Structure and Function in the Lexical System: Insights from Distributed Models of Word Reading and Lexical Decision

David C. Plaut

Departments of Psychology and Computer Science, Carnegie Mellon University, and the Center for the Neural Basis of Cognition, Pittsburgh, PA, USA

The traditional view of the lexical system stipulates word-specific representations and separate pathways for regular and exception words. An alternative approach views lexical knowledge as developing from general learning principles applied to mappings among distributed representations of written and spoken words and their meanings. On this distributed account, distinctions among words, and between words and nonwords, are not reified in the structure of the system but reflect the sensitivity of learning to the relative systematicity in the various mappings. Two computational simulations address findings that have seemed problematic for the distributed approach. Both involve a consideration of the role of semantics in normal and impaired lexical processing. The first simulation accounts for patients with impaired comprehension but intact reading in terms of individual differences in the division of labour between the semantic and phonological pathways. The second simulation demonstrates that a distributed network can reliably distinguish words from nonwords based on a measure of familiarity defined over semantics. The results underscore the importance of relating function to structure in the lexical system within the context of an explicit computational framework.

INTRODUCTION

Word reading is one of the most extensively investigated areas within cognitive psychology. A tremendous amount of data has been gathered on how various properties of a written word (frequency of occurrence, age of acquisition, numbers of letters and phonemes, orthographic neighbourhood, spelling–sound regularity and consistency, imageability, meaningfulness, prior context, etc.) influence the speed and accuracy with which a word is understood and pronounced by normally and abnormally developing readers, skilled readers and brain-damaged patients with selective reading impairments. Developing a comprehensive theory of lexical processing that can account for this full range of findings is a considerable challenge, one that has yet to be met.

A natural way of accounting for the effects of some variable or distinction on performance is to build the distinction directly into the structure of a model. A good example is the status of a letter string as a word. Few researchers would argue against the relevance to understanding human language performance of the distinction between letter strings that are words and those that are not. The fact that, in most contexts, skilled readers are fast and accurate at discriminating written words from nonwords with relatively little practice indicates that there must be *some* difference in how the reading system responds to words and nonwords. The simplest way of incorporating this distinction is to introduce words as explicit structures in a model. Well-known examples of such structures are Morton's (1969) logogens and McClelland and Rumelhart's (1981) word units in the Interactive Activation model. In a model with word-specific representations, words can be distinguished from nonwords simply by whether one of them reaches threshold or becomes fully active. Moreover, it becomes straightforward to account for effects of lexical variables by manipulating the word representations to reflect the variable explicitly. For example, effects of word frequency can be incorporated directly by setting the resting level of the logogen or word unit to be proportional to frequency.

Distinctions among words also lead researchers to postulate additional structure in the lexical system. The classic example is the distinction between regular words and exception words. Many researchers assume that knowledge of spelling–sound correspondences is encoded in terms of an explicit set of rules. The rules need to be general—for example, based on grapheme–phoneme correspondence—to support effective pronunciation of unfamiliar word-like letter strings (e.g. MAVE, RINT). However, in a language like English, in which spelling–sound correspondences are only partially systematic, such rules will apply to only a subset of words—so-called *regular* words (e.g. GAVE, HINT). The remaining *exception* words (e.g.

HAVE, PINT) are mispronounced—*regularised*—by the rules. Given that skilled readers can pronounce such words correctly, they must be using some other mechanism with word-specific knowledge, like the logogens or word units. In this way, a given set of spelling–sound correspondences rules defines a sharp dichotomy between the items which obey the rules and the items which violate them; theories that reify this dichotomy as two separate processing mechanisms are thus termed "dual-route" theories (e.g. Besner & Smith, 1992; Coltheart, 1978, 1985; Coltheart, Curtis, Atkins, & Haller, 1993; Marshall & Newcombe, 1973; Meyer, Schvaneveldt, & Ruddy, 1974; Morton & Patterson, 1980; Paap & Noel, 1991). Note that most such theories further subdivide the lexical mechanism into semantic and non-semantic components to account for the conditions under which normal and impaired word reading is (or is not) influenced by word meaning. Thus, despite the name, the essence of a dual-route theory is *not* that it has exactly two pathways from print to sound; rather, it is the claim that the mechanism that pronounces nonwords is functionally distinct from, and operates according to different principles than, the mechanism that pronounces exception words.

   Instead of accounting for effects of lexical variables by manipulating the structure of a model directly, an alternative theoretical approach is to start with a set of general assumptions on the nature of representation, processing and learning in the cognitive system, and to show how, when instantiated to perform specific tasks, these assumptions give rise to the relevant behavioural effects. This is the perspective adopted by researchers employing connectionist/parallel distributed processing (PDP) models of lexical processing (e.g. Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990; Van Orden & Goldinger, 1994). For example, in the framework proposed by Seidenberg and McClelland (1989; see Fig. 1), orthographic, phonological and semantic information is represented in terms of distributed patterns of activity over separate groups of simple neuron-like processing units. Within each domain, similar words are represented by similar patterns of activity. Lexical tasks involve transformations between these representations; for example, oral reading requires the orthographic pattern for a word to generate the appropriate phonological pattern. Such transformations are accomplished via the cooperative and competitive interactions among units, including additional *hidden* units that mediate between the orthographic, phonological and semantic units. In processing an input, units interact until the network as a whole settles into a stable pattern of activity—termed an *attractor*—corresponding to its interpretation of the input. Unit interactions are governed by weighted connections between them, which collectively encode the system's knowledge about how the different types of information are related. Weights that give rise to the appropriate

transformations are learned on the basis of the system's exposure to written words, spoken words and their meanings.

As Fig. 1 makes clear, the distributed connectionist approach does not entail a complete lack of structure within the lexical system. But the distinctions that are relevant relate to the different types of information that must be coordinated: orthographic, phonological and semantic. Given that such information may be based on input from different modalities (at least for the surface forms), it is natural to assume that the corresponding representations—and hence the pathways between them—are neuroanatomically distinct. In fact, such divisions play an important role in accounting for data on the selective effects of brain damage on reading. However, irrespective of these distinctions among types of representations, there is a uniformity in the processing mechanisms by which they are derived and interact. In this way, the distributed connectionist approach is fundamentally at odds with the core tenet of dual-route theories.

Seidenberg and McClelland (1989) provided computational support for the claim that word-specific representations and separate rule-based versus lexical processing routes are unnecessary to account for skilled oral reading. They trained a connectionist network (corresponding to the bottom portion of Fig. 1, termed the *phonological pathway*) to map from the orthography of 2897 monosyllabic English words—both regular and exception—to their phonology. After training, the network pronounced 97.7% of the words correctly, including most exception words. The network also exhibited the standard empirical pattern of an interaction of frequency and consistency in naming latency (Andrews, 1982; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987; Waters & Seidenberg, 1985) if its
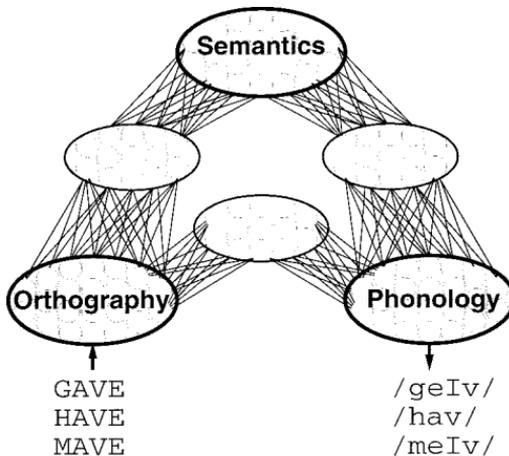

FIG. 1.   A connectionist framework for lexical processing, based on that of Seidenberg and McClelland (1989).

real-valued accuracy in generating correct responses was taken as a proxy for reaction time.

The theoretical impact of Seidenberg and McClelland's model is, however, undermined by certain inadequacies in its match to human performance, particularly in three respects. First, the model was much worse than skilled readers at pronouncing orthographically legal nonwords (Besner, Twilley, McCann, & Seergobin, 1990). Secondly, it was unable to perform lexical decision accurately under many conditions (Besner et al., 1990; Fera & Besner, 1992). Thirdly, it failed to exhibit central aspects of fluent surface dyslexia when damaged (Patterson, Seidenberg, & McClelland, 1989); specifically, the frequency $\times$ consistency interaction in naming accuracy, the high regularisation rates and (again) the near-perfect nonword reading (see Patterson, Coltheart, & Marshall, 1985).

More recently, Plaut et al. (1996) demonstrated that network implementations of the phonological pathway can learn to pronounce regular and exception words, and yet also pronounce nonwords as well as skilled readers, if they use more appropriately structured orthographic and phonological representations (based on graphemes and phonemes and embodying phonotactic and graphotactic constraints). Plaut and colleagues also provided a more adequate account of surface dyslexia based on the notion that semantics plays an important role in skilled reading. Specifically, the support for word pronunciations from the semantic pathway relieves the phonological pathway from having to master low-frequency exception words by itself. The combination of the two pathways is fully competent in skilled readers but, following damage to the semantic pathway, the limited competence of the isolated phonological pathway manifests as surface dyslexia.

In the current work additional simulations are presented that further develop the role of semantics to address two remaining problematic issues for the distributed connectionist approach to lexical processing. The first concerns a finding that would seem to challenge the above account of surface dyslexia; namely, that some brain-damaged patients have substantial semantic impairments but nonetheless can read low-frequency exception words accurately (W.L.P., Schwartz, Marin, & Saffran, 1979; M.B., Raymer & Berndt, 1994; D.R.N., Cipolotti & Warrington, 1995; D.C., Lambon Ralph, Ellis, & Franklin, 1995). The first set of simulations demonstrates that such patients are natural consequences of expected individual differences in the division of labour between the semantic and phonological pathways under parametric variations, even when the majority of individuals with semantic damage would be expected to exhibit some degree of surface dyslexia (Graham, Hodges, & Patterson, 1994; Patterson & Hodges, 1992). The second issue is the ability of distributed networks to perform lexical decision accurately. A second simulation, involving a full (although

feedforward) implementation of the framework in Fig. 1, demonstrates that a network that maps orthography to semantics directly and via phonology can reliably distinguish words from nonwords based on a measure of familiarity defined over semantics.

As background for the current work, the next section presents a brief overview of the approach taken by Plaut et al. (1996) in accounting for the effects of frequency and consistency in skilled word and nonword reading and in fluent surface dyslexia. The two simulation studies are then presented, followed by a general discussion.

AN ACCOUNT OF FREQUENCY AND CONSISTENCY EFFECTS IN NORMAL AND IMPAIRED WORD READING

Seidenberg and McClelland (1989) based their orthographic and phonological representations on context-sensitive triples of letters and phonemic features (Wickelgren, 1969). A central insight offered by Plaut et al. (1996) was that such representations suffer from what was termed the *dispersion problem*, and that it was this problem, rather than any more general failing of connectionist networks, that lead to the poor nonword reading performance of Seidenberg and McClelland's model. Specifically, the triples-based representations disperse the regularities between spelling and sound, thereby hindering generalisation. For example, in the orthographic representation of the word GAVE, the contribution made by the letter A depends entirely on the presence of the G and V; there is no similarity to the contribution made by the A in SAVE or GATE, and hence no basis for the training on the pronunciation of A in GAVE to support the same pronunciation in SAVE or GATE—or in a nonword like MAVE.

Plaut and colleagues hypothesised that the human reading system generalises effectively because it uses representations that condense spelling–sound regularities. To support this hypothesis, they designed orthographic and phonological representations in which, within each consonant and vowel cluster, letters and phonemes almost always activated the same units irrespective of context. The representations were not intended to be fully general, but rather to minimise the degree to which spelling–sound correspondences were dispersed across different units for different words. Plaut and colleagues found that, when networks were trained on essentially the same corpus as Seidenberg and McClelland's model, but using the new representations, they were able to pronounce all the words correctly and yet still generalise effectively to nonwords. Moreover, the networks exhibited the empirical frequency $\times$ consistency interaction pattern when trained on actual (as opposed to logarithmically

compressed) word frequencies, even when naming latencies were modelled directly by the settling time of a recurrent, attractor network (see Fig. 2a).

Importantly, Plaut et al. (1996) went beyond providing only empirical demonstrations that networks could reproduce accuracy and latency data on word and nonword reading, to offer a mathematical analysis of the critical
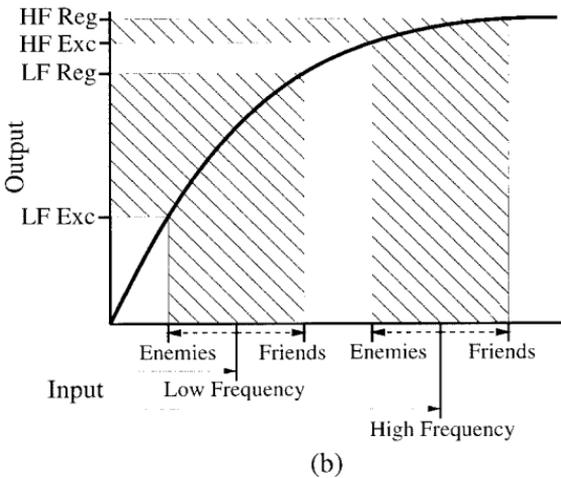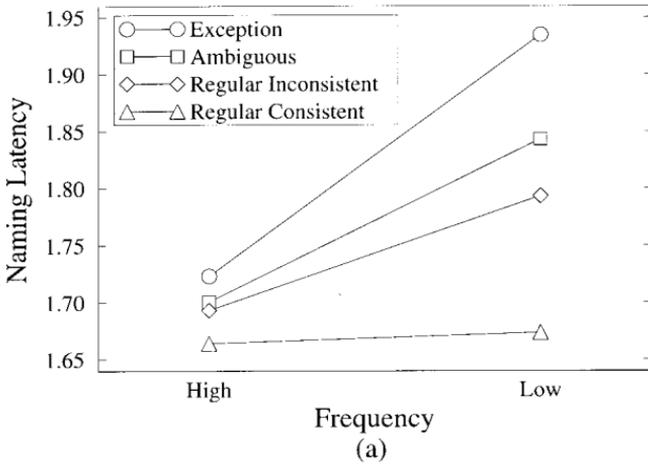

FIG. 2. (a) The frequency × consistency interaction exhibited in the settling time of an attractor network implementation of the phonological pathway in pronouncing words of varying frequency and spelling–sound consistency (Plaut et al., 1996, simulation 3). (b) Its explanation in terms of additive contributions of frequency and consistency subject to an asymptotic activation function—$\sigma(\cdot)$ in equation (1), only the top half of which is shown. From Plaut et al. (1996).

factors that govern *why* the networks (and, by hypothesis, subjects) behave as they do. This analysis was based on a network that, while simpler than the actual simulations—it had no hidden units and employed Hebbian rather than error-correcting learning—retained many of the essential characteristics of the more general framework. For this simplified network, Plaut and colleagues derived an expression for how the response of the network to any input (test) pattern depended on its experience with every pattern on which the network was trained, as a function of the pattern's frequency of training, its similarity with the test pattern, and the consistency of its output with that of the test pattern. Specifically, the response $s_j^{[t]}$ of any output unit j to a given test pattern t is given by:

$$s_j^{[t]} = \sigma \left(F^{[t]} + \sum_f F^{[f]} \mathcal{O}^{[ft]} - \sum_e F^{[e]} \mathcal{O}^{[et]} \right) \tag{1}$$

in which the standard smooth, non-linear sigmoidal input–output function for each unit, $\sigma(\cdot)$, is applied to the sum of three terms: (1) the cumulative frequency of training on the pattern t itself, $F^{[t]}$; (2) the sum of the frequencies $F^{[f]}$ of the "friends" of pattern t (patterns trained to produce the same response for unit j), each weighted by its similarity (overlap) with t, $\mathcal{O}^{[ft]}$; and (3) minus the sum of the frequencies $F^{[e]}$ of the "enemies" of pattern t (patterns trained to produce the opposite response), each weighted by its similarity to t, $\mathcal{O}^{[et]}$.

   Many of the basic phenomena in word reading can be seen as natural consequences of adherence to this "frequency–consistency" equation. Factors that increase the summed input to units (e.g. word frequency, spelling–sound consistency) improve performance as measured by naming accuracy and/or latency, but their contributions are subject to "diminishing returns" due to the asymptotic nature of the activation function, $\sigma(\cdot)$ (see Fig. 2b). As a result, performance on stimuli that are strong in one factor is relatively insensitive to variation in other factors. Thus, regular words show little effect of frequency, and high-frequency words show little effect of consistency, giving rise to the standard pattern of interaction between frequency and consistency in which the naming of low-frequency exception words is disproportionately slow or inaccurate.

   It is important to note, however, that equation (1) is only approximate for networks with hidden units and trained by error-correcting algorithms like back-propagation (Rumelhart, Hinton, & Williams, 1986a). These two aspects of the Plaut et al. (1996) simulations are critical in that they help to overcome interference from enemies (i.e. the negative terms in equation 1), thereby enabling the networks to achieve correct performance on exception words—that is, words with many enemies and few, if any, friends—as well as on regular words and nonwords.

Although Plaut and colleagues demonstrated that implementations of the phonological pathway on its own can learn to pronounce words and nonwords as well as skilled readers, a central aspect of their general theory is that skilled reading more typically requires the combined support of both the semantic and phonological pathways (for similar proposals, see Hillis & Caramazza, 1991; Van Orden & Goldinger, 1994), and that individuals may differ in the relative competence of each pathway (Seidenberg, 1992). Certainly, semantic involvement is necessary to pronounce homographs like WIND and READ correctly. Furthermore, a semantic variable—imageability— influences the strength of the frequency × consistency interaction in the naming latencies and errors of skilled readers (Strain, Patterson, & Seidenberg, 1995). Moreover, brain damage that impairs lexical semantics— typically to the left temporal lobe—can lead to surface dyslexia (see Patterson et al., 1985). In its purest "fluent" form (e.g. M.P., Behrmann & Bub, 1992; Bub, Cancelliere, & Kertesz, 1985; K.T., McCarthy & Warrington, 1986; H.T.R., Shallice, Warrington, & McCarthy, 1983), surface dyslexic patients read nonwords and regular words with normal accuracy and latency, but exhibit an interaction of frequency and consistency in word reading *accuracy* that mirrors the interaction shown by normal subjects in latency. That is, surface dyslexic patients are disproportionately poor at pronouncing low-frequency exception words, often giving a pronunciation consistent with more standard spelling–sound correspondences (e.g. SEW read as "sue", termed a *regularisation* error). In fact, there can be a close correlation for individual patients between the lack of comprehension of exception words and their likelihood of being regularised (Funnell, 1996; Graham et al., 1994; Hillis & Caramazza, 1991), and the surface dyslexic pattern may emerge gradually as lexical semantic knowledge deteriorates in patients with some types of progressive dementia, such as semantic dementia (Graham et al., 1994; Patterson & Hodges, 1992; Schwartz et al., 1979) or dementia of the Alzheimer's type (Patterson, Graham, & Hodges, 1994b).

The distributed connectionist framework depicted in Fig. 1 provides a natural formulation of how contributions from both the semantic and phonological pathways might be integrated in oral reading. At a more abstract level, given that phonological units simply sum their inputs from the two pathways, the influence of the semantic pathway can be incorporated into equation (1) by adding an additional term, $S^{[t]}$, to the summed input of unit j. [In fact, if it is assumed that this term scales with imageability, this can account for the three-way interaction of frequency, consistency and imageability in skilled reading found by Strain et al. (1995).] When formulated explicitly in connectionist terms, however, this integration has important implications for the nature of *learning* in the two pathways. During training, to the extent that the contribution of the semantic pathway

reduces the overall error, the phonological pathway will experience less pressure to learn. On its own, it may master only those items it finds easiest to learn: words high in frequency and/or consistency (i.e. those items with large positive terms in equation 1); low-frequency exception words may never be learned completely. As a result, brain damage that reduced or eliminated the semantic pathway would reveal the latent limitation of the phonological pathway, giving rise to surface dyslexia.

In further simulations, Plaut and colleagues explored the possibility that the surface dyslexic reading pattern might reflect the natural limitations of an intact but isolated phonological pathway that had learned to rely partially on semantic support. Given that a full implementation of the semantic pathway was beyond the scope of their work, they approximated the contribution that such a pathway would make to the oral reading of each word during training by providing the output (phoneme) units of the implemented phonological pathway with external input that pushed them towards their correct values for each word during training. Semantic damage, then, was modelled by weakening or removing this external input. Plaut and colleagues found that, indeed, a phonological pathway trained in the context of support from semantics exhibited the central phenomena of surface dyslexia when semantics was impaired. Figure 3 shows how different degrees of semantic impairment provide a reasonably good quantitative fit to the performance levels of individual surface dyslexic patients.

In summary, Plaut et al. (1996) provided connectionist simulations and mathematical analyses supporting a view of lexical processing in which the distinctions between words and nonwords, and between regular and exception words, are not reflected in the structure of the system, but rather in functional aspects of its behaviour as it brings all its knowledge to bear in processing an input. An important insight that emerges from the approach is that semantic and phonological processing are intimately related, over the course of reading acquisition, in normal skilled performance, and in the effects of brain damage. The following two simulation experiments aim to elaborate and clarify the role of semantics in oral reading and other lexical tasks. Although it might be more natural to consider normal performance (in lexical decision) before impaired performance (in surface dyslexia), we begin with the latter issue as the relevant simulations are more closely related to existing work.

SIMULATION 1: INDIVIDUAL DIFFERENCES IN DIVISION OF LABOUR

Plaut and co-workers' (1996) account of surface dyslexia involves a causal relationship between an impairment in semantic input to phonology and the occurrence of (regularisation) errors in reading low-frequency exception

Strength of Semantics

1.6 1.4 1.2 1.0 0.8 0.6 0.4 0.2 0.0

Percent Correct

100
90
80
70
60
50
40
30
20
10
0

● HF Reg
○ LF Reg
▲ HF Exc
△ LF Exc
✳ Regularizations
☐ Nonwords

(a)

HF Reg LF Reg HF Exc LF Exc Reg's Nonwords

Percent Correct

100
90
80
70
60
50
40
30
20
10
0

● Patient MP
▲ Patient KT
○ Sem = 0.9
△ Sem = 0.0

(b)
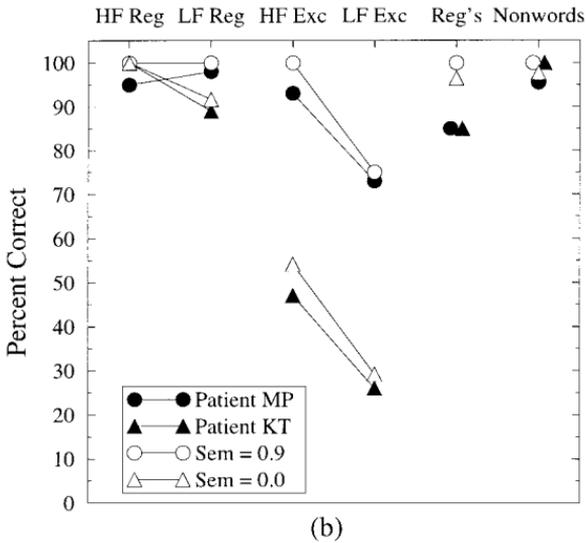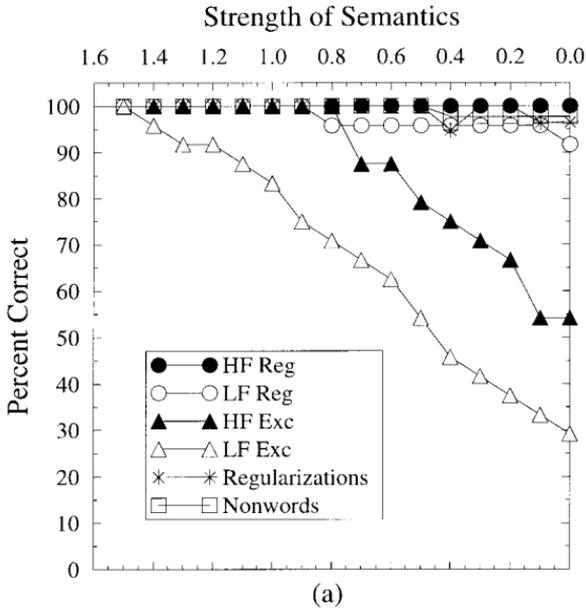
FIG. 3. (a) The effect of gradual elimination of semantics on the correct performance of Plaut and co-workers' (1996) network after 2000 epochs of training with semantics, for Taraban and McClelland's (1987) high-frequency (HF) and low-frequency (LF) regular consistent words (Reg) and exception words (Exc), and for Glushko's (1979) nonwords, and approximate percentage of errors on the exception words that are regularisations. (b) Performance of two surface dyslexic patients (M.P., Behrmann & Bub, 1992; Bub et al., 1985; K.T., McCarthy & Warrington, 1986) and the network with different amounts of semantic damage. From Plaut et al. (1996).

words. Note that such an impairment might arise from damage to semantics itself or from damage to the mapping from semantics to phonology in which semantics is spared. Thus, it is perfectly consistent with this account that some surface dyslexic patients have intact comprehension (see Graham, Patterson, & Hodges, 1995; Watt, Jokel, & Behrmann, 1997). Nonetheless, the semantics-to-phonology impairment should manifest in non-reading tasks, such as picture naming. Consistent with this expectation, all known surface dyslexic patients are anomic. In fact, when the anomia is restricted to abstract concepts, as in patient D.R.B. (Franklin, Howard, & Patterson, 1995), reading errors correspondingly occur only on exception words with abstract meanings.

Although Plaut and co-workers' account does not require that surface dyslexic patients have impaired comprehension, it does seem to imply that patients who do have a semantic impairment should exhibit some degree of surface dyslexia. Thus, the account is challenged by the finding that there are patients with semantic impairments who nonetheless read low-frequency exception words with normal or near-normal accuracy. The two clearest cases of this are D.R.N. (Cipolotti & Warrington, 1995) and D.C. (Lambon Ralph et al., 1995) (see also Coslett, 1991; Cummings, Houlihan, & Hill, 1986; Friedman, Ferguson, Robinson, & Sunderland, 1992; Raymer & Berndt, 1994; Schwartz et al., 1979).

Cipolotti and Warrington (1995) presented data on patient D.R.N., a 67-year-old biological scientist with focal atrophy of the temporal lobes (more pronounced on the left), most likely due to Pick's disease. Behaviourally he exhibited what is known as semantic dementia or progressive fluent aphasia (Hodges, Patterson, Oxbury, & Funnell, 1992; Snowden, Goulding, & Neary, 1989). Semantic dementia is a relatively pure impairment in semantic memory, as reflected by severe anomia and word comprehension difficulties, with relatively few other cognitive impairments. Cipolotti and Warrington compared D.R.N.'s ability to give definitions to words and to read them aloud, for sets of words that varied in frequency and spelling–sound consistency. We focus here on his performance on low-frequency exception words, as these items provide the most sensitive measure of surface dyslexia. In keeping with his semantic dementia, D.R.N. showed a substantial impairment in generating definitions for low-frequency exception words, producing 39% correct (11/28) and 14% correct (3/21) for two sets of words, using a very lenient scoring criterion. By contrast, his oral reading of low-frequency exception words was essentially intact; he produced 95% correct (46/48 and 20/21) on each of two lists.

Lambon Ralph et al. (1995) described the performance of patient D.C., who had attended school only until the age of 14. At the time of testing, D.C. was 85 years old and exhibited the clinical symptoms of dementia of the Alzheimer's type (see Schwartz, 1990), including very poor episodic memory

with fluent, well-structured speech and occasional word-finding difficulties. Like D.R.N., D.C. was impaired at defining low-frequency exception words but was within the normal range at reading them aloud. In generating definitions for the low-frequency exception words on Patterson and Hodges' (1992) "surface" list, D.C. was 31% correct (13/42) by a lax criterion and 57% correct (24/42) by a very lax criterion. By contrast, she was 95% correct (40/42) in reading the words aloud. Similar dissociations between generating definitions and pronunciations were observed for the object labels from the PALPA (Kay, Lesser, & Coltheart, 1992) and for word lists from Shallice et al. (1983) and from Strain et al. (1995).

One possible explanation for the lack of surface dyslexia in patients like D.R.N. and D.C. is that the task of generating definitions is far more sensitive to semantic or semantics-to-phonology damage than is the task of reading low-frequency exception words, and that surface dyslexia emerges only with very severe semantic damage. In fact, inspection of the results from progressive deterioration of semantics in Plaut and co-workers' (1996) simulation (Fig. 3a) indicates that low-frequency exception word reading is affected only when semantic strength is reduced to a level of 1.5 (from a maximum of 5.0). Consistent with this idea, patient W.L.P. (Schwartz et al., 1979) became impaired at reading exception words only when her semantic deterioration became very severe (Schwartz, Saffran, & Marin, 1980). Moreover, many fluent surface dyslexia patients show evidence of comprehension impairments and/or anomia that seem more severe than those of D.R.N. and D.C. As examples, M.P. (Bub et al., 1985) was at chance at word–word and picture–word semantic categorisation, only very slightly above chance at semantic category sorting, and exhibited no semantic priming in word naming, and K.T. (McCarthy & Warrington, 1986) scored 0 on the WAIS Vocabulary test, could not perform the Peabody Picture Vocabulary test (Dunn, 1965), could name only 11% (14/130) of a subset of the Snodgrass and Vanderwart (1980) pictures, and was only slightly above chance at picture–word matching. By comparison, on object naming, D.R.N. was 25% correct (5/20) and D.C. was 58% correct (128/220). However, some patients are no more anomic than D.R.N. and D.C. but show severe surface dyslexia. For example, G.C. (Patterson, Graham, & Hodges, 1994a) was correct in naming 45% (118/260) of the Snodgrass and Vanderwart pictures but could pronounce correctly only 38% (16/42) of low-frequency exception words (cf. 95% for both D.R.N. and D.C.). Thus, the severity of semantic impairment alone does seem to provide a full account of the differences in reading performance among all of the relevant patients.

An additional hypothesis is that individual differences in the severity of surface dyslexia can arise, not only from differences in the degree of semantic damage, but also from premorbid differences in the division of labour and overall competence of the reading system. Recall that the

simulations of Plaut et al. (1996) demonstrate that the phonological pathway on its own can, if necessary, master the entire training corpus, including low-frequency exception words, but that support from the semantic pathway relieves it from having to do so. Thus, the ultimate competence of the phonological pathway is a function both of its own learning properties—its sensitivity to frequency and consistency as expressed by equation (1)—and of the strength of the semantic contribution to phonology. Both of these factors are matters of degree that would be expected to vary across individuals. More generally, Plaut and co-workers pointed out that the competence and division of labour of the reading system would be expected to be influenced by a wide variety of factors, including the nature of reading instruction, the sophistication of preliterate phonological representations, the overall amount of reading practice, relative experience in reading aloud versus silently, the computational resources devoted to each pathway, and the reader's more general skill levels in visual pattern recognition and in spoken word comprehension and production. Broadly speaking, the overall competence of the phonological and semantic pathways together would be expected to improve gradually with reading experience, but that factors that aided the development of one pathway would be expected to limit the competence of the other.

In Plaut and co-workers' (1996) simulation of surface dyslexia, this trading relation between the semantic and phonological pathways depends primarily on two specific parameters. The first is the asymptotic strength over training of the external correct input provided to phoneme units, termed *semantic strength*, which corresponds directly to the competence of the putative semantic pathway. The second parameter is the magnitude of the pressure to keep weights in the phonological pathway small, termed *weight decay*. Weight decay is standardly used in connectionist modelling to improve generalisation (i.e. nonword reading) by preventing the network from developing arbitrarily large weights in learning the training data. With weight decay, the magnitude of each weight becomes proportional to its importance in reducing training error (Hinton, 1989). Weights can still grow large, however, if aspects of the task demand it. In fact, this property is critical in the current context if the network is to master exception words. Because the pronunciations of such words violate standard spelling–sound correspondences, they typically require the involvement of larger weights to override the standard mapping. Thus, the higher the amount of weight decay, the more difficult it is for the network to learn to pronounce exception words, particularly those of low frequency. In this way, the magnitude of weight decay controls the base level of competence that can be achieved by the phonological pathway in isolation. There are a number of possible neurophysiological analogues of weight decay that would be expected to vary across individuals. For instance, if a connection weight corresponds to

the number and strength of synapses between neurons, then the magnitude of weight decay can be thought of as the level of metabolic activity that can be supported for maintaining synapses. Low values of weight decay correspond to high levels of metabolic activity and thus a high level of competence in the network.

Under this view of individual differences as parametric variations of Plaut and co-workers' (1996) simulation, patients such as D.R.N. and D.C. may correspond to parameter combinations that lead to a high degree of competence in the phonological pathway. The goal of the current simulation was to determine the effects of parametric variations of semantic strength and weight decay on the severity of surface dyslexia exhibited by network implementations of the phonological pathway following semantic impairment.

Method

The method used in the simulation follows that used by Plaut et al. (1996, simulation 4), and readers are referred to that paper for additional details. A feedforward network was trained to map from orthographic representations of 2998 monosyllabic words to their phonological representations. The orthographic representations consisted of 108 grapheme units that code single- and multi-letter graphemes for the onset, vowel and coda of the written form of each word. The phonological representations were structured in an analogous fashion over 61 phoneme units. Each of the grapheme units was connected to each of 100 hidden units, which, in turn, was connected to each phoneme unit. Each of the hidden and phoneme units also had a *bias* weight that determined its "resting" state in the absence of other input. This bias is equivalent to the weight on a connection from a unit whose state is always equal to 1.0, and was learned in the same way as other weights.

During training, a word was presented to the network by clamping the states of units representing graphemes contained in the word to 1 and all others to 0. The hidden and phoneme units then computed their states based on this input. The state s_j of each unit j varied between 0 and 1 as a smooth, real-valued function of the summed weighted input from other units:

$$s_j = \frac{1}{1 + \exp(-\sum_i s_i w_{ij} + b_j)} \tag{2}$$

where w_{ij} is the weight on the connection from unit i to unit j; b_j is the bias weight of j; and $\exp(\cdot)$ is the exponential function. This is the standard sigmoid function, $\sigma(\cdot)$, referred to in the frequency–consistency equation (equation 1).

The summed input to phoneme units was augmented with an additional term, S, corresponding to the frequency-weighted contribution of the semantic pathway to phonology:

$$S = g \, \frac{\log (f + 2) \, t}{\log (f + 2) \, t + 2000} \qquad (3)$$

where f is the Kučera and Francis (1967) frequency of the word, and t indexes the training epoch (i.e. number of presentations of the training corpus). The parameter g, termed *semantic strength*, determined the asymptotic level of input from semantics and was varied from 1.0 to 5.0 in the current simulation. Figure 4 shows the magnitude of this input over time for each value of g using $f = 5.33$, the average Kučera and Francis (1967) frequency of the Patterson and Hodges (1992) low-frequency words. The sign of this external input was set to drive phoneme units towards their correct states.

The error, E, of the network was measured by the sum of two terms: (1) the *cross-entropy* (see Hinton, 1989) between the generated phoneme states s_j and the correct or target phoneme states t_j for the word; and (2) a proportion of the sum of the squares of the weights:

$$E = - \sum_j t_j \log_2 (s_j) + (1 - t_j) \log_2 (1 - s_j) \; + \; \lambda \sum_{i<j} w_{ij}^2 \qquad (4)$$

where λ is the *weight decay* parameter and was varied between 0.00005 and
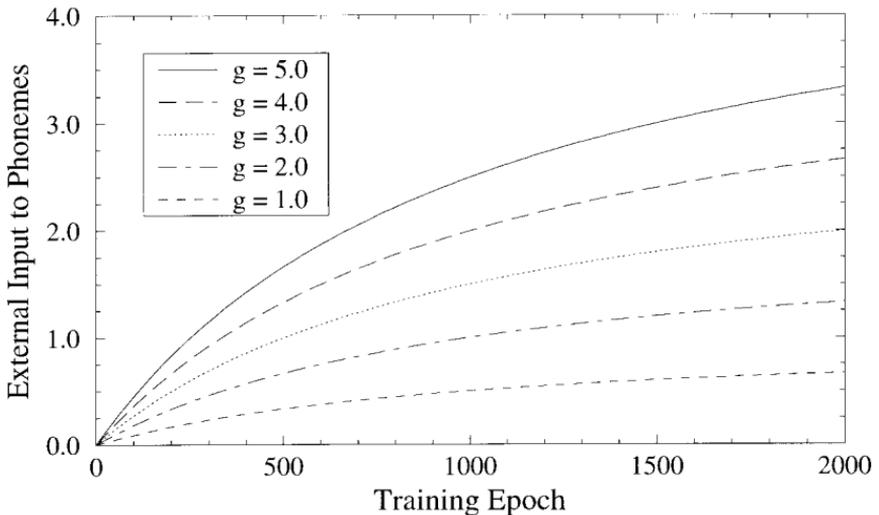
FIG. 4.   The magnitude of the additional external input supplied to phoneme units by the putative semantic pathway as a function of training epoch, for different values of semantic strength (g in equation 3) for a frequency $f = 5.33$ corresponding to the average Kučera and Francis (1967) frequency of the Patterson and Hodges (1992) low-frequency words.

0.001 in the current simulation. Back-propagation (Rumelhart, Hinton, & Williams, 1986b) was used to calculate the derivative of this error with respect to each weight in the network. This derivative was scaled by a factor reflecting the frequency of occurrence of the word, proportional to the square-root of its Kučera and Francis (1967) frequency.[1] Each of the 2998 words was presented in turn and their scaled derivatives were accumulated. At the end of each epoch, the weights were changed in proportion to the sum of the current accumulated derivatives and the past weight step (scaled by a *momentum* parameter of 0.9). The weight change was scaled both by a global learning rate of 0.001 and by a connection-specific learning rate that was initialised to 1.0 and adapted using the *delta-bar-delta* procedure (Jacobs, 1988) with an additive increment of 0.1 and a multiplicative decrement of 0.9.

For each combination of semantic strength and weight decay, the network was initialised with the same set of random weights (sampled uniformly between ± 0.1) and trained for 2000 epochs. At this point, it was subjected to a complete semantic "lesion" by removing the external input to the phoneme units. The network was then tested for its pronunciations of the high- and low-frequency regular and exception words from Patterson and Hodges' (1992) "surface" list, and 86 nonwords from Glushko (1979; see Plaut et al., 1996, for details on scoring).

Results and Discussion

Figure 5 displays the number of errors to the Patterson and Hodges words made by the network after the elimination of semantics, as a function of the semantic strength (g in equation 3) and weight decay (λ in equation 4) used during training. These error scores were entered into an ANOVA over items with frequency and consistency as between-item factors and semantic strength and weight decay as within-item factors.

As is apparent from Fig. 5, the network exhibited a strong frequency × consistency interaction in accuracy [$F(1,164) = 17.6$, $P < 0.001$]. Moreover, 86.6% (174/201) of the errors to exception words were regularisations, and 98.5% (1186/1204)[2] of presentations of the Glushko nonwords were pronounced correctly. Thus, overall, the network exhibited the hallmark characteristics of surface dyslexia.

However, as is also apparent from Fig. 5, the effects of frequency and consistency depended strongly on the levels of semantic strength and weight

[1]Note that this manipulation is equivalent, in the limit of small weight changes, to the more natural procedure of sampling words for training as a function of their frequency.

[2]This level of correct performance was calculated over only the 14 pairings of semantic strength and weight decay giving rise to non-perfect performance on the low-frequency exception words (see Fig. 5).

High-Frequency Regular Words (n=42)
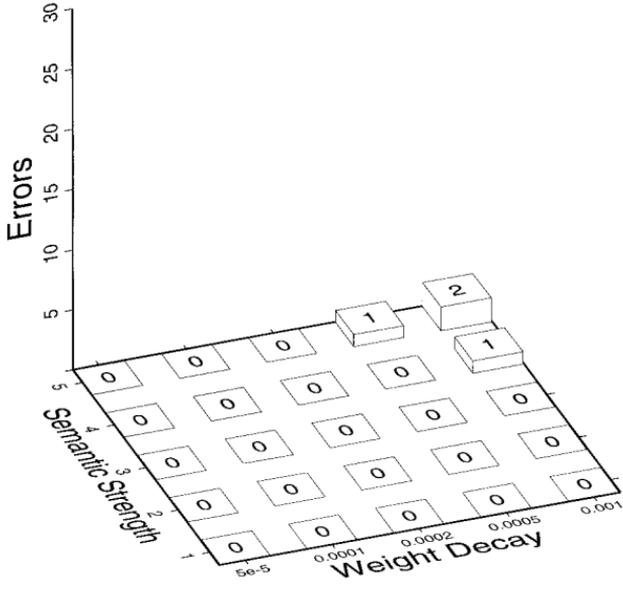
(a)

Low-Frequency Regular Words (n=42)
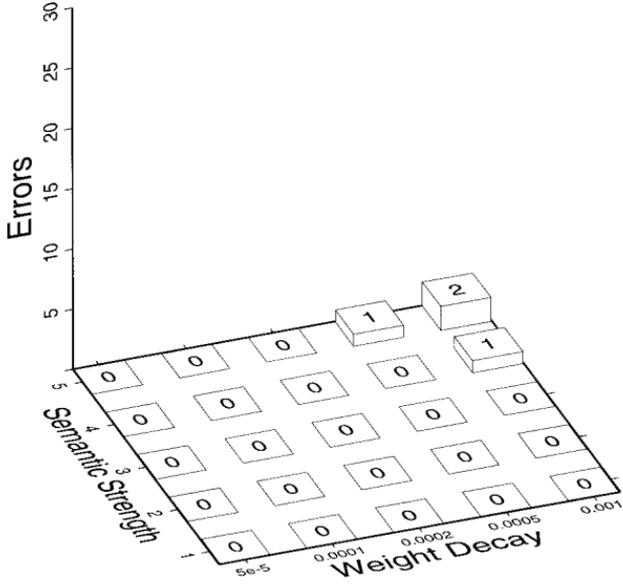
(b)

High-Frequency Exception Words (n=42)
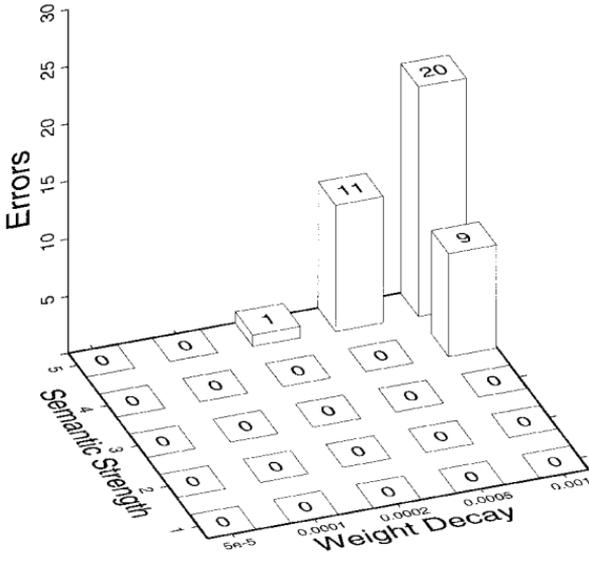
(c)

Low-Frequency Exception Words (n=42)
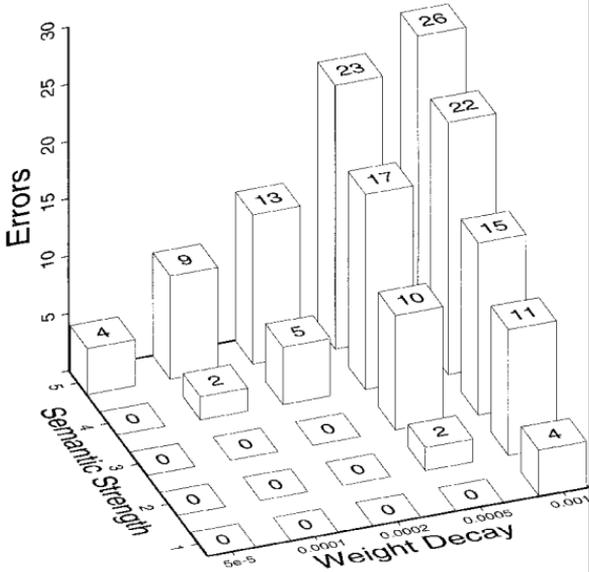
(d)

FIG. 5. Numbers of errors produced by the network to Patterson and Hodges' (1992) (a) high-frequency regular words, (b) low-frequency regular words, (c) high-frequency exception words and (d) low-frequency exception words, when semantics is removed completely after 2000 epochs of training, as a function of the semantic strength (g in equation 3) and weight decay (λ in equation 4) used during training.

decay. In fact, there was a four-way interaction among these factors [$F(16,2624) = 4.25$, $P < 0.001$]. In general, accuracy on low-frequency exception words improves if either semantic strength or weight decay is reduced, reaching perfect or near-perfect levels if either of these factors is extreme or if both are moderate (e.g. $g = 3.0$ and $\lambda = 0.0002$).

Thus, patients like D.R.N. and D.C., who have semantic impairments but no surface dyslexia, can be understood as falling at one end of a continuum of performance levels that would be expected under different divisions of labour within the reading system. Note that this perspective provides an account of how these two patients can exhibit qualitatively similar performance even though their educational backgrounds—and hence, presumably, their reading experience—differ so dramatically. D.R.N., with a high degree of education, may have a highly developed phonological pathway, corresponding to low levels of weight decay (e.g. $\lambda = 0.00005$). Even with a strong semantic pathway (e.g. $g = 5.0$), complete semantic damage yields 90.5% correct (38/42) on low-frequency exception words, which is comparable to D.R.N.'s performance of 95% on such words. By contrast, D.C., with little formal education, may have developed semantic and phonological pathways that are both weak (e.g. $g = 1.0$ and $\lambda = 0.001$). This combination also yields 90.5% correct (38/42) on low-frequency exception words, which is comparable to the 95% correct performance of D.C.

It should be kept in mind that the current demonstration involves the simplifying assumptions that the patients' brain damage completely eliminates the contribution of the semantic pathway and completely spares the phonological pathway. In general, the network's performance improves with partial sparing of the semantic pathway (see Fig. 3a) and degrades with partial impairment of the phonological pathway (see Plaut et al., 1996, fig. 20 and table 9, pp. 94–95). Thus, the performance of individual patients may also reflect a balance of these two factors. In particular, partial phonological pathway damage would seem to be implicated in patients in the later stages of semantic dementia. Such patients become impaired at pronouncing regular words as well as exception words, although an advantage for regular words remains (see Patterson et al., 1996). Interestingly, apart from visual errors (which occur equally often to regular and exception words), errors on regular words are predominantly what Patterson terms LARC errors—or Legitimate Alternative Reading of Components. The classic LARC error is a regularisation of an exception word, but such errors also occur to regular words with inconsistent bodies. As examples (from Patterson & Hodges, 1992), P.B. named HOOT to rhyme with "foot", YEAST like "breast", HEAR like "bear"; F.M. pronounced BROWN like "blown", HEAT like "threat", COST like "post", etc. Thus, although errors to regular and exception words differ in quantity, they do not seem to differ in nature, suggesting that they have a

common cause (see Patterson et al., 1996, for further discussion). On the Plaut et al. (1996) account, LARC errors to both regular and exception words derive from the sensitivity of learning in the phonological pathway to graded degrees of frequency and consistency, as reflected in equation (1).

In summary, the patterns of performance of patients like D.R.N. and D.C. indicate that low-frequency exception word reading does not always depend on semantic support. Nonetheless, the current simulation demonstrates that such patients fall within the distribution of performance patterns expected from parametric variations within a system in which the division of labour and mutual support of the semantic and phonological pathways play a critical role. Thus, the role of semantics in skilled reading, and the effects of semantic impairment on dyslexic reading, remain of central importance in understanding the organisation and function of the reading system. At a more general level, the simulation results also underscore the importance of considering, not just the performance of individual patients, but also the full distribution of performance patterns that can arise following brain damage (see Plaut, 1995a, for additional results and discussion).

We now turn to issues in normal reading that have seemed problematic for the distributed connectionist account—specifically, how skilled readers distinguish words from nonwords, and how this process is influenced by the nature of the stimuli in the task. Just as with the performance of brain-damaged patients, a consideration of the contribution of semantics will prove crucial in understanding the performance of normal subjects.

SIMULATION 2: LEXICAL DECISION

As mentioned in the Introduction, the distinction between words and nonwords is of fundamental importance to the lexical system, and many researchers incorporate this distinction into their models by representing words as explicit structural entities such as logogens or localist word units. Such units provide a natural account of how skilled readers can accurately distinguish written words from nonwords in lexical decision (LD) tasks. Given that the current distributed connectionist approach to lexical processing does not contain word-specific representations, it becomes important to establish that distributed models can, in fact, perform lexical decision accurately and that, in doing so, they are influenced by properties of the words and nonwords in the same way as human readers.

Seidenberg and McClelland (1989) attempted to demonstrate that distributed representations can provide a sufficient basis for lexical decision. They assumed that subjects make word/nonword decisions based on some measure of the "familiarity" of the stimulus (Atkinson & Juola, 1973; Balota

& Chumbley, 1984). This familiarity measure could potentially be computed from a variety of types of information derived by the lexical system— orthographic, phonological and semantic. Subjects are assumed to adopt a specific strategy and decision criterion that allows fast responding with acceptable error rates depending on the composition of the word and nonword stimuli. Given that orthographic information is provided directly in the input, subjects rely on orthographic familiarity when this provides a sufficient basis for distinguishing words from nonwords—for example, if the nonwords are orthographically illegal consonant strings (e.g. PSLR). If the nonword stimuli include very word-like, orthographically legal nonwords (e.g. NUST), subjects may consider phonological familiarity—whether the stimulus sounds like a word. If the nonwords include so-called "pseudohomophones" (e.g. BRANE) that have the same pronunciation as a word, subjects may have to base their decision on whether the orthography generates a familiar meaning.

Although Seidenberg and McClelland's general framework includes semantics, their specific implemented network did not. Accordingly, they focused on demonstrating that, under some conditions, lexical decision can be performed by a distributed network using only orthographic and/or phonological information. In their simulation, the orthographic familiarity of a stimulus was measured by the discrepancy between the orthographic pattern presented to the network and the one regenerated by the network from its hidden representation. Seidenberg and McClelland showed that this *orthographic error score* tended to be smaller for trained stimuli (words) than for novel stimuli (nonwords), although there was some overlap in the distributions. Moreover, the degree of overlap depended on properties of the words and nonwords in a way that corresponded to the effects of these properties on LD accuracy and latency in empirical studies. For example, the overlap is increased and, hence, lexical decision is slower and less accurate when the words are of lower frequency (e.g. Gordon, 1983) or include so-called "strange" words with unusual spelling patterns (e.g. AISLE, GAUGE; Waters & Seidenberg, 1985), or when the nonwords include pseudo-homophones (e.g. Coltheart, Davelaar, Jonasson, & Besner, 1977; McCann, Besner, & Davelaar, 1988). Seidenberg and McClelland assumed that, as the overlap in orthographic error scores for words and nonwords increases, the network (and subjects) would have to turn to phonological and/or semantic information to perform lexical decision accurately. In particular, the presence of strange words causes subjects to rely on phonological information, giving rise to effects of spelling–sound consistency in lexical decision (Waters & Seidenberg, 1985).

However, Besner and colleagues (Besner et al., 1990; Fera & Besner, 1992) challenged the claim that Seidenberg and McClelland's model provides an adequate account of LD performance, even under conditions in

which orthographic information is deemed sufficient (i.e. in the absence of strange words). Besner et al. (1990) reported that, when a decision criterion over orthographic error scores is set to yield an error rate on words of 5.2% (as observed by Waters & Seidenberg, 1985, experiment 2), the false-positive rate on nonwords exceeded 28%. Although Waters and Seidenberg did not report nonword error rates, this value is certainly higher than would be expected of subjects, especially if given unlimited time to respond. Moreover, Fera and Besner (1992) demonstrated that, for human subjects, the degree of overlap in orthographic error scores for words and nonwords had no effect on LD accuracy and latency, nor on the magnitude of a pseudohomophone effect among nonwords. These negative findings concerning Seidenberg and McClelland's implemented model and its predictions call into question the more general claim that words can be distinguished from nonwords by a distributed system lacking word-specific representations.

The goal of the current simulation was to demonstrate that lexical decision can be performed accurately when based on a familiarity measure applied to semantics and that, moreover, such a measure exhibits the appropriate sensitivity to properties of the words and nonwords in the task. It is important to be clear at the outset that this demonstration should not be interpreted as implying that lexical decision always relies on semantics. The perspective taken is exactly that of Seidenberg and McClelland (1989)—subjects performing lexical decision may adopt a variety of strategies, including those that depend only on orthographic and/or phonological information, as a function of the composition of the stimuli. In fact, subjects are assumed to rely on these more surface representations as much as possible. It is acknowledged, however, that in some contexts, orthographic and phonological representations will not support sufficiently accurate LD performance on their own. The current work aimed to establish that semantic familiarity provides a sufficient basis for lexical decision within a distributed lexical system, and supports the hypothesis that subjects can perform lexical decision accurately by relying on semantics when necessary.

Method

A feedforward network was trained to map from the orthographic representations of the 2998 monosyllabic words in Plaut et al.'s (1996) corpus to their phonological representations and to newly created semantic representations. The architecture of the network, shown in Fig. 6, cor-responded to an instantiation of the full framework for lexical processing in Fig. 1, except that the depicted hidden units between orthography and
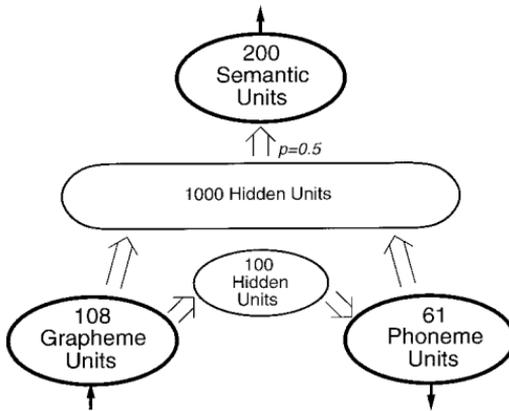
FIG. 6.   The network architecture used to model naming and lexical decision. Large arrows represent full connectivity between layers except where indicated. Small arrows indicate input units (incoming arrow) or output units (outgoing arrow).

semantics and between phonology and semantics were combined.[3] Specifically, 108 grapheme units were connected to 100 hidden units, which, in turn, were connected to 61 phoneme units. The grapheme and phoneme units were also connected to a second group of 1000 hidden units, which were then connected to 200 semantic units such that each hidden unit had a probability of 0.5 of being connected to each semantic unit. A much larger number of hidden units was needed to map to semantics than to map from orthography to phonology because there is no systematicity between the surface forms of words and their meanings, and connectionist networks find unsystematic mappings particularly difficult to learn (see Hinton, McClelland, & Rumelhart, 1986, for discussion).[4] Including bias weights, the network contained a total of 283,998 connections.

   The fact that the network has a feedforward architecture should not be interpreted as a theoretical claim about the structure of the human reading system. The underlying theory incorporates interactivity, and its use in forming attractors, as an important computational principle. Thus, the current feedforward network should be thought of as an approximation to a

---

   [3]This alteration was introduced only to reduce the time required to train the network—it is not intended as an important theoretical claim about the organisation of the semantic pathway. In fact, a network with two separate sets of 500 hidden units was developed subsequently and replicated the current results.

   [4]Note that the portion of the network that maps to semantics still has fewer hidden units ($n = 1000$) than words on which it is trained ($n = 2998$), so the network cannot solve the task by devoting a single hidden unit to each word (i.e. by developing word-specific representations). In fact, networks with sigmoidal units and trained with error-correcting procedures like back-propagation do not develop this "localist" solution even when given far more hidden units than training examples.

fully recurrent one. Training a recurrent version of the network was, however, infeasible due to limitations in the available computational resources.[5] Nonetheless, there are certain aspects of the feedforward approximation that are important to point out. In particular, processing is unidirectional from phonology to semantics, so that phonological representations (derived from orthography) influence semantics in performing lexical decision, but semantic representations cannot influence phonology in naming. As a result, unlike the first simulation, the naming performance of the current network does not reflect a contribution of semantics, and thus constitutes only a coarse approximation of human performance. Even so, Plaut et al. (1996) demonstrated that, with regard to normal performance, networks trained without semantics produce patterns of accuracy and latency results that are very similar to those of networks trained with semantics.

The semantic representations for words were designed to capture only the most abstract characteristics of word meanings—namely, that they cluster into categories and are arbitrarily related to orthographic and phonological representations. They were created in the following way. First, 120 random prototype patterns were generated over 200 semantic features such that each feature had a probability $P_a = 0.1$ of being active. Then each prototype was used to generate 25 exemplars that cluster around it by regenerating each semantic feature (using $P_a = 0.1$) with a probability $P_r = 0.05$, under the constraint that each exemplar had to differ from every other exemplar by at least three features. This procedure created 3000 semantic exemplars, each with an average number of active features of 20.12 (SD 1.712, range 14–26; the mean is slightly greater than 20 due to the constraint on the minimum difference between semantic patterns). Fifteen of these exemplars were chosen at random and discarded, and the remaining 2985 patterns were assigned randomly to the words in the training corpus (to avoid the difficulties involved in learning one-to-many mappings from orthography-to-semantics, the 13 pairs of homographs in the corpus, such as WIND and READ, were assigned only one meaning). The orthographic and phonological representations for the words were the same as used in Simulation 1 and by Plaut et al. (1996).

Although the semantic patterns used in the simulation do not reflect the relative similarities among the actual meanings of these words, their random assignment to words ensures that there is no systematic relationship between the written and spoken form of each word and its meaning. On the current approach, it is only this property that is critical for demonstrating that

---

[5] Training the feedforward network took just under 10 days of CPU time on a Silicon Graphics R4400 (150 MHz) processor. Based on prior experience, training a fully recurrent version would have taken approximately 20 times this amount of processing, or about 6 months.

semantics can support lexical decision. In fact, abstract semantic representations like those used in the current simulation have been used successfully to model empirical phenomena in a number of psycholinguistic domains, including word learning (Chauvin, 1988), inflectional morphology (Cottrell & Plunkett, 1991; Hoeffner, 1992), lexical ambiguity resolution (Joordens & Besner, 1994; Kawamoto, 1988, 1993; Kawamoto, Farrar, & Kello, 1994), semantic and associative priming (Masson, 1995; Plaut, 1995b) and rehabilitation of impaired reading via meaning (Plaut, 1996).

Given that the number of incoming connections to units varied widely in parts of the network, some sets of connections were initialised differently than others. Specifically, weights in the network were initialised to random values drawn uniformly between ± 1, except that hidden-to-phoneme weights were initialised between ± 0.2, hidden-to-semantics weights were initialised between ± 0.1, and the bias weights for phoneme and semantic units were pre-set to -2.94444 (the magnitude of input producing an initial state of 0.05). These conditions ensured effective learning at the very outset of training, but are otherwise irrelevant to the results reported below.

The network was trained with back-propagation using a learning rate of 0.0001, momentum of 0.9 (set to 0.0 for the first 10 epochs), adaptive connection-specific learning rates (Jacobs, 1988, increment of 0.1, decrement of 0.9), and the cross-entropy error function without weight decay (see equation 4). During training, each presentation of a word generated both a phonological pattern and a semantic pattern as output. The error computed over the phoneme units was back-propagated to the grapheme units and used to calculate derivatives for weights in the orthography-to-phonology pathway, exactly as in Simulation 1. The error computed over the semantic units was back-propagated both to the grapheme units and to the phoneme units and used to calculate derivatives for the orthography-to-semantics and phonology-to-semantics weights. Semantic error was not back-propagated through the phoneme units to influence the derivatives for the orthography-to-phonology weights, however, to prevent the network from trading off phonological accuracy for semantic accuracy. These error derivatives were scaled by a factor proportional to the logarithm of the word's Kučera and Francis (1967) frequency and accumulated for each word before being used to change the weights at the end of each epoch.

After 1300 epochs of training, the network was tested for its performance both at naming and at lexical decision, using the words in the training corpus and two lists of nonwords. The first list consists of 591 pronounceable nonwords created by Seidenberg et al. (1994) from each unique body in Seidenberg and McClelland's (1989) corpus, and is referred to below as the "body-matched" nonword list. It was used by Seidenberg and colleagues to compare the nonword reading performance of 24 human subjects with the

performance of Plaut and McClelland's (1993) network (a precursor to the simulations of Plaut et al., 1996) and the performance of the pronunciation rules in Coltheart and co-workers' (1993) Dual Route Cascaded model. This list was chosen because its size and diversity allows a thorough evaluation of overall LD accuracy. The second list of nonwords (Seidenberg, Petersen, MacDonald, & Plaut, 1996) contains 64 pseudohomophones and 64 nonpseudohomophone control nonwords, and is referred to below as the "PH/nonPH" nonword list. The two sets of items are very closely matched orthographically because they were constructed in groups of four by exchanging onsets and bodies (e.g. pseudohomophones JOAK and HOAP, nonpseudohomophones HOAK and JOAP). Thus, these nonwords provide a stringent test for the existence of pseudohomophone effects in the absence of orthographic confounds (see Seidenberg et al., 1996, for discussion).

Lexical decision was based on a measure of the familiarity of the semantic pattern generated by a word or nonword. A commonly used measure of familiarity in distributed networks is the negative of the "energy", $\Sigma_{i\,<\,j}, s_i s_j w_{ij}$ (Hopfield, 1982), and a number of researchers (e.g. Besner & Joordens, 1995; Borowsky & Masson, 1996; Joordens & Becker, 1997; Masson & Borowsky, 1995; Rueckl, 1995) have proposed recently that it may be possible to perform lexical decision in distributed models on the basis of differences in the energy for words versus nonwords. A serious drawback of this measure, however, is that it requires decision processes to have explicit access to the weights between units (analogous to the size and number of synapses between neurons), which is far less neurobiologically plausible than a procedure that need only access unit states.

An appropriate alternative measure, termed *stress*, is based only on the states of units. Specifically, the stress S_j of unit j is a measure of the information content (entropy) of its state s_j, corresponding to the degree to which it differs from rest:

$$S_j = s_j \log_2(s_j) + (1 - s_j) \log_2(1 - s_j) - \log_2(0.5) \tag{5}$$

The stress of a unit is 0 when its state is 0.5 and approaches 1 as its state approaches either 0 or 1. The target semantic patterns for words are binary, and thus have maximal stress. Because, over the course of training, the semantic patterns generated by words increasingly approximate their target patterns, the average stress of semantic units approaches 1 for words. By contrast, nonwords are novel stimuli that share graphemes with words that have conflicting semantic features. As a result, nonwords will typically fail to drive semantic units as strongly as words do, producing semantic patterns with much lower average stress. Accordingly, the average stress of semantic units, here termed simply *semantic stress*, should provide an adequate basis for performing lexical decision. We assume that LD responses are actually generated by a stochastic decision process (e.g. Ratcliff, 1978; Usher

& McClelland, 1995) in which a decision criterion is adopted such that stimuli with stress values farther from this criterion are responded to more quickly.

Results and Discussion

Naming.   After the 1300 epochs of training, the network pronounced all of the 2998 words in the training corpus correctly (where homographs were considered correct if they elicited either correct pronunciation). Moreover, the cross-entropy error the network produced when tested on Patterson and Hodges' (1992) high- and low-frequency regular and exception words replicated the standard empirical finding of a frequency $\times$ consistency interaction in naming latency [means: high-frequency regular = 0.0022, low-frequency regular = 0.0051, high-frequency exception = 0.0051, low-frequency exception = 0.0203; $F(1,164) = 20.56$, $P < 0.001$].

With regard to nonword reading, 91.4% (540/591) of body-matched nonwords were given a pronunciation that either matched the pronunciation given by at least one of Seidenberg and co-workers' (1994) 24 human subjects, or was consistent with the pronunciation of a word in the training corpus with the same body. For the PH/nonPH nonwords, the network's pronunciations of 96.9% (62/64) of the pseudohomophones and 95.3% (61/64) of the nonpseudohomophones matched that of some word in the training corpus with the same body.

Thus, overall, the network exhibited the appropriate pattern of skilled performance in naming words and nonwords.

Lexical Decision.   Semantic stress values were calculated using equation (5) for the body-matched nonwords and for sets of high-, medium- and low-frequency words. The words sets consisted of the 600 words in the training corpus with the highest Kučera and Francis (1967) frequency (mean 653.4, median 158), the 600 with median frequency (mean 8.807, median 9) and the 600 with the lowest frequency (mean 0.6298, median 1). Figure 7 displays the distributions of semantic stress for these words and nonwords.

As Fig. 7 shows, there is very little overlap between the semantic stress values for words and those for nonwords. In fact, if a decision criterion is adopted such that a stimulus is accepted as a word if it generates semantic stress in excess of 0.955, then the network produces error rates of 1.5% on both words (27/1800) and nonwords (9/591) and a d' value of 4.33. Considering only low-frequency words, a criterion of 0.95 yields 2.5% misses (15/600) and 4.4% false-alarms (26/591) and a d' of 3.67. For high-frequency words, a criterion of 0.965 yields 0.167% misses (1/600) and false-alarms (1/591) and a d' of 5.87. These very high levels of discriminability between words and nonwords should be interpreted as reflecting asymptotic
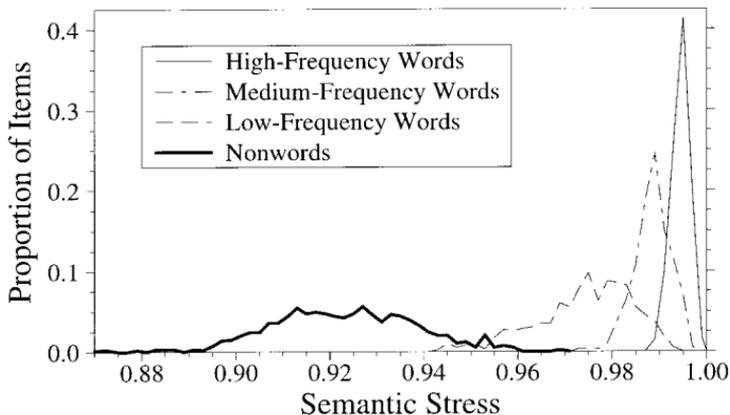
FIG. 7. Distributions of semantic stress values for the body-matched nonwords ($n = 591$; Seidenberg et al., 1994) and for high-, medium- and low-frequency words from the training corpus ($n = 600$ for each).

performance in the absence of time pressure. We assume that subjects can trade accuracy for reduced latency under conditions in which they are encouraged to respond as quickly as possible while keeping error rates acceptably low (e.g. by adjusting the response threshold in Ratcliff's, 1978, diffusion model).

Moreover, Fig. 7 illustrates that the distributions of stress values for words vary systematically as a function of their frequency. Specifically, the overlap with nonwords increases as word frequency decreases. This property provides an account of the *frequency blocking effect* in lexical decision. Gordon (1983; see also Glanzer & Ehrenreich, 1979) compared LD latency to high-, medium- and low-frequency words when presented in mixed blocks versus when blocked by frequency. The blocking manipulation did not influence the error rates for words in each frequency band, nor the latencies to low-frequency words. By contrast, LD latencies to medium- and high-frequency words were faster (by 19 and 40 msec, respectively) when presented in pure blocks than when mixed with low-frequency words. These findings make sense in the context of the distributions of semantic stress values shown in Fig. 7 if subjects adjust their decision criterion to optimise their performance within each block given the composition of the stimuli. A conservative criterion is required to achieve acceptable levels of accuracy when low-frequency words are present in the block (either mixed or pure). However, when only higher-frequency words are present, a more aggressive criterion can be adopted that, for the same error rate, produces faster responding.

The network also produced reliably higher semantic stress values for the pseudohomophones in the PH/nonPH list compared with their non-

pseudohomophone control nonwords [means: PH = 0.9246, nonPH = 0.9184; paired $t(63)$ = 2.408, P = 0.019]. The network tends to produce greater semantic stress for a pseudohomophone (e.g. HOAP) compared with a control nonword (e.g. JOAP) because the pronunciation it derives for the pseudohomophone (HOPE) was trained to help drive semantic units to extreme value—specifically, the (binary) semantic representation of the base word. Given that higher stress values for nonwords decrease their discriminability from words, this result corresponds to the empirical finding of increased latency and/or error rates to pseudohomophones compared with control nonwords in lexical decision under time pressure (e.g. Coltheart et al., 1977; McCann et al, 1987). Nonetheless, in the absence of time pressure, both types of nonwords are easily discriminated from even the lowest-frequency words. For example, a decision criterion of 0.95 produces 2.5% errors to words (15/600), 6.25% errors to pseudohomophones (4/64) and no errors to nonpseudohomophones (0/64), corresponding to a d' of 3.82.

In summary, a network that maps orthography to semantics both directly and via phonology performed lexical decision accurately if word/nonword decisions were based on a measure of semantic familiarity, termed *stress*, that reflects the degree to which generated semantic patterns are binary. Moreover, the distributions of stress values for different types of words and nonwords accounted for empirical findings concerning the effects of word frequency and nonword pseudohomophony on LD performance. In this way, the findings establish clearly that words can be distinguished from nonwords based on the functional properties of distributed semantic representations without recourse to word-specific structural representations.

GENERAL DISCUSSION

The traditional view of the lexical system stipulates rather complicated and domain-specific structures and processes, including those that apply to individual words but not to nonwords, or to some words (regulars) but not to others (exceptions). The current article adopts an alternative view in which lexical knowledge and processing develop through the operation of general learning principles as applied to distributed representations of written and spoken words and their meanings. Distinctions between words and nonwords, and among different types of words, are not reified in the structure of the system, but rather reflect the functional implications of the statistical structure relating orthographic, phonological and semantic information. The structural divisions within the system are assumed to arise from the neuroanatomic localisation of input and output modalities, not

from differences in representational content (for similar arguments, see
Farah, 1994; Farah & McClelland, 1991).

This distributed view of lexical processing has been championed most
explicitly by Van Orden et al. (1990) and by Seidenberg and McClelland
(1989). Seidenberg and McClelland supported the view with an explicit
computational model, but this support was limited by inadequacies in the
model's ability to account for skilled performance in nonword reading and in
lexical decision (Besner et al., 1990) and impaired performance in fluent
surface dyslexia (Patterson et al., 1989). Plaut et al. (1996) provided a more
adequate account of nonword reading and of surface dyslexia by improving
the orthographic and phonological representations and by incorporating an
influence of semantics on word reading. However, the claimed role of
semantics in reading, as reflected in the account of surface dyslexia, has been
challenged by the performance of patients whose word reading is unaffected
by semantic impairment (e.g. Cipolotti & Warrington, 1995; Lambon Ralph
et al., 1995). Moreover, the limitations of Seidenberg and McClelland's
model in performing lexical decision remained unaddressed.

The current simulations provide additional support for a distributed
theory of lexical processing by addressing these two challenges. The
approach taken to both involves a reconsideration and elaboration of the
role of semantics in naming and lexical decision.

Semantics in Word Naming: Individual Differences

The first simulation demonstrated that parametric variations within reading
models give rise to individual differences in the overall competence and
division of labour between the semantic and phonological pathways, such
that individuals who are able to pronounce low-frequency exception words
without semantic support fall at one end of a continuum. In particular, the
semantic and phonological pathways together learn to support skilled word
reading, but the specific division of labour between the two—particularly for
low-frequency exception words—depends on factors that either improve the
competence of the semantic pathway or impede the competence of the
phonological pathway. Two factors were investigated. The first, the
asymptotic strength of semantic support for phonology, summarises a
variety of factors that would be expected to influence learning in the
semantic pathway, particularly the mapping from orthography to semantics.
The second factor, weight decay in the phonological pathway, can be
thought of as reflecting the degree to which the underlying physiology can
support large numbers of synapses and, hence, strong interactions between
neurons. Simulations demonstrated that the phonological pathway can
pronounce low-frequency exception words without semantics if either
semantic strength or weight decay is particularly low during training, but

that, in general, semantic damage leads to some degree of reading impairment (surface dyslexia).

On the one hand, this explanation would seem to be at odds with claims advanced by Patterson and colleagues (Graham et al., 1994, 1995; Patterson et al., 1994a, b; Patterson and Hodges, 1992) that semantic support is critical to the integrity of phonological representations. The current findings suggest that this property may still hold for most individuals but not for those with highly developed phonological pathways. On the other hand, the explanation for individual differences in division of labour relies critically on the supposition that the semantic and phonological pathways combine to support skilled reading, and that the semantic contribution to phonology has important implications for the nature of learning in the phonological pathway. In this way, even though the current account allows for the possibility that semantics plays a minimal role in the skilled reading performance of some individuals (and, hence, the preservation of reading performance despite semantic damage), it nonetheless incorporates the fundamental insight of Patterson and colleagues—that a consideration of semantic–phonological interactions is critical for a general under- standing of the organisation and operation of the reading system. The account also retains the ability to explain why, in individuals who do exhibit surface dyslexia following semantic damage, there is a close relationship between the comprehension and correct pronunciation of individual exception words (Funnell, 1996; Graham et al., 1994; Hillis & Caramazza, 1991).

The introduction of individual differences into the current explanation raises the question of whether Plaut and co-workers' (1996) account of surface dyslexia is underconstrained. After all, a variety of types and degrees of impairment to the models give rise to the qualitative pattern of surface dyslexia. It should be pointed out, though, that the approach is no different than the traditional dual-route model in this respect (see Coltheart & Funnell, 1987). To be clear, what is in question about the current account is the specific claim regarding the relationship between premorbid differences in division of labour and the quantitative severity of surface dyslexia following semantic or semantic-to-phonological damage. More extensive testing of the non-reading capabilities of surface dyslexic patients is needed to address this concern to establish independent patterns of performance that predict the severity of their reading impairment.

What is *not* left underspecified by the approach is the computational basis for the surface dyslexia pattern itself. This pattern arises directly from the intrinsic sensitivity of learning in distributed networks to word frequency and spelling–sound consistency, as expressed by the frequency–consistency equation (equation 1) for a two-layer Hebbian network. In the normal system, this sensitivity manifests as a frequency $\times$ consistency interaction in

naming latency. Low-frequency exception words (e.g. PINT) are named disproportionately slowly because their vowel pronunciations (I ⇒ /aI/) have the least support from friends and suffer the greatest competition from enemies (e.g. I ⇒ /I/ in PRINT, HINT, MINT, etc.). However, when the performance of the system is limited (e.g. by weight decay and/or strong semantic support during learning, or by direct damage), these weakly supported vowels may now lose the competition to their enemies, resulting in regularisation errors (PINT ⇒ "pihnt"). High-frequency exceptions (e.g. HAVE) are more immune to these influences because their own frequency ($F^{[t]}$ in equation 1) counterbalances the effect of enemies (e.g. GAVE, SAVE, etc.). However, with greater impairment, the vowels in these items also begin to lose the competition and are regularised (HAVE ⇒ "haive"). With still greater impairment, low-frequency regular words (e.g. SOUR) begin to elicit errors. If they have inconsistent bodies, they are also subject to competition from enemies (e.g. POUR, FOUR) and their errors reflect this competition (SOUR ⇒ "sore")—so-called LARC errors (Patterson et al., 1996). Thus, on the current account, the errors made by surface dyslexic patients to all types of words and the pattern of naming latencies exhibited by skilled readers have a common underlying cause: the inherent sensitivity of distributed networks to frequency and consistency. This is why the simulations of surface dyslexia, in particular, account for the full pattern of performance across a range of levels of severity of impairment, even when the specific factors that lead to a particular level of impairment in a specific patient may be in need of further specification.

It is important to emphasise that the postulated individual differences in division of labour do not give rise to all possible patterns of reading impairment following semantic damage, allowing *any* observed pattern to be explained. For instance, no parametric variation in the network will cause it to be severely impaired on low-frequency exception words without also showing some impairment on high-frequency exception words (see Fig. 5), or to be completely unable to read exception words while remaining unimpaired on regular words. (Note that the standard dual-route theory has no difficulty exhibiting either of these patterns.) Rather, the patterns of performance that result from semantic impairment are highly constrained. That is, they all correspond to the specific pattern of fluent surface dyslexia—a frequency × consistency interaction in accuracy with poor reading of low-frequency exception words, high rates of regularisations, and normal nonword reading. Individual differences serve only to locate patients along a single continuum—severity of reading impairment—with the possibility that some, like D.R.N. and D.C., fall at one end of the continuum and are essentially unimpaired. This extreme pattern, however, belies the commonality of the effects that are observed when the full distribution of patients is considered. This commonality derives, on the current account,

from inherent properties of connectionist learning applied to distributed representations. Essentially, the effects of word frequency and spelling–sound consistency are inextricably related because they have the same underlying cause: frequency-weighted sensitivity to input–output similarity, with frequency reflecting the most similar item (the stimulus itself) and consistency reflecting the less similar items (its friends and enemies).

Semantics in Lexical Decision: Stress as a Measure of Familiarity

The second simulation addressed a further challenge to a distributed theory of lexical processing, by demonstrating that, in a network that mapped orthography to semantics directly and via phonology, a measure of familiarity derived from semantic activation—termed *semantic stress*—provide a sufficient basis for distinguishing words from nonwords. Moreover, this measure was shown to account for some aspects of how lexical decision performance is influenced by the nature of the words and nonwords in the task; namely, the frequency blocking effect (Gordon, 1983) and the pseudohomophone effect (Coltheart et al., 1977; McCann et al., 1988).

The stress measure reflects the strength with which semantic units are driven from their "resting" activation level (0.5) towards an extreme value (0 or 1). The reason why semantic stress distinguishes words from nonwords stems from the lack of systematicity in the mapping from orthography to semantics. This lack of systematicity means that orthographic similarity is unpredictive of semantic similarity. Thus, during training, the network must learn to map visually similar words (e.g. HAVE, GAVE, SAVE) to completely unrelated sets of semantic features. The presentation of a nonword (e.g. MAVE) partially engages the mappings for visually similar words, but because these mappings are inconsistent with each other, they generate conflicting input to the semantic units, resulting in only weak (non-binary) semantic activation. By contrast, the considerable systematicity between orthography and phonology enables the mappings for visually similar words to cooperate and collectively produce strongly active, correct pronunciations for nonwords.

Other researchers (e.g. Besner & Joordens, 1995; Borowsky & Masson, 1996; Joordens & Becker, 1997; Masson & Borowsky, 1995; Rueckl, 1995) have proposed performing lexical decision in distributed networks based on the negative of the "energy" in the network ($\Sigma_{i < j} s_i s_j w_{ij}$; Hopfield, 1982). The current stress measure has the advantage of being based entirely on unit activations rather than requiring decision processes to have access to weights on connections between other units. However, it should be pointed out that the two measures are closely related: To the extent that unit states are on the

appropriate side of "rest" given the sign of the weight between them,[6] then increasing stress (by moving states towards their extreme values) will also decrease energy. Thus, the present results would also be expected to hold, at least qualitatively, if lexical decision were based on the negative of energy.

Although the simulation establishes that lexical decision can be performed accurately based on semantics, it should not be interpreted as implying that subjects always do so. In fact, following Seidenberg and McClelland (1989), we assume that subjects can base their decisions on any available information in the lexical system, and that they adopt a strategy that optimises their performance given the composition of the stimuli. We also assume that subjects will rely on orthographic information when this suffices, given that such information is more reliable and more rapidly available than either phonological or semantic information. Thus, the simulation is not intended as a fully adequate account of lexical decision under all conditions. In particular, it is not intended to account for the influence of orthographic factors, such as neighbourhood density, on lexical decision (see, e.g. Andrews, 1992; Sears, Hino, & Lupker, 1995). However, the simulation does serve to demonstrate that subjects can fall back on semantic information when necessary—for example, when orthographically strange words like AISLE and GAUGE or highly word-like pseudohomophones like HOAP and JOAK are included among the stimuli.

An important aspect of LD performance that is not treated in detail in the current work is a specification of the actual processing mechanism that computes semantic stress and uses it to make word/nonword decisions in real time. One of the advantages of stress over an alternative measure of familiarity like energy (Hopfield, 1982) is that the computation of stress does not require access to the values of connection weights between units (which are presumably inaccessible to other units). It is fairly straightforward for a decision process to compute semantic stress if it has access to the semantic units as input. Moreover, Usher and McClelland (1995) have demonstrated recently that competition between linear, stochastic, time-averaging units representing alternative responses gives rise to a number of basic properties of empirical findings in standard choice reaction time tasks, including the shapes of time-accuracy functions, latency-probability functions, hazard functions and reaction-time distributions. Their approach could be applied to generate LD latencies in the current context by creating a "yes" unit whose input is the current level of semantic stress and a "no" unit whose input constitutes a decision criterion that is adjusted to optimise

---

[6]What is meant by "appropriate" here is that both unit states should be on the same side of "rest" if the weight between them is positive and on opposite sides if the weight is negative, assuming that unit states are normalised relative to the rest value (i.e. 0.5 for standard [0,1] units) before entering into the energy equation.

performance, and having the network respond when one of the units exceeds a response criterion.

CONCLUSION

The simulations described in this article, in conjunction with those developed previously (Plaut et al., 1996; Seidenberg & McClelland, 1989), illustrate how connectionist computational principles—distributed representation, structure-sensitive learning, and interactivity—can provide insight into central empirical phenomena in normal and impaired lexical processing. Moreover, they make it clear that distinctions in the function of the lexical system—as manifest in the behaviour of experimental subjects— need not reflect corresponding distinctions in the structure of the system. Thus, networks exhibit word-frequency effects and word/nonword discrimination without word representations, and spelling–sound consistency effects without separate mechanisms for regular and exception items. In this way, gaining insight into the structure and function of the cognitive system by observing its normal and impaired behaviour—the central goal of cognitive psychology and neuropsychology—may depend critically on developing theories and explicit simulations in the context of a specific computational framework that relates structure to function.

The current work demonstrates how the distributed connectionist approach can provide an effective theoretical framework for understanding word naming and lexical decision. This is not to say that the existing distributed models are fully adequate and account for all of the relevant data in sufficient detail—this is certainly not the case. In fact, given that they are *models*, they are abstractions from the actual processing system and are certainly wrong in their details. Nonetheless, their relative success at reproducing key patterns of data in the domain of word reading, and the fact that the very same computational principles are being applied successfully across a wide range of linguistic and cognitive domains, suggests that these models capture important aspects of representation and processing in the human language and cognitive system.

REFERENCES

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory and Cognition*, *10*, 565–575.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 234–254.

Atkinson, R.C., & Juola, J.F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV*, pp. 583–612. New York: Academic Press.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357.

Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, a single lexicon. *Cognitive Neuropsychology*, *9*, 209–258.

Besner, D., & Joordens, S. (1995). Wrestling with ambiguity—Further reflections: Reply to Masson and Borowsky (1995) and Rueckl (1995). *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 515–301.

Besner, D., & Smith, M.C. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 468–482.

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

Borowsky, R., & Masson, M.E.J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 63–85.

Bub, D., Cancelliere, A., & Kertesz, A. (1985). Whole-word and analytic translation of spelling-to-sound in a non-semantic reader. In K. Patterson, M. Coltheart, & J.C. Marshall (Eds), *Surface dyslexia*, pp. 15–34. Hove: Lawrence Erlbaum Associates Ltd.

Chauvin, Y. (1988). *Symbol acquisition in humans and neural (PDP) networks.* PhD thesis, University of California, San Diego, CA.

Cipolotti, L., & Warrington, E.K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society*, *1*, 104–110.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing*, pp. 151–216. New York: Academic Press.

Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In M.I. Posner & O.S.M. Marin (Eds), *Attention and performance XI*, pp. 3–37. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Coltheart, M., & Funnell, E. (1987). Reading writing: One lexicon or two? In D.A. Allport, D.G. MacKay, W. Printz, & E. Scheerer (Eds), *Language perception and production: Shared mechanisms in listening, speaking, reading and writing*, pp. 313–339. New York: Academic Press.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI*, pp. 535–555. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Coslett, H.B. (1991). Read but not write "idea": Evidence for a third reading mechanism. *Brain and Language*, *40*, 425–443.

Cottrell, G.W., & Plunkett, K. (1991). Learning the past tense in a recurrent network: Acquiring the mapping from meaning to sounds. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society*, pp. 328–333. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cummings, J.L., Houlihan, J.P., & Hill, M.A. (1986). The pattern of reading deterioration in dementia of Alzheimer type: Observations and implications. *Brain and Language*, *29*, 315–323.

Dunn, L.M. (1965). *The Peabody picture vocabulary test*. Circle Pines: American Guidance Service.

Farah, M.J. (1994). Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behavioral and Brain Sciences*, *17*, 43–104.

Farah, M.J., & McClelland, J.L. (1991). A computational model of semantic memory impairment: Modality-specificity and emergent category-specificity. *Journal of Experimental Psychology: General*, *120*, 339–357.

Fera, P., & Besner, D. (1992). The process of lexical decision: More words about a parallel distributed processing model. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 749–764.

Franklin, S., Howard, D., & Patterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology* , *12*, 549–566.

Friedman, R.B., Ferguson, S., Robinson, S., & Sunderland, T. (1992). Dissociation of mechanisms of reading in Alzheimer's disease. *Brain and Language*, *43*, 400–413.

Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology*, *49A*, 417–314.

Glanzer, M., & Ehrenreich, S.L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *18*, 381–398.

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.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, *22*, 24–44.

Graham, K.S., Hodges, J.R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia* , *32*, 299–316.

Graham, K.S., Patterson, K., & Hodges, J.R. (1995). Progressive pure anomia: Insufficient activation of phonology by meaning. *Neurocase*, *1*, 25–38.

Hillis, A.E., & Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, *114*, 2081–2094.

Hinton, G.E. (1989). Connectionist learning procedures. *Artificial Intelligence*, *40*, 185–234.

Hinton, G.E., McClelland, J.L., & Rumelhart, D.E. (1986). Distributed representations. In D.E. Rumelhart, J.L. McClelland, & the PDP Research Group (Eds), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*, pp. 77–109. Cambridge, MA: MIT Press.

Hodges, J.R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, *115*, 1783–1806.

Hoeffner, J. (1992). Are rules a thing of the past? The acquisition of verbal morphology by an attractor network. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pp. 861–866. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science*, *USA*, *79*, 2554–2558.

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

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 1083–1105.

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1051–1062.

Kawamoto, A. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S.L. Small, G.W. Cottrell, & M.K. Tanenhaus (Eds), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* . San Mateo, CA: Morgan Kaufmann.

Kawamoto, A.H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing approach. *Journal of Memory and Language*, *32*, 474–516.

Kawamoto, A.H., Farrar, W.T., & Kello, C.T. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1233–1247.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Palpa: Psycholinguistic assessments of language processing in aphasia*. Hove: Lawrence Erlbaum Associates Ltd.

Kučera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lambon Ralph, M., Ellis, A.W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase*, *1*, 363–369.

Marshall, J.C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*, 175–199.

Masson, M.E.J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 3–23.

Masson, M.E.J., & Borowsky, R. (1995). Unsettling questions about semantic ambiguity in connectionist models: Comment on Joordens and Besner (1994). *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 509–514.

McCann, R.S., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 693–706.

McCarthy, R., & Warrington, E.K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex*, *22*, 359–380.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

Meyer, D.E., Schvaneveldt, R.W., & Ruddy, M.G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory and Cognition*, *2*, 309–321.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165–178.

Morton, J., & Patterson, K. (1980). A new attempt at an interpretation; or, an attempt at a new interpretation. In M. Coltheart, K. Patterson, & J.C. Marshall (Eds), *Deep dyslexia*, pp. 91–118. London: Routledge and Kegan Paul.

Paap, K.R., & Noel, R.W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, *53*, 13–24.

Patterson, K., & Hodges, J.R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *30*, 1025–1040.

Patterson, K., Coltheart, M., & Marshall, J.C. (Eds) (1985). *Surface dyslexia*. Hove: Lawrence Erlbaum Associates Ltd.

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. Oxford: Oxford University Press.

Patterson, K., Graham, N., & Hodges, J.R. (1994a). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience*, *6*, 57–69.

Patterson, K., Graham, N., & Hodges, J.R. (1994b). Reading in Alzheimer's type dementia: A preserved ability? *Neuropsychology*, *8*, 395–412.

Patterson, K., Plaut, D.C., McClelland, J.L., Seidenberg, M.S., Behrmann, M., & Hodges, J.R. (1996). Connections and disconnections: A connectionist account of surface dyslexia. In J. Reggia, R. Berndt, & E. Ruppin (Eds), *Neural modeling of cognitive and brain disorders*, pp. 177–199. New York: World Scientific.

Plaut, D.C. (1995a). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 291–321.

Plaut, D.C. (1995b). Semantic and associative priming in a distributed attractor network. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pp. 37–42. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Plaut, D.C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language*, *52*, 25–82.

Plaut, D.C., & McClelland, J.L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pp. 824–829. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Raymer, A.M., & Berndt, R.S. (1994). Models of word reading: Evidence from Alzheimer's disease. *Brain and Language*, *47*, 479–482.

Rueckl, J.G. (1995). Ambiguity and connectionist networks: Still settling into a solution. Comment on Joordens and Besner (1994). *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 501–508.

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

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986b). Learning representations by back-propagating errors. *Nature*, *323*, 533–536.

Schwartz, M.F. (Ed.) (1990). *Modular deficits in Alzheimer-type dementia*. Cambridge, MA: MIT Press.

Schwartz, M.F., Marin, O.S.M., & Saffran, E.M. (1979). Dissociations of language function in dementia: A case study. *Brain and Language*, *7*, 277–306.

Schwartz, M.F., Saffran, E.M., & Marin, O.S.M. (1980). Fractioning the reading process in dementia: Evidence for word-specific print-to-sound associations. In M. Coltheart, K. Patterson, & J.C. Marshall (Eds), *Deep dyslexia*, pp. 259–269. London: Routledge and Kegan Paul.

Sears, C.R., Hino, Y., & Lupker, S.J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 876–900.

Seidenberg, M.S. (1992). Beyond orthographic depth: Equitable division of labour. In R. Frost & K. Katz (Eds), *Orthography, phonology, morphology, and meaning*, pp. 85–118. Amsterdam: Elsevier.

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

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 and Verbal Behaviour*, *23*, 383–404.

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

Seidenberg, M.S., Petersen, A., MacDonald, M.C., & Plaut, D.C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 48–62.

Shallice, T., Warrington, E.K., & McCarthy, R. (1983). Reading without semantics. *Quarterly Journal of Experimental Psychology*, *35A*, 111–138.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 6, 174–215.

Snowden, J.S., Goulding, P.J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioral Neurology*, 2, 167–182.

Strain, E., Patterson, K., & Seidenberg, M.S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 1140–1154.

Taraban, R., & McClelland, J.L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language*, 26, 608–631.

Usher, M., & McClelland, J.L. (1995). *On the time course of perceptual choice: A model based on principles of neural computation*. Technical Report PDP.CNS.95.5. Pittsburgh, PA: Carnegie Mellon University, Department of Psychology.

Van Orden, G.C., & Goldinger, S.D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1269.

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

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

Watt, S., Jokel, R., & Behrmann, M. (1997). Surface dyslexia in progressive aphasia. *Brain and Language*, 56, 211–233.

Wickelgren, W.A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76, 1–15.