Using Parallel Distributed Processing Models to Simulate Phonological Dyslexia: The Key Role of Plasticity-related Recovery

Stephen R. Welbourne and Matthew A. Lambon Ralph

Abstract

■ PMSP96 [Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115, 1996, Simulation 4] is an implementation of the triangle model of reading, which was able to simulate effects found in normal and surface dyslexic readers. This study replicated the original findings and explored the possibility that damage to the phonological portion of the model might produce symptoms of phonological dyslexia. The first simulation demonstrated that this implementation of PMSP96 was able to reproduce the standard effects of reading, and that when damaged by removal of the semantic input to phonology, it produced the kind of frequency/consistency interactions and regularization errors typical of surface dyslexia. The second simulation explored the effect of phonological damage. Phonological damage alone did not result in a convincing simulation of phonological dyslexia. However, when the damage was followed by a period of recovery, the network was able to simulate large lexicality and imageability effects characteristic of phonological dyslexia—the first time that both surface and phonological dyslexia have been simulated in the same parallel distributed processing network. This result supports the view that plasticity-related changes should be a significant factor in our understanding of chronic behavioral dissociations. ■

INTRODUCTION

Phonological dyslexia is a disorder of reading characterized by impairment in nonword reading ability. The characteristics of phonological dyslexia are closely related to those of deep dyslexia, with the important distinction that phonological dyslexics do not make any of the semantic errors that are diagnostic of deep dyslexia. The first case of phonological dyslexia was reported by Beauvois and Derouesné (1979) who coined the term. Since then there have been numerous reports of individual cases as well as two case series (Crisp & Lambon Ralph, 2006; Berndt, Haendiges, Mitchum, & Wayland, 1996). Analysis of these shows that there is a wide continuum of reading performance both for words and nonwords. At one end there are patients whose word reading is near ceiling and have only slightly impaired nonword reading; then there are patients with relatively ''pure'' deficits whose word reading is still reasonably preserved, but whose nonword reading is almost at floor. Finally, there are the very severe cases with abolished nonword reading, and also very poor reading of words.

At first, it was thought that the only important factor in phonological dyslexia was lexicality. More recently, it has been shown that imageability/concreteness also affects word reading. Traditionally, this variable has been associated with reading performance in deep dyslexics, and most of the early reports do not associate imageability effects with phonological dyslexia. The first suggestion of this possible association comes from patient LB (Derouesné & Beauvois, 1985). Friedman (1995) also reported some cases who exhibited imageability effects; however, it was not until the most recent case-series study (Crisp & Lambon Ralph, 2006) that it became clear that the occurrence of imageability effects in phonological dyslexia was widespread. In that study, all except one of the 12 patients (the mildest) were significantly more accurate when reading high imageability words. This gradual realization that many ''deep dyslexic'' symptoms also occur in phonological dyslexia is part of an emerging trend in which deep and phonological dyslexia are viewed as points on a continuum rather than as separate disorders (Crisp & Lambon Ralph, 2006; Friedman, 1996).

The vast majority of phonological–deep dyslexic patients also present with other associated phonological impairments, an observation which has led many to attribute phonological dyslexia to a generalized phonological impairment (e.g., Farah, Stowe, & Levinson, 1996; Patterson & Marcel, 1992). Indeed, we are only aware of three potential exceptions to this rule: LB (Derouesné & Beauvois, 1985), RR (Bisiacchi, Cipolotti, & Denes, 1989), and, more recently, RG (Caccappolo-van Vliet,

University of Manchester, UK

Miozzo, & Stern, 2004). All of these cases had only mild nonword reading impairments, and as nonword reading is one of the more difficult phonological tasks, we would argue that their apparent lack of phonological impairment is merely a reflection of the fact that tests for other phonological deficits were not sufficiently sensitive to identify such a mild impairment. Although the current simulations cannot speak directly to this issue (they are restricted to modeling performance in reading tasks), they respect the notion that phonological dyslexia is the result of generalized damage to the phonological system. Accordingly, the expectation for this study is that lexicality and imageability effects should emerge from generalized damage to the phonological portion of a neural network that processes words using interconnected orthographic, phonological, and semantic systems.

Previous models of reading have focused on modeling surface rather than phonological dyslexia (Plaut, McClelland, Seidenberg, & Patterson, 1996; Patterson, Seidenberg, & McClelland, 1989); as yet, there has been no account of acquired phonological dyslexia within a connectionist framework. Harm and Seidenberg (1999) have explored the phenomenon of developmental phonological dyslexia with some success. They trained a single-route network in two stages. First, they trained the phonological portion of the network so that it learned the phonological representations of the words in the training corpus. They then trained the network to read, interleaving this new training with continued exposure to phonological only trials from the first phase of training. To model developmental phonological dyslexia, they damaged the phonological portion of the network after the first stage of training. Although they successfully modeled varying severities of developmental dyslexia, none of their simulations were intended to produce, nor did they come near to producing the very large lexicality effects found in cases of pure acquired phonological dyslexia. In fact, there are no reported parallel distributed processing (PDP) models of acquired phonological dyslexia that produce lexicality effects of the required magnitude. (Harm & Seidenberg, 2001 modeled acquired phonological dyslexia, but the focus of the article was on orthographic influences on RTs, and lexicality effects were not reported.) It is difficult to come to any definite conclusion as to why this should be, but we suspect that a key factor is the difficulty in obtaining large lexicality effects. Attempting to model large performance dissociations as a result of damage to a PDP network can be a very frustrating task. Damage to these networks tends to affect all processing tasks to a similar degree (although this will also depend critically on the architecture of the network and the nature of the tasks involved). This was certainly the case with early attempts to model surface dyslexia (Patterson et al., 1989). PMSP96 successfully captured surface dyslexia, but it achieved this without directly damaging the network. It modeled surface dyslexia by encouraging a division of labor between the direct activation of phonology by orthography and the additional constraint supplied by word meaning (semantics). Semantic contributions were modeled by applying an external input to ''push'' the output of the phonological units toward their targets. Semantic damage could then be modeled by the removal of this input, thus producing surface dyslexia.

This article adopts an additional factor when modeling the effects of brain damage (Welbourne & Lambon Ralph, 2005a). Under this approach, patient performance is assumed to be the combined result of brain damage and plasticity-related recovery. The period of recovery (corresponding to the period of spontaneous recovery in patients) is critical because it allows the brain to reoptimize its remaining connections, thus allowing it to make the best use of what resources it has left. Therefore, recovery after brain damage may be, at least in part, attributable to synaptic weight changes. If accurate performance depends on the pattern of synaptic weights (as it does in PDP models), then it seems reasonable to assume that the disconnection of some weights, after brain damage, will not leave the remaining synapses optimally configured to perform the task. Further, provided that there is an optimization process by which the synaptic weights can change (learning), then it seems inevitable that some of the recovery that we observe in patients after brain damage must be attributable to synaptic change. This kind of mature synaptic plasticity has been studied mostly in the context of cortical sensory maps (for a review, see Buonomano & Merzenich, 1998), and it is clear that these maps are capable of undergoing extensive modification, presumably as a result of some learning process operating at the synaptic level. Further evidence for this comes from studies that demonstrate an altered pattern of activation in recovered aphasic patients—often involving a shift of processing from the left to right hemisphere (e.g., Blank, Bird, Turkheimer, & Wise, 2003; Weiller et al., 1995). Leff et al. (2002) provided a particularly strong example of this. They looked at the relationship between rate of speech presentation and cerebral activity. In normal subjects, this function showed a left–right asymmetry in the posterior superior temporal sulcus with the left hemisphere showing a steeper response curve. However, in a group of recovered aphasic patients with lesions encompassing the left posterior superior temporal sulcus, the steep response curve had transferred to the homologous area in the right hemisphere. The key strength of this study is that it demonstrates a shift in activation response curves (rather than just activations), suggesting that the activation in the homologous area is performing a similar function to the original left hemisphere activations. The authors attribute this transfer to synaptic reorganization—exactly the process that we have introduced into the model.

Our previous simulations (Welbourne & Lambon Ralph, 2005a, 2005b) demonstrated that at least some

behavioral dissociation exhibited by patients might result from this plasticity-related recovery. Plaut (1996) also investigated the effects of retraining after damage. Using a network linking orthography to semantics, he showed that retraining could produce a shift in the error patterns from deep to phonological dyslexia. The purpose of the present study was to apply our new methodology to a suitable reading model in an attempt to simulate acquired phonological dyslexia. We selected Simulation 4 from Plaut et al. (1996) as the most appropriate for our purposes. This model consists of a feedforward network trained on a set of monosyllabic words, with the training weighted by the square root of word frequency. Input to the phonological units comes, in a great part, direct from the orthographic units, but is supplemented by an external ''semantic'' contribution. In their article, Plaut et al. demonstrated that removal of this semantic contribution resulted in typical surface dyslexic reading patterns. We speculated that damage to the phonological side of the network, followed by a suitable period of retraining, might result in typical phonological patterns of impairment.

METHODS

The architecture, training, and representations used in these simulations were modeled on those used by Plaut et al.'s (1996) Simulation 4.[1] Each of these key features is summarized below. Figure 1 shows the architecture of the network that was used throughout this set of simulations. There were three sets of units: 105 grapheme units, 100 hidden units, and 61 phoneme units. The input layer was connected to the hidden layer, with a probability of 40%, and the hidden layer was connected to the output layer, with a probability of 80%. This sparse connection is a modification from the original simulation where every layer was fully connected to the next layer up. The purpose of this modification was to reduce the competence of the phonological part of the model so that word reading would require a division of labor between semantic and phonological systems. Plaut et al. achieved the same result by using a very high value of weight decay in the phonological part of the model. The current method was chosen in preference because it is a more realistic description of synaptic connectivity in the human brain: If every neuron in the brain were connected with every other, the size of the brain would need to increase to a sphere of radius 10 km (Plaut, 2002; Nelson & Bower, 1990). Histological studies in mice and rats (Young, Scannell, & Burns, 1995) have demonstrated that the probability of neuronal connection is extremely dependent on distance. Neurons that are in close proximity are connected with a high probability using intracortical connections. Neurons that are in different systems may still be connected intercortically via white matter connections, but the density of the connection projection will be much sparser. Hence, in our model, connections between orthography and phonology are less likely than connections occurring within the phonological system.

The activity level of each unit was set to vary between 0 and 1 as a nonlinear (logistic) function of the unit's total input. The initial weights on the connections were set to random values between −0.1 and +0.1. The network was then trained using the standard backpropagation learning algorithm with momentum enabled only if the gradient of the error slope was less than 1. Cross entropy was used as the error function as in PMSP96. The learning rate for the network was set to 0.05 and the momentum was 0.9.

It should be noted that this learning procedure differs slightly from that used in PMSP96, where each connection was allowed to modify its own learning rate in a procedure known as delta bar delta learning (Jacobs, 1988). The procedure used here, however, is computationally simpler and results in very similar performance to that found in PMSP96.

Orthographic and Phonological Representations

The network used the same representations as PMSP96, designed to minimize the dispersion problem (a problem

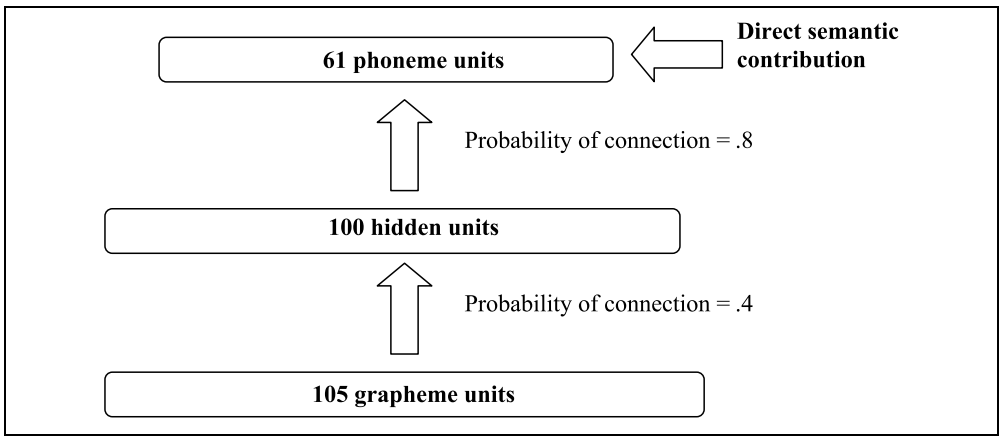

Figure 1. Network architecture.

whereby the same letters in different positions cannot mutually benefit from phonological knowledge). These representations divide each word into three parts (onset, vowel, and coda) and then use specific units to code for particular graphemes or phonemes occurring within each part. In addition, the phonological onset and coda are further divided into groups of mutually exclusive phonemes so that when reading off the unit activations only the most active member of each group is a candidate for inclusion in the output phoneme string. Table 1 shows the representation scheme used in this simulation (phonological subgroups are separated by extra spaces). In general, words are coded from left to right so that if more than one unit is active in the onset or coda, then the output is read in the order that they appear in the table. The only exception to this occurs for the phonemes pairs p–s, k–s, and t–s, which can occur either way round in the phonological coda. To cater for this, special units k–s, p–s, and t–s are used to determine the order. If both s and p are active, then they are taken in the order s–p, unless the p–s unit is active, in which case the order is reversed.

Interpretation of Output to Allow for Omissions

In the original PMSP96 model, it was not possible for the network to make omission errors. At the very least, every input would generate an output consisting of a single vowel sound (from the most active vowel unit). Although this is not a serious problem, where the target is surface dyslexia, it acquires much more significance when attempting to model phonological dyslexia, where omissions constitute one of the largest categories of error. In order to avoid this problem, we adopted a slightly different interpretation of the network's output. We decided to adopt a range of activations that would be classed as omissions. For the onset and coda, this range was set to 0.35–0.65, whereas for the vowel units, it was set to 0.15–0.25 (this maintains the distinction made in the original model where vowel units require less activation to be considered as ''ON'').[2] Under this scheme, units with activations above the top end of the range are regarded as ''ON,'' whereas units whose activations fall below the bottom of the range are regarded as ''OFF.'' Any units with activations falling within the range are considered to be in an undecided state. The presence of a single output unit in this undecided state is sufficient to flag the whole response as an omission. This new method of interpreting the network's output was used in all of the simulations including the replication of the original PMSP results. The exact settings for these parameters is not critical, using a wider range results in more omission and fewer nonword errors, narrower ranges give more nonword and fewer omission errors.

Table 1. Orthographic and Phonological Representations

Orthographic units

Onset	Y S P T K Q C B D G F V J Z L M N R W H CH GH GN PH PS RH SH TH TS WH
Vowel	E I O U A Y AI AU AW AY EA EE EI EU EW EY IE OA OE OI OO OU OW OY UE UI UY
Coda	H R L M N B D G C X F V J S Z P T K Q BB CH CK DD DG FF GG GH GN KS LL NG NN PH PP PS RR SH SL SS TCH TH TS TT ZZ U E ES ED

Phonological units

Onset	s S C z Z j f v T D p b t d k g m n h l r w y
Vowel	a e i o u @ ^ A E I O U W Y
Coda	r l m n N b g d ps ks ts s z f v p k t S Z T D C j

Imageability Ratings

Imageability ratings for words in the corpus were obtained from the MRC Psycholinguistic database and from Cortese and Fugett (2004). Between them these sources provided ratings for 2719 of the 2998 words in the corpus (1529 words had ratings from both sources). For purposes of this study, both these ratings were converted to z scores and averaged if necessary. Words without an imageability rating were given an average imageability value (z score = 0).

Semantic Input

Semantic input to the phonological units was provided, such that it tended to push the phonological units towards the correct activations. Throughout training, the strength of this contribution was gradually increased to mimic the effect of learning. The strength of this input at any given developmental stage was modulated by word frequency and imageability according to the following formula:

$$\mathrm{INPUT}_{\mathrm{Semantic}} = \left(\frac{0.5}{1 + e^{-(1.14\mathrm{Log}(\mathrm{Freq}+2)-1)}} + \frac{0.5}{1 + e^{-(\mathrm{Image}_z+1.5)}} \right) \times \mathrm{Modulation}$$

Equation 1. Formula for calculating the semantic input to phonological units

Frequency was taken from Kuçera and Francis (1967) (for consistency with PMSP96) and imageability z score was calculated as above. (The constants in this formula were selected to provide a sensible distribution across

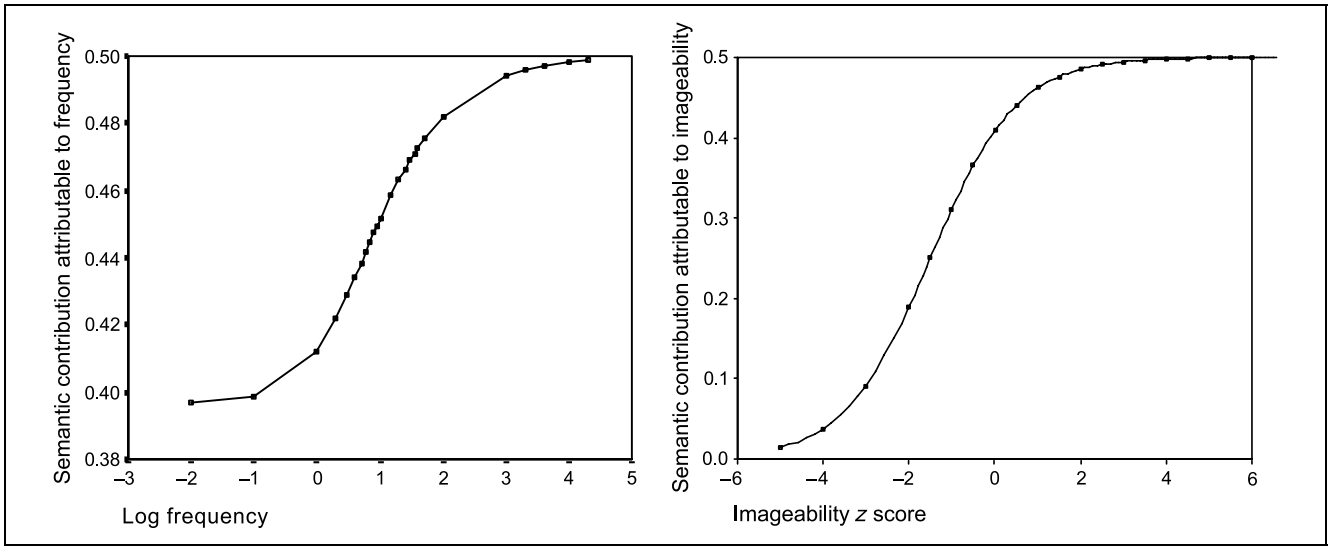

Figure 2. Contribution to semantics from variations in frequency and imageability.

the frequency and imageability values in the corpus with more of the variation originating in imageability.) Figure 2 shows how this semantic input varies with frequency and imageability, whereas Figure 3 shows the distribution of the semantic contribution for items in the training corpus.

Over the course of development, the total semantic input was modulated by an epoch-dependent stepwise modulation factor that varied from 0.6 to 4.8 in steps of 0.6, where a step occurred after every 200 epochs. This is slightly different to the continuous function used in PMSP96, but the key point is that, through training, semantics makes an increasing contribution to the activation of the phonological units.

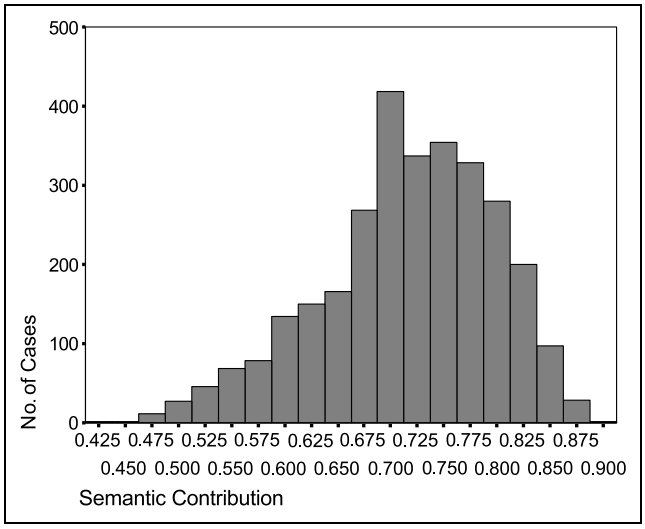

Figure 3. Distribution of semantic contribution for all items in training corpus.

In the case of the nonwords, Plaut et al. did not provide any semantic contribution. We adopted a different approach for the following reason: In the brain, the connections between O and S (either direct O → S or indirect O → P → S) cannot be selectively turned off for nonwords. Hence, nonwords will generate some kind of activation across the semantic units, which will, in turn, contribute to the activation of phonological units. This nonword semantic activation will not correspond to any known semantic targets (except in the case of lexicalization errors); rather, it will represent some kind of average semantic activation for all the visually similar words. This will result in a contribution from semantics to phonology that is effectively random noise. Accordingly, for nonword reading, semantic input was randomly added to the phonological units, where the input for each unit varied between −0.5 and +0.5 modulated by the same function as for the real words.

Training Procedure

The network was trained using full batches with the same corpus of 2998 monosyllabic words used in PMSP96. The square-root frequency (Kuçera & Francis, 1967) of each word was used to scale the error derivatives for the purposes of backpropagation. This has the same effect as using frequency to determine the probability of a word being presented for training; however, it has the considerable advantage that every word can still be presented once every epoch, thus considerably compressing the required training time (see Plaut et al., 1996 for a fuller discussion of this issue). To eliminate the possibility that the results might be a consequence of one particular set of initial weights, the network was trained 10 times; each time using a different random set of weights as the

starting point. These 10 trained networks then formed the starting point for further investigations.

Testing Procedure

The performance of the network was tested every 10 epochs throughout the training period. Seven sets of test stimuli were used in this testing process:

1. High-frequency regular words (n = 24) (BEST, BIG, CAME. . .)
2. High-frequency irregular words (n = 24) (ARE, BOTH, BREAK. . .)
3. Low-frequency regular words (n = 24) (BEAM, BROKE, BUS. . .)
4. Low-frequency irregular words (n = 24) (BOWL, BROAD, BUSH. . .)
5. Regular nonwords (n = 43) (BEED, BELD, BINK. . .)
6. High imageability words (n = 24) (BANK, BED, BLUE. . .)
7. Low imageability words (n = 24) (BEAU, DRAB, FATE. . .)

The regular and irregular words were taken from Taraban and McClelland (1987) and were matched across groups for frequency. The regular nonwords were taken from Glushko (1979) and were created by changing the onset of an existing regular word. These stimuli are the same as those used in PMSP96 so that it is possible to make a direct comparison of results.

The high and low imageability word sets were constructed for the purposes of this simulation. Low imageability words (imageability rating 200–400) were selected from the training corpus and matched, pairwise, on frequency with high imageability words (imageability rating 500–700) also selected from the corpus. A list of these words is provided in Appendix A.

Classification of Errors

Where the network's response was incorrect, it was classified into one of five possible error types: omissions, nonwords, visual/phonological, regularizations, unrelated lexical responses.[3] As we have already indicated, omissions were recorded whenever the activation of any of the output units fell into a predefined range. Nonword errors occurred where the network produced a valid output which did not correspond to any of the words in its training corpus. Visual/phonological errors were recorded where the network produced an incorrect word that contained at least 50% of the correct target phonology (these words were also visually similar to the target). Regularization errors could only occur when the target word was irregular. They were calculated by matching the actual output of the network against a list of possible regularizations for each word. For most words, this list consisted of just one possible regularization, but some words (e.g., FLOOD or LOSE) can be regularized in two different ways. The list of pronunciations treated as regularizations was taken from those used in PMSP96. Finally, unrelated lexical responses were recorded where the output corresponded to a word from the training corpus that contained less than 50% of the correct phonology.

Simulation Structure

Initially, 10 versions of the network (each with a random set of initial weights) were trained until performance reached asymptote (2000 epochs). These 10 networks then formed the starting point of two additional simulations, exploring the effects of semantic and phonological damage. Semantic damage was simulated by decreasing the strength of the ''semantic'' contribution to phonology while simultaneously adding increasing amounts of noise. Phonological damage was simulated by lesioning links into the hidden layer of phonological units while simultaneously adding noise to the output of those units. Of course, in this network, it is not possible to differentiate between damage to the O → P pathway and damage to the phonological system, but in view of the overwhelming association between phonological dyslexia and other phonological impairments, it seems more parsimonious to interpret this damage as generalized damage to the phonological system. The addition of noise served two purposes: (1) it helped to ensure that nature of the damage was, as near as possible, that of a generalized phonological impairment; (2) it meant that, in addition to damaging the network's performance, we were also damaging its ability to relearn. This is likely to be the case in the human brain where learning and representation are intrinsically linked. After phonological damage, the network was allowed to ''recover'' for 200 epochs of further training.

RESULTS

Initial Training

By epoch 2000, the network had reached asymptote performance for all of the stimuli sets except nonwords, which reached asymptote sooner (epoch 300). At this point, the network correctly pronounced all of the words in its corpus, including all of the homographs. This is slightly better than the performance achieved by PMSP96, which was 99.7% accurate in word reading. For nonword reading, the model was correctly reading 90.5% of the regular nonwords. This is not as good as the 96.5% achieved by PMSP96, but it is still close to human performance, which averages 93.8% (Glushko, 1979).

It was important to verify that this model could replicate the standard frequency/consistency interaction found in the naming latencies of normal human populations (e.g., Seidenberg, 1985; Seidenberg, Waters,

Barnes, & Tanenhaus, 1984). Error scores from the network at epoch 2000 were submitted to a 2 × 2 analysis of variance, where frequency and consistency were treated as between-group variables. This confirmed that there was indeed a significant frequency/consistency interaction [$F(1, 1916) = 238, p < .001$]. In addition, there were significant main effects of both frequency [$F(1, 1916) = 306, p < .001$] and consistency [$F(1, 1916) = 522, p < .001$]. Figure 4 illustrates this interaction and it is clear that the effect of frequency was almost completely modulated by consistency. For irregular words, low frequencies resulted in much higher error scores, but for regular words, there was almost no effect of frequency. This is consistent with the standard effect found in human reading latencies and with the results found for PMSP96.

In addition to standard effects of consistency and frequency, one might also expect to see an effect of imageability (Strain, Patterson, & Seidenberg, 1995) with high imageability items having lower error scores than low imageability ones. To test this, error scores from the high and low imageability word sets were compared. The mean error score for high imageability items was 0.0082 (SD = 0.013), whereas the mean error score for the low imageability items was 0.0223 (SD = 0.0416). Submitting these scores to a t test revealed that there was, as predicted, a significant difference ($t = -7.08$, $df = 570$, $p < .001$).

Simulation 1: Effect of Semantic Damage—Replication of PMSP96

Before investigating the possibility that this model could simulate the symptoms of phonological dyslexia, it was important to verify that, like PMSP96, it was capable of replicating the symptoms of surface dyslexia. Surface dyslexia is characterized by poor reading of low-frequency exception words, coupled with accurate reading of nonwords. Errors made in reading irregular words tend to be regularizations or LARCs (legitimate alternate reading of components; Patterson, Suzuki, Wydell, & Sasanuma, 1995); for example, reading PINT to rhyme with MINT. To mimic the effect of semantic damage, we gradually reduced the strength of the semantic contribution while simultaneously adding noise to it. This was achieved by decreasing the strength of the modulation factor from 4.8 to 0 in 10 steps of 0.48 while simultaneously adding increasing amounts of Gaussian noise with a standard deviation increasing in 10 steps of 0.75.

Figures 5 and 6 show the results of this simulation; Figure 5 gives the accuracy rates and Figure 6 shows the percentage of responses to irregular stimuli that were regularizations. For clarity, the regular high-frequency, high imageability and low imageability word sets have been omitted—performance on these word sets is very similar to that for low-frequency regular words. Low-frequency irregular words are the most affected by the semantic lesions, with performance dropping to 47% for the worst damage. At this point, performance on high-frequency irregular words is reduced to 83%, whereas accuracy rates on all other word sets fall between 90% and 93%. Note that for nonwords this represents a slight improvement on the undamaged performance. Figure 6 shows how the proportion of responses that are regularizations increases with the degree of semantic lesion. Again, low-frequency words are the most affected. For the worst level of damage, 34% of responses to low-frequency irregular words and 7% of responses to high-frequency irregular words are regularizations. This pattern of results is consistent with that found in surface dyslexic patients (Woollams, Lambon Ralph, Plaut,

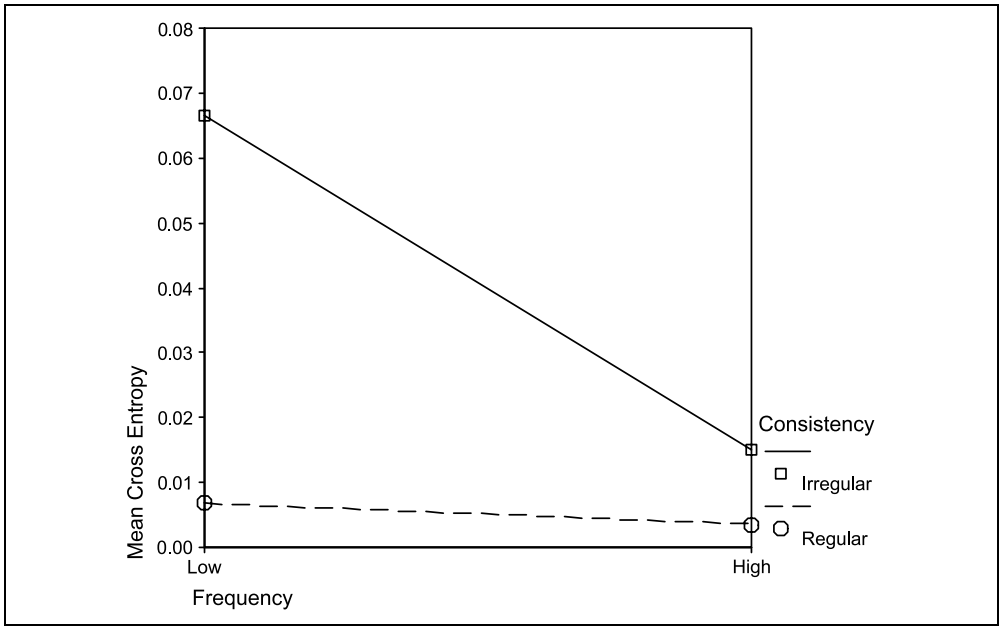

Figure 4. Interaction between frequency and consistency.

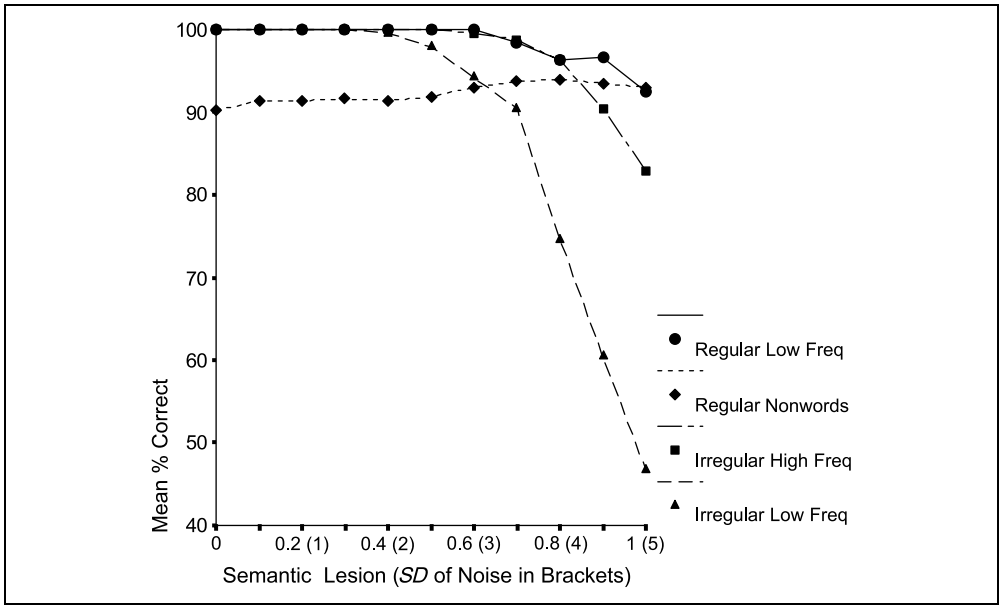

Figure 5. Effect of removal of semantic input on accuracy rates.

Hodges, & Patterson, in press) and with the results of PMSP96 Simulation 4.

Simulation 2: Effect of Phonological Damage

Phonological damage was simulated by lesioning the links between input and hidden layers while simultaneously adding noise to the output of the hidden layer. Twenty levels of severity were tested with lesions ranging from 5% to 100% and noise ranging from 0.05 to 1, both in equal intervals. For each level of severity, two random lesions were administered to each of the 10 networks so that each data point represents the average of 20 trials.

Figure 7 shows the effect of this kind of damage on network performance. Clearly, there is some degree of lexicality effect; throughout the range of damage severity, nonwords are read less accurately than any of the word sets. However, the size of the effect is relatively small, with the difference ranging between 20% and 40% so that there is no point at which the network's performance resembles that of pure phonological dyslexia. Throughout the range of damage severity, there is a small advantage for high imageability words versus low imageability words. However, the magnitude of this difference is generally smaller than one would expect to see in patients with phonological dyslexia. In summary, phonological damage on its own does not result in

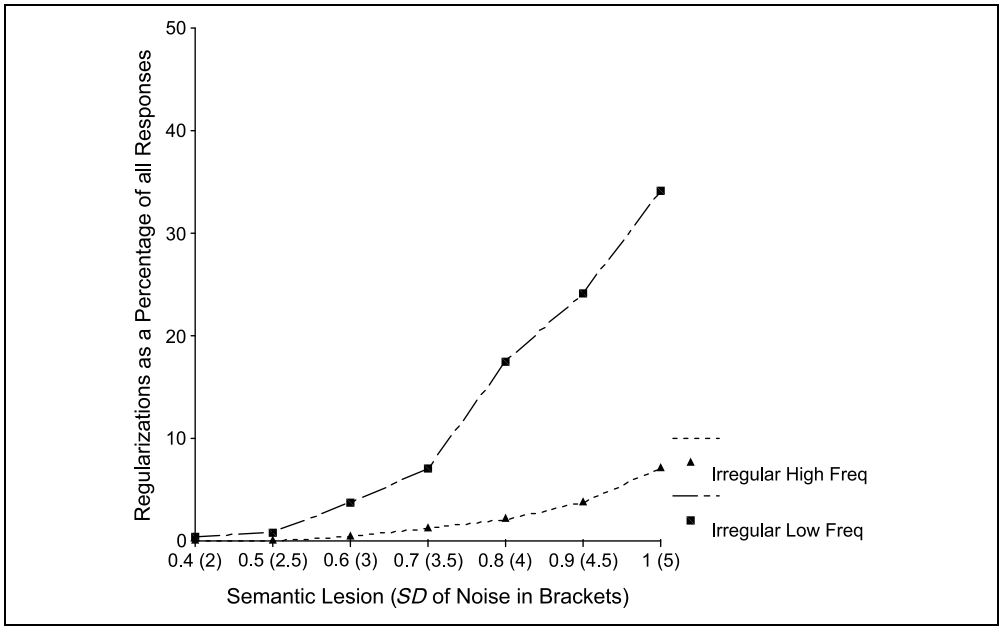

Figure 6. Effect of removal of semantic input on regularization rates.

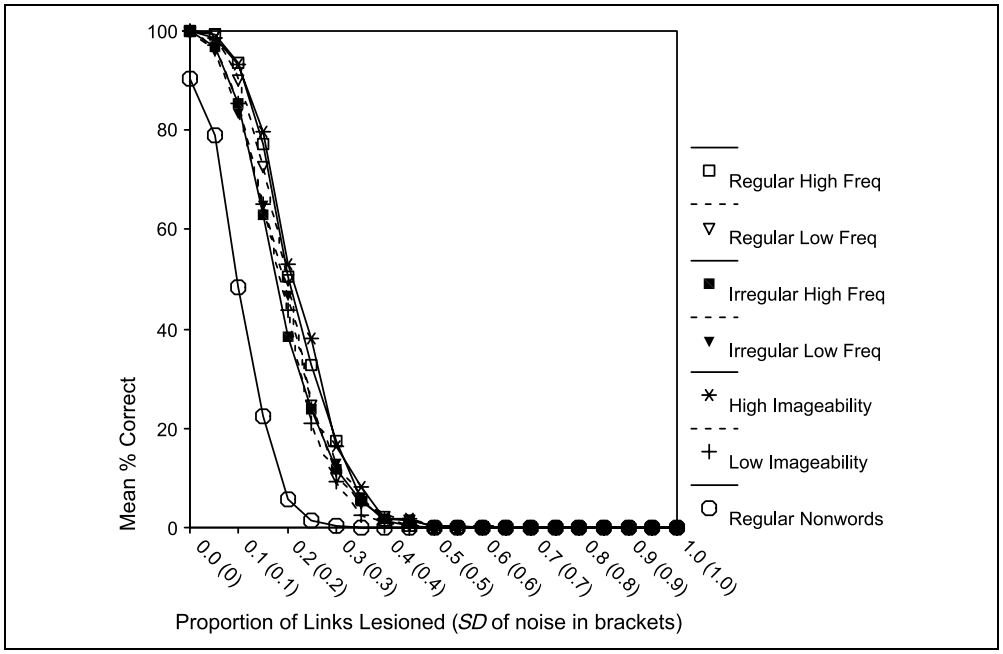

Figure 7. Effect of phonological damage.

performance that is typical of phonological dyslexia: Although it produces data which are qualitatively similar to phonological dyslexia, it cannot produce sufficiently wide ranges of lexicality or imageability effects.

So far, the results are in accordance with our expectations: Phonological damage on its own does not produce sufficiently large dissociations to make a convincing case that the network is modeling phonological dyslexia. This is an analogous situation to that found when trying to model surface dyslexia with a single-route network (Welbourne & Lambon Ralph, 2005a; Plaut et al., 1996; Patterson et al., 1989), where damage alone did not result in sufficiently large dissociations. The key question is whether the addition of a period of recovery will change this pattern of results. To test this, we selected three levels of damage severity (15%, noise *SD* = 0.15; 30%, noise *SD* = 0.3 and 70%, noise *SD* = 0.7) and allowed the network to recover for 200 epochs by re-exposing it to the original learning environment. Figure 8 shows the results of this investigation, in terms of reading accuracy, for the three levels of damage severity. At the most severe level, nonword reading is abolished while word reading

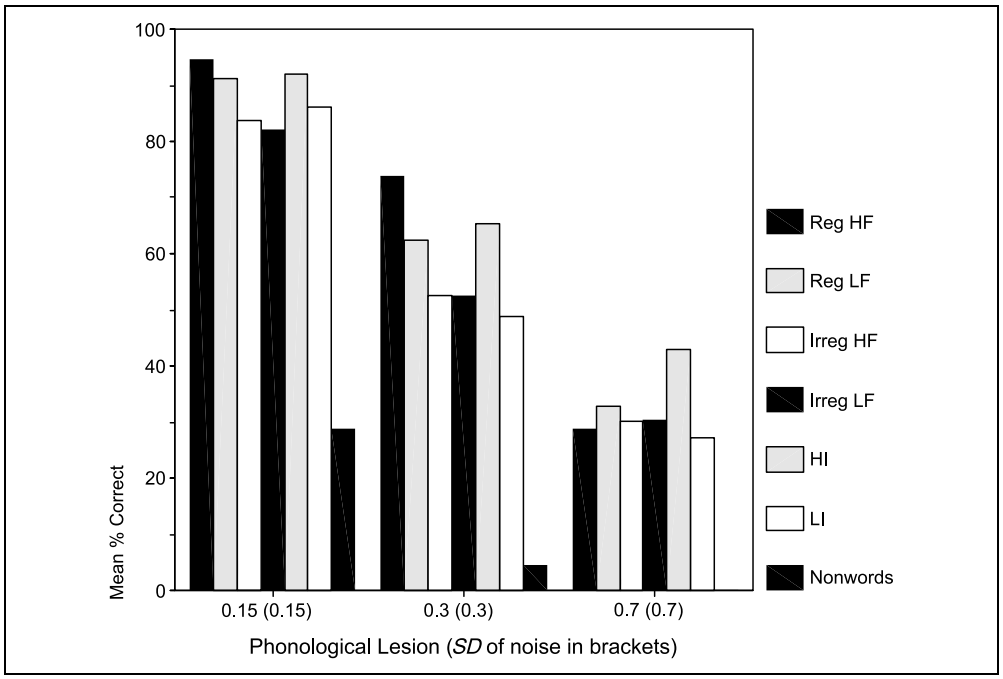

Figure 8. Performance after phonological damage and recovery.

accuracy varies between 27% and 43%, depending on the stimuli set. The high imageability and regular high-frequency words are read with the highest accuracy. This pattern of results is exactly what one might expect to see in a rather severe case of phonological dyslexia.

For the medium and mild levels of damage, the pattern of performance is similar to that for severe damage except that it is centered around progressively higher mean scores: In the case of moderate damage (30% of links removed), scores range from 49% to 73%, whereas for milder damage (15% of links removed) they range from 82% to 94%. In all cases, irregular and low imageability words are read less accurately than regular and high imageability words. Nonword reading is seriously impaired for all levels of damage, with overall level of nonword reading accuracy decreasing with increasing damage severity. Even at mild levels of damage, nonword reading accuracy is still only 29%. For the medium level of damage, the network's performance resembles that of a ''pure'' phonological dyslexic patient: Word reading performance is relatively preserved (over 70% for regular high frequency words), and at 4% accuracy nonword reading is almost abolished. For mild damage, the performance of the network is similar to that of a mild case of phonological dyslexia. In addition to the expected effects of lexicality and imageability, there appears also to be a small effect of consistency, such that for the mild and moderate levels of damage, the network is slightly less accurate when reading inconsistent items.

To confirm the significance of the apparent effects of lexicality, imageability, and consistency, we submitted the results to a series of t tests: Lexicality was tested by comparing performance on high-frequency regular words with performance on regular nonwords; imageability was tested by comparing performance on the high and low imageability word sets; consistency was tested by comparing the low-frequency regular and irregular word sets. To compensate for the increased possibility of Type 1 errors due to multiple comparisons, a more severe significance criterion of 0.01 was adopted. Table 2 shows the results of these tests and it is immediately obvious that all of the comparisons for lexicality and imageability produce highly significant differences (all ps $< .001$). At mild levels of damage, there is also a significant effect of consistency, however, this effect reduces so that it is only marginal for moderate damage and disappears completely for severe damage.

Table 2. Lexicality, Imageability, and Consistency Effects at Three Severities of Damage

Effect Type	*Lesion Severity*	*Mean Difference*	*MSE*	*t Value*	*df*	*p*
Lexicality: Reg HF–NW	*Mild*	65.75	2.1	31.8	19	<.001
	Moderate	69.3	2.0	34.5	19	<.001
	Severe	28.8	1.4	19.9	19	<.001
Imageability: HI–LI	*Mild*	5.83	1.8	5.1	19	<.001
	Moderate	10	3.4	2.9	19	<.001
	Severe	15.8	1.3	12.5	19	<.001
Consistency: LFR–LFI	*Mild*	9.2	1.8	5.1	19	<.001
	Moderate	10.0	3.4	2.9	19	.009
	Severe	2.5	2.9	.85	19	.405

Reg = Regular; HF = high frequency; NW = nonwords; HI = high imageability; LI = low imageability; LFR = low-frequency regular; LFI = low-frequency imageability.

Comparison of Lexicality Effects in Patients and Model

So far, the comparison with patient data has been of a qualitative nature. We have identified three prototypical patient profiles and demonstrated that the model can reproduce these patterns. However, we have not, as yet, quantified this comparison. To achieve this, we combined data from the two extant case-series studies (Crisp & Lambon Ralph, 2006; Berndt et al., 1996). These patients were all tested more than 3 months postonset (mean = 52 months), so for all of the patients there had clearly been ample time for the operation of plasticity-related recovery processes. For each patient, we took the total percentage accuracy on all words in the study and the total percentage accuracy on all nonwords in the study (note that the exact composition of these lists was not the same for both studies). Patients were then divided into three groups corresponding to the three prototypical profiles already identified. Patients with lexicality effect sizes of more than 60% were identified and placed in the ''pure'' group ($n = 5$). The remaining patients were split into two groups according to their word reading ability: Those with accuracy rates in excess of 70% were placed in the mild group ($n = 6$), with the remainder being placed in the severe group ($n = 12$). These groups were then matched on word reading accuracy with the model (summing across all word types), both immediately after damage alone and after damage and then retraining. To obtain accurate fits, different levels of damage were required for the retraining and no retraining cases. Figure 9 shows the results of this matching process. For the damage-only case, damage levels of 5%, 10%, and 20% were used, whereas for the damage and then retraining case, the best fitting levels were 10% 15%, and 40%, corresponding to the mild, pure, and severe patient groups, respectively. t Tests revealed no significant difference between any of the word reading scores of the patient groups and the matching scores from the model (all ps $> .05$). Having established matched word reading performance, we then compared the size of the lexicality effects across

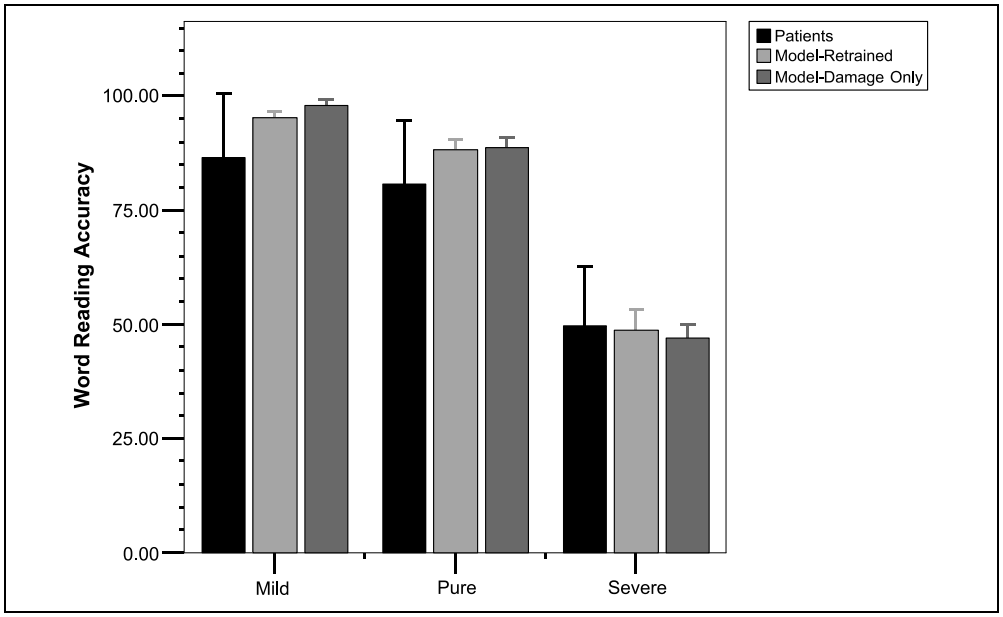

Figure 9. Fit between patients and model on word reading accuracy.

the groups (Figure 10). The lexicality effect size from the patients and the model appears very similar in the case where the model has been allowed to recover, but where no recovery has been allowed, the models' lexicality effect sizes seem substantially smaller than the patients for both the mild and the pure groups. *t* Tests confirm this impression; there is no significant difference in the size of lexicality effects between the patients and the retrained model (all *p*s > .05). However, for the damage-only case, there is a significant difference between the patients and the model in both the mild and the pure cases (all *p*s < .001).

These results are as predicted; in the damage-only case, the model is not able to produce sufficiently large lexicality effects to capture the patient data adequately. However, when the possibility of plasticity-related recovery is introduced, the model can capture the full range of data.

Analysis of Errors

The previous section has shown how the model can simulate the full range of lexicality effects that one would expect to see in patients. However, it is also important

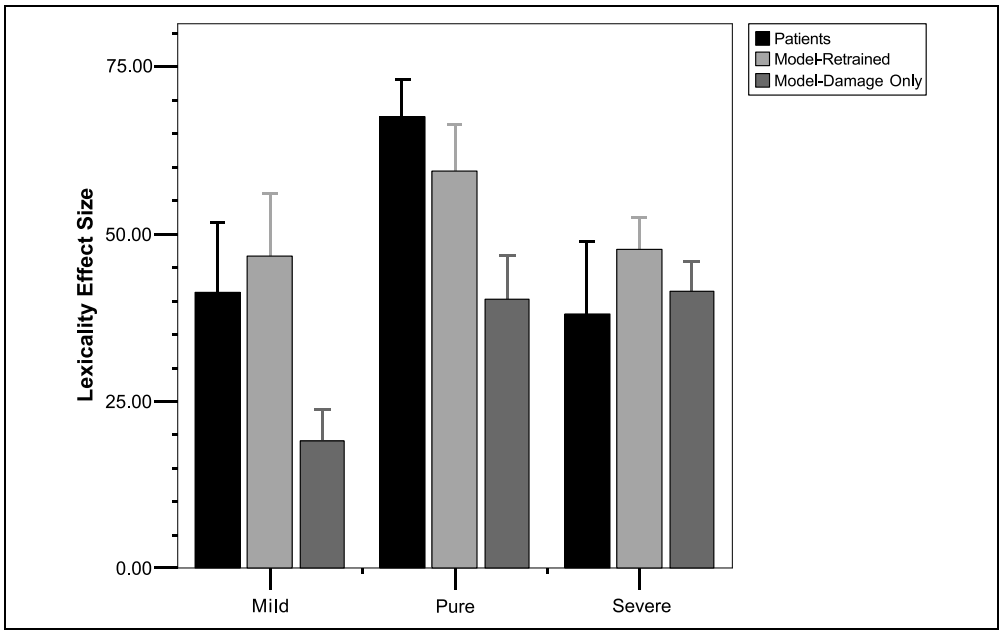

Figure 10. Comparison of lexicality effects across patients, damage-only model and model with damage + retraining.

to explore how this kind of damage affects the variety of errors that the network produces. The pattern of errors that one expects to see in patients suffering from phonological dyslexia is very different from that observed in surface dyslexic patients. Instead of regularizations, the dominant error responses are usually omissions, nonwords (not regularizations), and visual/phonological errors. Data from the Crisp and Lambon Ralph (2006)[4] case series of 12 patients suggest that omissions are the most common form of errors in these patients, followed by nonwords and then visual/phonological errors, with these three categories between them forming the bulk of all reported errors. Figure 11 shows how the frequency of error responses in the network varies with increasing levels of phonological damage. The results are strikingly similar to the expected phonological dyslexic pattern. The most common error type is omissions followed by nonwords and then visual/phonological errors. In contrast to the simulation of semantic damage, regularization errors form only a small proportion of responses, and this proportion actually decreases with increasing lesion severity.

DISCUSSION

Two simulations were conducted using a network architecture similar to PMSP96 (Simulation 4). The first simulation demonstrated that our implementation performs similarly to PMSP96 in that it can reproduce the cardinal features of normal reading, as well as the symptoms of surface dyslexia. The second simulation explored the possibility that damage to the phonological portion of the model would lead to performance resembling that found in phonological dyslexia. Damage alone did not produce sufficiently large lexicality effects, but if the network was allowed a period of plasticity recovery, then a full range of lexicality effects could be modeled, coupled with the imageability effects that are characteristic of phonological dyslexia. This is the first time that such large lexicality effects have been modeled in a network which also has the capacity to learn. Moreover, it is the first time that simulations of surface and phonological dyslexia have been produced from the same connectionist architecture.

These results pose two important questions: (1) What are the critical components in these simulations that are essential to successfully modeling phonological dyslexia? (2) How do these results mesh with those reported by Welbourne and Lambon Ralph (2005a)?

Two features of these simulations seem likely to have significantly contributed to their success in modeling phonological dyslexia. The first of these is the inclusion of a period of plastic recovery after damage. Welbourne and Lambon Ralph (2005a) found that including a period of recovery was helpful when modeling surface dyslexia because it magnified the effect of small preexisting processing biases into large performance dissociations. Exactly the same effect is produced in these simulations, but this time the biases are toward lexicality and imageability effects rather than a frequency/consistency interaction. The fact that in two cases this manipulation contributes to different behavioral dissociations, both matching known patient behavior, is indicative that these plastic recovery processes play a critical role in determining the behavior of chronic stage patients. This new evidence from modeling neatly complements the imaging studies (Blank et al., 2003; Leff et al., 2002; Weiller et al., 1995) that find altered

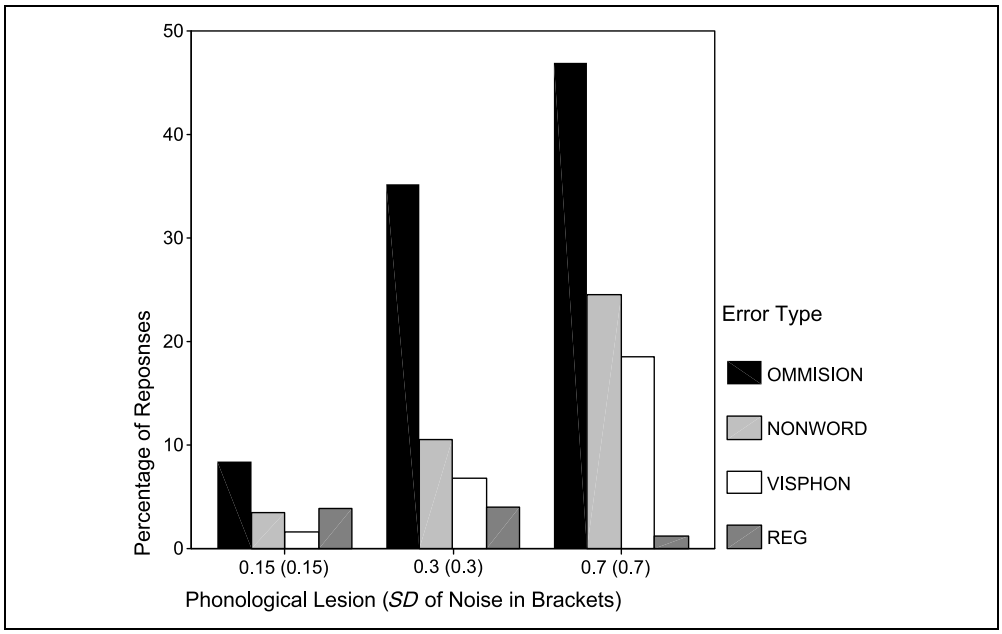

Figure 11. Showing the proportion of different error responses following phonological damage and retraining.

activation patterns in recovered patients. The obvious conclusion is that the synaptic reorganization that supports changes in brain activation patterns and response curves also underpins behavioral changes that occur in the period of spontaneous recovery, and contributes to the formation patterns of dissociation typical of chronic stage patients.

The second key factor in this simulation is the fact that the phonological damage was generalized in nature, affecting both the ability of the network to map from orthography to phonology and the integrity of its phonological representations. This was achieved by combining damage to the connections in the O → P pathway with noise added to the output of the phonological hidden units. Without the addition of noise, it is probable that the network would have been able to recover by finding solutions that relied more on the regularities in the training set, resulting in reduced lexicality effects and an increased influence of consistency. The idea that phonological dyslexia arises from generalized phonological damage is consistent with the primary systems hypothesis (Patterson & Lambon Ralph, 1999), which assumes that reading is subserved by the more general preexisting language systems and that the acquired dyslexias arise from generalized damage to one of these systems. Indeed, the current model could be regarded as a first step toward an implementation of the primary systems hypothesis. Of course, a full implementation would require a model that was able to perform additional linguistic tasks such as speech, comprehension, and repetition.

It is important to consider how the results of this simulation mesh with the results reported by Welbourne and Lambon Ralph (2005a). In that simulation, damage to an isolated phonological network resulted in a surface dyslexic performance; here, on the other hand, surface dyslexia arises from damage to the semantic portion of the network, whereas damage to the phonological portion produced the symptoms of phonological dyslexia. At first glance, this seems somewhat inconsistent; how is it that surface dyslexia can arise from two different damage loci? In reality, there is no inconsistency; in both cases, the endpoint is the same. Surface dyslexia occurs where the phonological system has insufficient computational resources to successfully process all of the words in its corpus and has no available support from semantics. Welbourne and Lambon Ralph (2005a) achieved this situation by damaging a phonological system that was initially overcompetent in that it could read without any support from semantics. In the current simulation, the same situation was achieved, more realistically, by removing semantics from a network where reading was supported by a division of labor between phonology and semantics (Plaut et al., 1996). Only in this latter situation, where there is the potential for a division of labor, can damage to the phonological system result in phonological dyslexia.

One slightly unexpected aspect of these results is the presence of a small consistency effect following mild or moderate phonological damage. This is not traditionally associated with phonological dyslexia. However, although it is not often reported, phonological dyslexics do often exhibit consistency effects. A reanalysis of data from Berndt et al. (1996)[5] reveals that 9 out of 10 of the patients in the series showed more accurate reading of regular than of irregular words with the performance difference ranging from 2% to 20%. When data from all of the patients are submitted to statistical analysis, these differences are shown to be significant ($t = 2.32$, $df = 9$, $p = .023$, one-tailed). Data from the only other case series of phonological dyslexics (Crisp & Lambon Ralph, 2006) are even more emphatic; 10 out of 12 patients showed a superiority for regular words, varying from 5% to 33%, and the group as a whole showed a very significant consistency effect ($t = 4.41$, $df = 11$, $p < .001$, one-tailed). The mean size of the consistency effect for the two sets of patients (including those who did not exhibit a consistency effect) was 5% for the Berndt et al. set and 14% for the Crisp and Lambon Ralph set. This compares with a mean consistency effect of 7.2% for the network (averaged across all damage severities). In light of this, it seems reasonable to suggest that this simulation has captured a hitherto unremarked feature of phonological dyslexia.

This study represents a considerable step forward in that it is the first time that any single PDP reading model has been able to produce both the frequency/consistency interactions typical of surface dyslexia and the lexicality/imageability effects associated with phonological dyslexia. However, there still remain a number of important questions which lie beyond the scope of the present model. As we indicated in the Introduction, the vast majority of phonological dyslexics present with generalized phonological deficits that are not specific to reading. This would be very interesting to explore computationally and this study takes a step toward this goal in that generalized phonological damage is shown to produce patterns of reading typical of phonological dyslexics. However, a thorough exploration of this issue would require a more complex network that could reproduce speech, repetition, and comprehension behaviors. In addition, it has been suggested that phonological and deep dyslexia form a continuum (Friedman, 1996). This study has demonstrated a continuum of symptoms within phonological dyslexia but it is not able to explore the relationship between severe phonological dyslexia and deep dyslexia. It is limited in this regard because the implementation of semantics does not allow for the production of semantic errors that are the defining symptom of deep dyslexia. Future studies should concentrate on models that include both semantic and phonological representations and should be trained on a variety of language tasks rather than just reading.

APPENDIX A—HIGH AND LOW IMAGEABILITY WORD LISTS

High Imageability			*Low Imageability*		
Word	*Imageability*	*KF Freq*	*Word*	*Imageability*	*KF Freq*
BANK	560	83	BEAU	394	2
BED	635	127	DRAB	340	5
BLUE	569	143	FATE	343	33
BRAIN	572	45	FIND	370	399
CHAIN	559	50	FOURTH	384	74
CHART	531	22	GRADE	397	35
CHEST	556	53	KEPT	300	186
CLAY	575	100	LOON	348	2
COAL	581	32	LUCK	399	47
CURB	556	13	NEED	327	360
HOUSE	606	591	PART	340	500
LINT	513	4	PAT	386	35
MINE	522	59	PRIME	386	45
MUG	574	1	SCORN	364	4
PLANE	556	114	SLOE	276	2
RIDGE	543	18	SLOUGH	245	1
RING	601	47	SMART	396	21
SHRIMP	618	2	STRICT	383	11
SOUP	604	16	THWART	324	3
SPADE	578	10	TREAT	360	26
SPIRE	541	5	TURN	384	233
SQUINT	515	1	TYPE	395	200
TACK	546	4	WARN	359	11
YELL	501	9	WISH	399	110
Mean	**563**	**65**	**Mean**	**358**	**98**
SD	**34**	**120**	***SD***	**41**	**142**

Acknowledgments

This research was supported by grants from NIMH (No. P50-MH64445), BBSRC (S20390), the Gatsby Charitable Foundation, and the University of Manchester. We are grateful for the helpful comments made by the two reviewers of this article.

Reprint requests should be sent to Stephen R. Welbourne, School of Psychological Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, UK, or via e-mail: Stephen.Welbourne@manchester.ac.uk.

Notes

1. We are grateful to David Plaut for sharing his training patterns with us.
2. In the original PMSP simulation, a distinction was made between the required activation levels for onsets and offsets as opposed to vowels: Onsets and offsets had to have activation levels of greater than 0.5, whereas the most active vowel was taken as the output regardless of its activation level. This reflects the fact that all monosyllabic words must, by definition, include a vowel.
3. Due to the abstract nature of the semantic implementation, it was not possible to characterize errors as being semantically related to the target word.
4. Data taken from responses to the PALPA 31 list of words varying in imageability and frequency.
5. We are grateful to Rita Berndt for allowing us access to some of the data that formed the basis of the article.

REFERENCES

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

Berndt, R. S., Haendiges, A. N., Mitchum, C. C., & Wayland, S. C. (1996). An investigation of nonlexical reading impairments. *Cognitive Neuropsychology, 13,* 763–801.

Bisiacchi, P. S., Cipolotti, L., & Denes, G. (1989). Impairment in processing meaningless verbal material in several modalities: The relationship between short-term memory and phonological skills. *Quarterly Journal of Experimental Psychology, Series A, Human Experimental Psychology, 41,* 293–319.

Blank, S. C., Bird, H., Turkheimer, F., & Wise, R. J. S. (2003). Speech production after stroke: The role of the right pars opercularis. *Annals of Neurology, 54,* 310–320.

Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience, 21,* 149–186.

Caccappolo-van Vliet, E., Miozzo, M., & Stern, Y. (2004). Phonological dyslexia without phonological impairment? *Cognitive Neuropsychology, 21,* 820–839.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behaviour Research Methods, Instrumentation, & Computers, 36,* 384–387.

Crisp, J., & Lambon Ralph, M. A. (2006). Unlocking the nature of the phonological-deep dyslexia continuum: The keys to reading aloud are in phonology and semantics. *Journal of Cognitive Neuroscience, 18,* 348–362.

Derouesné, J., & Beauvois, M. F. (1985). The ''phonemic'' stage in the non-lexical reading process: Evidence from a case of phonological alexia. In K. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia* (pp. 399–457). Hove, UK: Erlbaum.

Farah, M. J., Stowe, R. M., & Levinson, K. L. (1996). Phonological dyslexia: Loss of a reading-specific component of the cognitive architecture? *Cognitive Neuropsychology, 13,* 849–868.

Friedman, R. (1995). Two types of phonological dyslexia. *Cortex, 31,* 397–403.

Friedman, R. B. (1996). Recovery from deep alexia to phonological alexia: Points on a continuum. *Brain & Language, 52,* 114–128.

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

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106,* 491–528.

Harm, M. W., & Seidenberg, M. S. (2001). Are there orthographic impairments in phonological dyslexia? *Cognitive Neuropsychology, 18,* 71–92.
Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks, 1,* 295–307.
Kuçera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.
Leff, A., Crinion, J., Scott, S., Turkheimer, F., Howard, D., & Wise, R. (2002). A physiological change in the homotopic cortex following left posterior temporal lobe infarction. *Annals of Neurology, 51,* 553–558.
Nelson, M. E., & Bower, J. M. (1990). Brain maps and parallel computers. *Trends in Neurosciences, 13,* 403–408.
Patterson, K., & Lambon Ralph, M. A. (1999). Selective disorders of reading? *Current Opinion in Neurobiology, 9,* 235–239.
Patterson, K., & Marcel, A. J. (1992). Phonological ALEXIA or PHONOLOGICAL Alexia? In J. Alegria, J. Holender, J. Junça de Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 259–274). Amsterdam: Elsevier Science.
Patterson, K., Seidenberg, M. S., & McClelland, J. L. (1989). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neuroscience* (pp. 131–181). London: Oxford University Press.
Patterson, K., Suzuki, T., Wydell, T., & Sasanuma, S. (1995). Progressive aphasia and surface alexia in Japanese. *Neurocase, 1,* 155–165.
Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain & Language, 52,* 25–82.
Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology, 19,* 603–639.
Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.
Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19,* 1–10.
Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning & Verbal Behaviour, 23,* 383–404.
Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 1140–1154.
Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26,* 608–631.
Weiller, C., Isensee, C., Rijntjes, M., Huber, W., Müller, S., Bier, D., et al. (1995). Recovery from Wernicke's aphasia: A positron emission tomographic study. *Annals of Neurology, 37,* 723–732.
Welbourne, S. R., & Lambon Ralph, M. A. (2005a). Exploring the impact of plasticity-related recovery after brain damage in a connectionist model of single-word reading. *Cognitive, Affective & Behavioral Neuroscience, 5,* 77–92.
Welbourne, S. R., & Lambon Ralph, M. A. (2005b). Using computational, parallel distributed processing networks to model rehabilitation in patients with acquired dyslexia: An initial investigation. *Aphasiology, 19,* 789–806.
Woollams, A., Lambon Ralph, M. A., Plaut, D. C., Hodges, J. R., & Patterson, K. (in press). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*.
Young, M. P., Scannell, J. W., & Burns, G. (1995). *The analysis of cortical connectivity.* Heidelberg: Springer-Verlag.