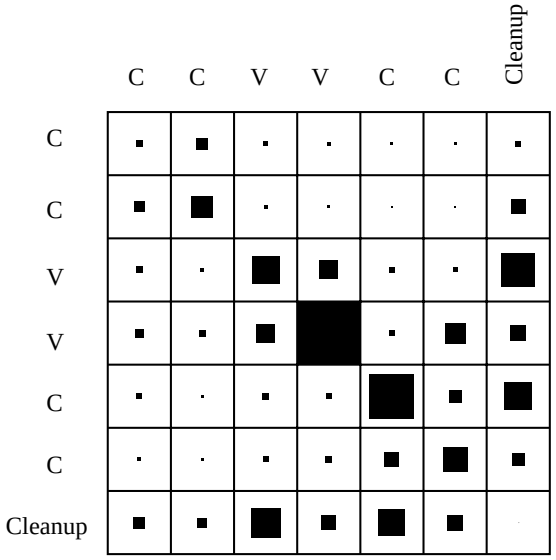

Figure 30. Difference between weights in literate and overtrained literate network, averaged over segment. The rows are the "from" segments of the connection matrix, and the columns are the "to" segments. Dark areas show greatest change. Each cell is scaled to magnitude 0.5.

Rhyme Detection. To test whether the literate network was more sensitive to the onset-rime structure of words, we examined the model's performance on rhyming words. All pair of words in the training corpus that rhyme (N= 22,175 pairs) were identified algorithmically. As a control, an additional set of 22,175 pairs of nonrhyming words was creating by permuting the list of rhyming words. For each word pair, the phonological units were initialized and clamped with the sound pattern of one word in the pair. The phonological network was run for 5 ticks without influence from the reading component. The activity values of the cleanup units at the end of cycling were recorded. The network was then initialized with the sound pattern of the second member of the pair, it was run, and the activity of the cleanup units was also recorded. The distance between the cleanup unit activity for the two words was computed. This was done for each of the rhyming and control pairs, for the literate and nonliterate network. These distances were analyzed in a 2x2 design, using literacy of network and pair type (rhyming or control) as factors.

The results are summarized in Table 10. A strong effect of rhyming was obtained: words that rhyme were overall much closer in their cleanup unit activation state than were the controls. There was also an interaction between literacy and rhyme was also observed ($F(1, 44348) = 1179$, $p < 0.0001$). The interaction is due to the fact that the lit-

Table 10
Literate and Illiterate Network Rhyme Similarities

	Pair		
	Rhyme	Control	Difference
Literate	4.48(1.76)	10.12(3.11)	5.64
Illiterate	4.74(1.79)	9.96(2.58)	5.22

Note. Means are shown, with standard deviations in parenthesis.

erate net represented rhymes more closely than the illiterate net.

Table 11
Orthographic Influences on Rhyme Detection

	Pair		
	Similar Rhyme	Dissimilar Rhyme	Difference
Literate	4.52(1.72)	4.43(1.82)	0.09
Illiterate	4.76(1.72)	4.72(1.87)	0.04
Orthographic Presentation	5.28(2.10)	5.61(2.28)	-0.33

Note. Means are shown, with standard deviations in parenthesis.

The next question is how literacy affected the representation of similarly spelled (e.g., KITE-BITE) and dissimilarly spelled (KITE-NIGHT) rhymes. Our initial intuition was that the effect of training on the orthography→phonology task would be to cause the phonological representations of dissimilarly spelled rhymes to differ more than the similarly spelled rhymes. This prediction was based on a study of college student participants' rhyming judgments by Seidenberg and Tanenhaus (1979). They found that, with auditory presentation of stimuli, participants took longer to judge that two dissimilarly-spelled words rhymed than two similarly-spelled rhymes. One interpretation of this result is that the phonological representations of dissimilarly spelled rhymes differ more than those of the similarly spelled rhymes because of the influence of orthographic knowledge. The 22,175 rhyming pairs used in the previous analysis were split into pairs that share the same orthographic word body (12,416) and those that do not (9,759). These words were tested for similarity in the same manner as the rhyme detection test. The results are summarized in Table 11. There was a reliable effect of orthography, but it was because the literate network was representing differently spelled rhyming words *more* similarly to each other than the similarly spelled words.

Instead of pulling the representations of dissimilarly-spelled rhymes away from each other, the phonological attractor is apparently compensating for the orthographic differences. The mapping from orthography to phonology (i.e., the input to the phonological attractor network from orthography) is more similar for similarly-spelled rhymes than dissimilarly-spelled rhymes. Nonetheless, all three words rhyme. Thus, the effect of learning to read is to separate dissimilarly-spelled rhymes which the phonological cleanup units have to counteract for the net to converge on the same rhyme representations.

If this account is correct, then the phonological differences between similarly and dissimilarly spelled rhymes should depend on whether they are presented phonologically or orthographically. Table 11 also presents data concerning the distances between the phonological representations of similarly and dissimilarly spelled rhymes when these codes were computed from orthography. With orthographic input, similar rhymes are closer than dissimilar rhymes. With phonological input, the pattern is the opposite in both the literate and illiterate nets; the effect is larger in the literate net. These data indicate that phonological representations in the literate net are shaped by the fact that they are also the target for the reading task.

Returning to the Seidenberg and Tanenhaus (1979) results, the model suggests that the fact that dissimilarly-spelled rhymes are harder to judge as rhymes than similarly-spelled rhymes is not due to differences in the similarity of their phonological representations. Rather, the effect seems to reflect the feedback ("resonance": Van Orden et al., 1990) between phonology and orthography that occurs in a fully interactive system. In the rhyming tasks participants hear familiar stimuli that rapidly activate several types of associated information, including meaning and spelling. This information in turn feeds back on the phonological system. We have not implemented this feedback in our model, but it has been assumed by this theoretical framework since Seidenberg and McClelland (1989); other evidence for this type of feedback is provided by Stone, Vanhoy, and Van Orden (1997).

Discussion

By examining the phonological component of the model with and without training on the reading task it was possible to examine whether the reading task changed the representation of phonological information. The weights in the literate network were larger than in the other networks, indicating that it has developed stronger phonological attractors. The changes to the weights produced better performance on feature and segmentation restoration tasks and sharpened the representation of the rime. Additionally, rhyming words were more similar to each other in the literate model than the non-literate ones. Orthographic influences were seen on the phonological representation, reflecting the differing demands of the reading task: the phonological representation compensated for differences in spelling of rhyming words.

The results we have described in this section are preliminary in that we have not exhaustively examined all of the ways in which the reading task affected phonological representation. However, this was the first attempt to provide a computationally explicit account of how phonolog-

ical representations might be shaped by their participation in reading and it opens the way toward an interesting set of issues that can be pursued in future research.

5. General Discussion

We have described the results of connectionist simulations addressing several issues concerning the representation of phonological knowledge and its role in learning to read. The model that we employed was based on the framework introduced by Seidenberg and McClelland (1989) and subsequently extended by Plaut et al. (1996). Our further extension involved using an attractor network as the phonological representation and exploring normal and atypical development of reading skill. The main results of the simulations can be summarized as follows:

1. Phonological representation. The phonological attractor architecture acquired knowledge of the segmental structure and constraints on sequences of phonemes based on exposure to phonological word forms. This knowledge allowed the model to fill in missing features and segments in realistic ways. Our primary interest in this aspect of the model was in providing a more realistic target for the orthography to phonology mapping task; the representation is limited in various ways. It would nonetheless be interesting to pursue further the use of this type of architecture as a phonological representation. The tests of categorical perception that we have described were an initial step in this direction.

2. Learning to read. The main finding here was that having a phonological attractor architecture facilitated learning the orthography to phonology mapping task; however, phonological knowledge did not have to be in place prior to reading acquisition because it could be acquired very rapidly anyway. The simulations confirm that the quality of phonological representations mainly affects the ability to generalize not the acquisition of a finite reading vocabulary, as suggested by Seidenberg and McClelland (1990) and Plaut et al. (1996).

3. Developmental dyslexia. Two types of developmental dyslexia were simulated by introducing different types of anomalies in the model. Phonological dyslexia derives from an impairment in phonological representation that has a greater effect on nonword generalization than on learning the training vocabulary. We provided analyses of why this effect obtained: degrading the phonological representation causes the orthography to phonology part of the network to overfit the training data, impairing generalization. A second type of dyslexia represents a general delay in the acquisition of word processing skills. The simulations suggest that this kind of delay can have several causes, including a shortage of computational resources, lack of experience, or failures to learn efficiently from experience. This behavioral pattern has sometimes been termed "surface dyslexia," but this is a vestige of the dual-route theory that we have abandoned because it incorrectly implies that the impairment is specific to exception words, it misses the similarity between these children's behavior and that of younger normal readers, and it does not derive from an impairment to a "lexical" route.

4. Effects of literacy on phonological representation. Finally, we presented simulations in which the acquisition of skill in translating from orthography to phonology had an impact on phonological representation itself, consistent with other evidence that the formation of segmental phonological representations may result in part from learning to read an alphabetic orthography.

Conclusions

The work we have described is part of an on-going effort to develop a general, computationally explicit account of visual word recognition, normal and atypical acquisition of this skill, and impairments that are caused by brain injury. Our research strategy is to develop models that account for important characteristics of behavior using theoretical and computational principles that are general rather than specific to the reading domain. The principles utilized in the present research are the same ones as in Seidenberg and McClelland (1989) and Plaut et al. (1996). The models have evolved as we discover more about the nature of word recognition in reading, about the properties of connectionist networks, and about the limitations of implemented models, but the theoretical framework has remained the same. The present work contributes to understanding reading acquisition and dyslexia both by providing a computationally explicit account of phenomena that had been described by others (e.g., effects of phonological impairment on reading acquisition) and by providing new insights about reading phenomena (e.g., the causes of different types of dyslexia).

References

Adams, M. (1990). *Beginning to read.* Cambridge, MA: MIT Press.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The celex lexical database (cd-rom).* (Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA)

Beauvois, M. F., & Derouesné, J. (1979). Phonological alexia: Three dissociations. *Journal of Neurology, Neurosurgery and Psychiatry, 42*, 1115-1124.

Besner, D., Twilley, L., McCann, R., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? *Psychological Review, 97*, 432-446.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford: Clarendon Press.

Bishop, D. V. M. (1992). The underlying nature of specific language impairment. *Journal of Child Psychology and Psychiatry, 33*(1), 3-66.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children.* Hove: Psychology Press.

Bradley, L., & Bryant, P. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature, 271*, 746-747.

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read - A causal connection. *Nature, 301*, 419-421.

Brown, G. D. A. (1997). Connectionism, phonology, reading and regularity in developmental dyslexia. *Brain and Language, 59*(2), 207-235.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*(2), 149-180.

Castles, A., & Coltheart, M. (1996). Cognitive correlates of developmental surface dyslexia: A single case study. *Cognitive Neuropsychology, 13*(1), 25-50.

Chase, C., & Jenner, A. R. (1993). Magnocellular visual deficits affect temporal processing of dyslexics. *Annals of the New York Academy of Sciences, 682*, 326-330.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 589-608.

Cossu, G., Rossini, F., & Marshall, J. C. (1993). When reading is acquired but phonemic awareness is not: A study of literacy in Down's syndrome. *Cognition, 46*(2), 129-138.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory* (1 ed.). New York: John Wiley & Sons, Inc.

De Weirdt, W. (1988). Speech perception and frequency discrimination in good and poor readers. *Applied Psycholinguistics, 9*, 163-183.

Di Lollo, V., Hanson, D., & McIntyre, J. S. (1983). Initial stages of visual information processing in dyslexia. *Journal of Experimental Psychology: Human Perception and Performance, 9*(6), 923-935.

Farmer, M., & Klein, R. (1996). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin and Review, 4*(2), 460-493.

Galaburda, A., & Livingstone, M. (1993). Evidence for a magnocellular deficit in developmental dyslexia. *Annals of the New York Academy of Sciences, 682*, 70-82.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* NJ: Hillsdale.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 674-691.

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32*, 401-424.

Gorecka, A. (1992). *Passive articulator features in phonology.* (Unpublished manuscript, University of Southern California)

Harm, M., Altmann, L., & Seidenberg, M. S. (1994). Using connectionist networks to examine the role of prior constraints in human language. In *Proceedings of the sixteenth annual conference of the cognitive science society* (Vol. 16, p. 392-396). Hillsdale, NJ: Erlbaum.

Harm, M. W., & Seidenberg, M. S. (1996). Computational bases of two types of developmental dyslexia. In *Proceedings of the eighteenth annual conference of the cognitive science society* (Vol. 18, p. 364-369). Mahwah, NJ: Erlbaum.

Harnad, S. (Ed.). (1987). *Categorical perception.* Cambridge: Cambridge University Press.

Hetherington, P., & Seidenberg, M. S. (1989). Is there "catastrophic interference" in connectionist networks? In *Proceedings of the 11th annual conference of the cognitive science society* (p. 26-33). Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*(1), 74-95.

Howard, D., & Best, W. (1996). Developmental phonological dyslexia: Real word reading can be completely normal. *Cognitive Neuropsychology, 13*(6), 887-934.

Hurford, D. P., & Sanders, R. E. (1990). Assessment and remediation of a phonemic discrimination deficit in reading disabled second and fourth graders. *Journal of Experimental Child Psychology, 50*(3), 396-415.

Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks, 1*, 295-307.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120*(4), 358-394.

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (1998). *Speech perception, phonology, morphology: Heterogeneity of language deficits in dyslexic children.* (Manuscript submitted for publication)

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific Language Impairment: A deficit in grammar or processing? *Trends in Cognitive Science, 2*, 240-246.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4*, 103-147.

Leonard, L. B., McGregor, K. K., & Allen, G. D. (1992). Grammatical morphology and speech perception in children with specific language impairment. *Journal of Speech and Hearing Research, 35*, 1076-1085.

Leonard, L. B., Sabbadini, L., Leonard, J. S., & Volterra, V. (1987). Specific language impairments in children: A cross-linguistic study. *Brain and Language, 32*, 233-252.

Liberman, A., Harris, K., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education, 6*(6), 8-17.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fisher, W. F. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* (p. 207-225). Hillsdale, NJ: Erlbaum.

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123*(2), 107-128.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandanavian Journal of Psychology, 21*, 159-173.

Manis, F., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Child Psychology, 66*, 211-235.

Manis, F., Seidenberg, M., Doi, L., McBride-Chang, C., & Peterson, A. (1996). On the basis of two subtypes of developmental dyslexia. *Cognition, 58*, 157-195.

Mann, V. A. (1984). Longitudinal prediction and prevention of early reading difficulty. *Annals of dyslexia, 34*, 115-136.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*, 313-330.

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception* (p. 254-283). Cambridge: Cambridge University Press.

McBride-Chang, C., Manis, F. R., Seidenberg, M. S., Custodio, R. G., & Doi, L. M. (1993). Print exposure as a predictor of word reading and reading comprehension in disabled and nondisabled readers. *Journal of Educational Psychology, 85*(2), 230-238.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 14-24.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 23). New York, NY: Academic Press.

Merzenich, M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271*(5245), 77-81.

Minsky, M., & Papert, S. (1969). *Perceptrons.* Cambridge, MA: MIT Press.

Morais, J., Alegria, J., & Content, A. (1987). The relationship between segmental analysis and alphabetic literacy: An interactive view. *Cahiers de Psychologie Cognitive, 7*(5), 415-438.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*, 45-64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323-331.

Morton, J., & Frith, U. (1993). What lesson for dyslexia from Down's syndrome? Comments on Cossu, Rossini and Marshall (1993). *Cognition, 48*(3), 289-296.

Murphy, L., & Pollatsek, A. (1994). Developmental dyslexia: Heterogeneity without discrete subgroups. *Annals of Dyslexia, 44*, 120-146.

Olson, R., Wise, B., Conners, F., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities, 22*(6), 339-348.

Olson, R. K., Forsberg, H., & Wise, B. (1994). Genes, environment, and the development of orthographic skills. In V. W. Beringer (Ed.), *The varieties of orthographic knowledge 1: Theoretical and developmental issues.* The Netherlands: Kluwer.

Patterson, K. E., Marshall, J. C., & Coltheart, M. (Eds.). (1985). *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading.* London: Erlbaum.

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30*, 473-485.

Perfetti, C. A., Bell, L., & Delaney, S. (1988). Automatic phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*, 59-70.

Plaut, D. C., McClelland, J. L., Seidenberg, M., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

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

Pollack, L., & Pisoni, D. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science, 24*, 299-300.

Pollatsek, A., Lesch, M., Morris, R., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 148-162.

Rack, J. (1985). Orthographic and phonetic coding in developmental dyslexia. *British Journal of Psychology, 76*, 325-340.

Read, C., Yun-Fei, Z., Hong-Yin, N., & Bao-Qing, D. (1987). The ability to manipulate speech sounds depends on knowing alphabetic writing. In P. Bertelson (Ed.), *The onset of literacy* (p. 31-44). Cambridge, MA: MIT Press.

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology, 48*, 270-292.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing, vol. 1.* Cambridge, MA: MIT Press.

Seidenberg, M. S. (1985). The time course of information activation and utilization in visual word recognition. In D. Besner, T. G. Waller, & E. M. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (p. 199-252). New York: Academic Press.

Seidenberg, M. S. (1992). Dyslexia in a computational model of word recognition in reading. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (p. 243-274). Hillsdale, NJ: Erlbaum.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568.

Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review, 97*(3), 447-452.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1177-1196.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 546-554.

Shankweiler, D., & Liberman, I. (Eds.). (1989). *Phonology and reading disability: Solving the reading puzzle.* Ann Arbor, Michigan: University of Michigan Press.

Share, D. L. (1995). Phonological recoding and self-teaching: *sine qua non* of reading acquisition. *Cognition, 55*, 151-218.

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*(6), 1309-1324.

Snowling, M. J. (1991). Developmental reading disorders. *Journal of Child Psychology and Psychiatry, 32*, 49-77.

Stanovich, K., Siegel, L., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89*(1), 114-127.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory and Cognition, 20*, 51-68.

Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology, 85*(2), 211-229.

Steriade, D. (1993). Closure, release and nasal contours. In M. Huffman & R. Krakow (Eds.), *Nasals, nasalization, and the velum* (p. 401-470). Academic Press.

Stone, G., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*(3), 337-359.

Tallal, P. (1980). Auditory temporal perception, phonics and reading disabilities in children. *Brain and Language, 9*, 182-198.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271*(5245), 81-83.

Tallal, P., & Piercy, M. (1973a). Defects of nonverbal auditory perception in children with developmental aphasia. *Nature, 241*, 468-469.

Tallal, P., & Piercy, M. (1973b). Developmental aphasia: impaired rate of nonverbal processing as a function of sensory modality. *Neuropsychologia, 11*, 389-398.

Tallal, P., Stark, R. E., & Curtiss, B. (1976). Relation between speech perception and speech production impairment in children with developmental dysphasia. *Brain and Language, 3*, 305-317.

Temple, C. M., & Marshall, J. C. (1983). A case study of developmental phonological dyslexia. *British Journal of Psychology, 74*, 517-533.

Treiman, R. (1992). The role of intrasyllabic units in learning to read and spell. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition.* Hillsdale, NJ: Erlbaum.

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology, 77*, 417-427.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*(4), 488-522.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392-393.

Warren, R. M. (1996). Auditory illusions and perceptual processing of speech. In N. J. Lass (Ed.), *Principles of experimental phonetics* (p. 435-466). St. Louis: Mosby.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory and Cognition, 13*(6), 557-572.

Werker, J., & Tees, R. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41*(1), 48-61.

Williams, R. J., & Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation, 2*, 490-501.