# Computational Modeling of Word Reading, Acquired Dyslexia, and Remediation

David C. Plaut
Carnegie Mellon University and the Center for the Neural Basis of Cognition

In R. M. Klein & P. A. McMullen (Eds.), *Converging methods in reading and dyslexia* (pp. 339-372). Cambridge, MA: MIT Press, 1999.

Many researchers assume that the most appropriate way to express the systematic aspects of language is in terms of a set of rules. For instance, there is a systematic relationship between the written and spoken forms of most English words (e.g., GAVE ⇒ /geɪv/), and this relationship can be expressed in terms of a fairly concise set of grapheme-phoneme correspondence (GPC) rules (e.g., G ⇒ /g/, A_E ⇒ /eɪ/, V ⇒ /v/). In addition to being able to generate accurate pronunciations of so-called *regular* words, such rules also provide a straightforward account of how skilled readers apply their knowledge to novel items—for example, in pronouncing word-like nonwords (e.g., MAVE ⇒ /meɪv/). Most linguistic domains, however, are only partially systematic. Thus, there are many English words whose pronunciations violate the standard GPC rules (e.g., HAVE ⇒ /hæv/). Given that skilled readers can pronounce such *exception* words correctly, GPC rules alone are insufficient. More generally, skilled language performance at virtually every level of analysis—phonological, morphological, lexical, syntactic—requires both effective handling of exceptional items and the ability to generalize to novel forms.

In the domain of reading, there are three broad responses to this challenge. The first, adopted by traditional "dual-route" theories (Besner, this volume; Besner & Smith, 1992; Coltheart, 1978; 1985; Coltheart, Curtis, Atkins, & Haller, 1993; Marshall & Newcombe, 1973; Meyer, Schvaneveldt, & Ruddy, 1974; Morton & Patterson, 1980; Paap & Noel, 1991), is to add to the GPC system a separate, *lexical* system that handles the exceptions. The second response, adopted by "multiple levels" theories (Norris, 1994; Shallice & McCarthy, 1985; Shallice, Warrington, & McCarthy, 1983), is to augment the GPC rules with more specific, context-sensitive rules, (e.g., OOK ⇒ /ʊk/ as in BOOK), including rules that apply only to individual exceptions (e.g., PINT ⇒ /paɪnt/). Both of these approaches retain the general notion that language knowledge takes the form of rules (although such rules may be expressed in terms of connections between localist connectionist units; see, e.g., Norris, 1994; Reggia, Marsland, & Berndt, 1988).

The third response to the challenge, adopted by distributed connectionist theories (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990; Van Orden & Goldinger, 1994) and elaborated in the current chapter, is more radical. It eschews the notion that the knowledge supporting online language performance takes the form of explicit rules. Of course, such performance can certainly be described approximately in terms of rules, and language users can sometimes even state verbally certain of these "rules" (i.e., "I before E…"). The connectionist claim is that the language mechanism itself does not contain a set of rules and a rule interpreter. Rather, language knowledge is inherently graded, and the language mechanism is inherently a learning device that gradually picks up on the statistical structure among written and spoken words and the contexts in which they occur. On this view, there is no sharp division between the regular items which obey the rules and the exception items which violate them. Instead, the emphasis is on the degree to which the mappings among the spelling, sound, and meaning of a given word are *consistent* with those of other words (Glushko, 1979).[1]

[1] The relationship between regularity and consistency is often a source of confusion. *Regularity* is a dichotomous variable that expresses whether or not the pronunciation of a given word obeys a particular set of spelling-sound correspondence rules. Such rules are most typically described as grapheme-phoneme correspondence (GPC) rules, although the only set of spelling-sound rules that have actually been implemented (Coltheart et al., 1993) involve a considerably greater degree of context sensitivity. By contrast, *consistency* is a continuous variable that expresses the degree to which the pronunciation of a word agrees with those of similarly spelled words. Here, similarity is typically cast in terms of word endings or *bodies* (i.e., the vowel and any following consonants), in part on the basis of empirical evidence that this unit accounts for considerable variance in monosyllabic word pronunciations (see Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). Of course, similarity in terms of smaller orthographic and phonological units—including graphemes and phonemes—would also be expected to influence performance. Consequently, although the terms regularity and consistency entail rather different theoretical commitments concerning the nature of spelling-sound knowledge, their empirical implications are

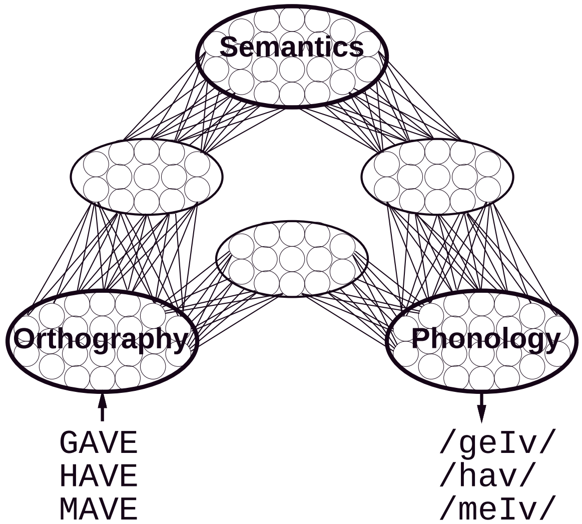


Figure 1: A connectionist framework for lexical processing, based on that of Seidenberg and McClelland (1989).

To make this third perspective concrete, consider the connectionist/parallel distributed processing (PDP) framework for lexical processing depicted in Figure 1, based on that of Seidenberg and McClelland (1989). 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 the figure makes clear, the connectionist approach does not entail a complete lack of structure within the reading 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 differentially 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 will turn out to be critical 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. To be clear, the essence of a dual-route theory is *not* that it has two pathways from print to sound (in fact, most such theories have three such pathways: sublexical, lexical semantic, and lexical nonsemantic; see Besner, this volume); rather, it is the claim that the mechanism that processes nonwords (typically GPC rules and an interpreter) is functionally distinct from, and operates according to different principles than, the mechanism that processes exception words (typically a look-up table or an associative network). It is this inhomogeneity of processing mechanisms that the distributed connectionist approach rejects.

In this context, it is important to note that it is perfectly feasible to build a dual-route mechanism out of connectionist hardware. For example, Zorzi, Houghton, and Butterworth (1998) have recently described simulations in which direct connections from letter units to phoneme units support the pronunciation of regular words and nonword, whereas a separate pathway, composed either of hidden units or localist word units, supports the pronunciation of exception words (also see Reggia et al., 1988). Although the mechanisms employed for the two pathways are more homogeneous than in more traditional, rule-based implementations (e.g., Coltheart et al., 1993), the models nonetheless retain a categorical distinction between items which obey spelling-sound rules and items which violate them.

This chapter describes a series of computational simulations based on the framework depicted in Figure 1. The value of computational modeling is greatest when properties of the formalism guide and constrain simulation work, and lead to insight into counterintuitive findings. Along these lines, one can identify three general computational principles on which the current connectionist approach to word reading is based:

**Distributed representation:** Orthography, phonology, and semantics are represented by distributed patterns of activity such that similar words are represented by similar patterns.[2]

notoriously difficult to distinguish.

[2] A representation is *localist* if there is a one-to-one relationship between processing units and entities in the domain; it is *distributed* if the relationship is many-to-many (i.e., each entity activates many units and each unit participates in representing many entities). Thus, a representation is localist or distributed *only relative to a specific set of entities.* For

**Structure-sensitive learning:** Knowledge of the relationships among orthography, phonology, and semantics is encoded across connection weights that are learned gradually through repeated experience with words in a way that is sensitive to the *statistical structure* of each mapping.

**Interactivity:** Mapping among orthography, phonology, and semantics is accomplished through the simultaneous interaction of many units, such that familiar patterns form stable *attractors*.[3]

These principles are claimed to be general in that versions of them apply across all cognitive domains. However, the challenge is to apply the principles to account for detailed behavioral data. The goal of the current chapter is to demonstrate that, when instantiated in a particular domain—single word reading—these principles provide important insights into the patterns of normal and impaired cognitive behavior. Word reading is a particularly appropriate domain of study because there is a wealth of detailed empirical data on reading acquisition and developmental dyslexia, normal skilled reading, acquired dyslexia from brain damage, and rehabilitation after brain damage. We will touch on each of these areas, but will focus on issues in normal reading, acquired dyslexia, and rehabilitation.

# 1 Skilled Oral Reading

Although the distributed connectionist framework for word reading depicted in Figure 1 may seem reasonable at a general level, it actually reflects a radical departure from traditional theorizing about lexical processing, particularly in two ways. First, there is nothing in the structure of the system that corresponds to individual words *per se*, such as a lexical entry, localist word unit (McClelland & Rumelhart, 1981) or "logogen" (Morton, 1969). Rather, words are distinguished from nonwords only by *functional* properties of the system—the way in which particular orthographic, phonological, and semantic patterns of activity interact (also see Plaut, 1997; Van Orden et al., 1990). Second, there are no separate mechanisms for lexical and sublexical processing (cf. Coltheart et al., 1993), such that, for instance, regular words (e.g., MINT) are pronounced by one route and exceptions (e.g., PINT) by another. Rather, all parts of the system participate in processing all types of input, although, of course, the contributions of different parts may be more or less important for different inputs.

In an attempt to demonstrate that the structural reification of words and of lexical/sublexical processing routes is unnecessary to account for skilled oral reading, Seidenberg and McClelland (1989) trained a connectionist network to map from the orthography of about 3000 monosyllabic English words—both regular and exception—to their phonology (i.e., the bottom portion of the framework in Figure 1, referred to as the *phonological pathway*). After training, the network pronounced correctly 97.7% of the words, 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, 1984a; Taraban & McClelland, 1987; Waters & Seidenberg, 1985) if its real-valued accuracy in generating a response is taken as a proxy for response time (under the assumption that an imprecise phonological representation would be less effective at driving an articulatory system). However, the model was much worse than skilled readers at pronouncing orthographically legal nonwords (Besner, Twilley, McCann, & Seergobin, 1990) and at lexical decision under some conditions (Besner et al., 1990; Fera & Besner, 1992). Thus, the model failed to refute traditional claims that localist, word-specific representations and separate mechanisms are necessary to account for skilled reading.

More recently, Plaut, McClelland, Seidenberg, and Patterson (1996, also see Plaut & McClelland, 1993; Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994) have shown that the limitations of the Seidenberg and McClelland model in pronouncing nonwords stems not from any general limitation in the abilities of connectionist networks in quasi-regular domains (as suggested by, e.g., Coltheart et al., 1993), but from its use of poorly structured orthographic and phonological representations. The original simulation used representations based on context-sensitive triples of letters or phonemic features. When more appropriately structured representations are used—based on graphemes and phonemes and embodying phonotactic and graphotactic constraints—network implementations of the phonological pathway can learn to pronounce regular words, exception words, and nonwords as well as skilled readers. Furthermore, the networks also exhibit the empirical frequency-by-consistency interaction pattern when trained on actual word frequencies.[4]

example, the letter layer of the Interactive Activation model (McClelland & Rumelhart, 1981) is localist with respect to letters but distributed with respect to words. Despite the terminology used by Besner (this volume), both localist and distributed models can be "connectionist" in the sense that the system's knowledge is encoded in terms of weights on connections between simple, neuron-like processing units (Feldman & Ballard, 1982).

[3] An attractor is a stable pattern of activity within a network such that unit interactions cause similar patterns to move towards and settle into the exact attractor pattern.

[4] Seidenberg and McClelland (1989) trained their model using logarithmically compressed word frequencies in order to ensure sufficient sampling of the lowest frequency words. Plaut et al. (1996) avoided this problem by using word frequency to scale weight changes directly.

This remains true if naming latencies are modeled directly by the settling time of a recurrent, attractor network (see Figure 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 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 learning—retained many of the essential characteristics of the more general framework (e.g., distributed representations and structure-sensitive learning). For this simplified network, it was possible to derive an expression for how the response of the network to any input (test) pattern depends on its experience with every pattern on which the network is trained, as a function of its 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

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

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 $t$ (similar patterns trained to produce the same response for unit $j$), each weighted by its similarity (overlap) with $t$, $O^{[ft]}$; and (3) minus the sum of the frequencies $F^{[e]}$ of the *enemies* of pattern $t$ (similar patterns trained to produce the opposite response), each weighted by its similarity to $t$, $O^{[et]}$.

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 (see Figure 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. Equation 1 is only approximate, however, for more complex networks—those with hidden units and trained with an error-correcting algorithm like

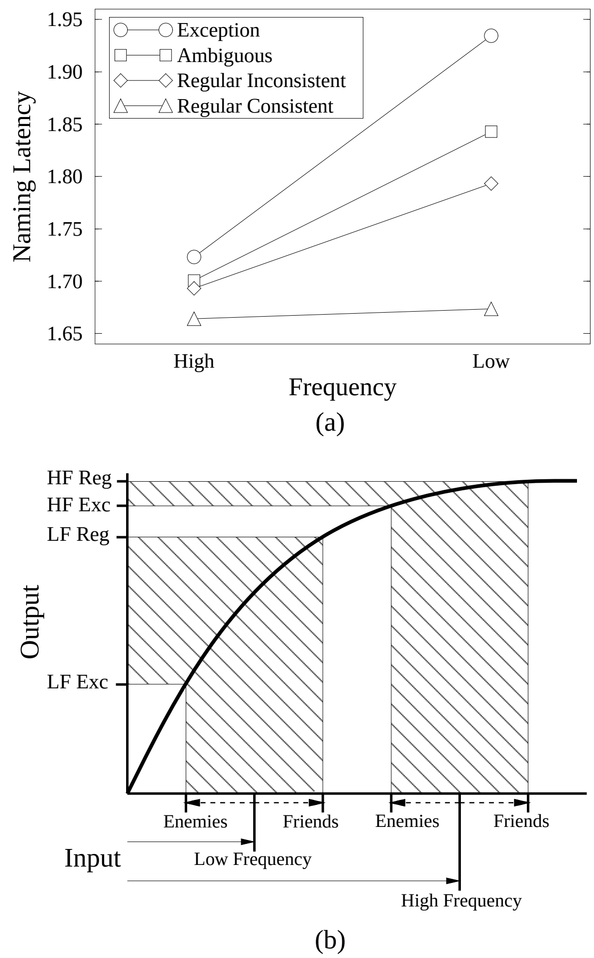


Figure 2: (a) The frequency-by-consistency interaction exhibited in the settling time of an attractor network implementation of the phonological pathway in pronouncing words of varying frequency and spelling-sound consistency (Plaut et al., 1996, Simulation 3); and (b) its explanation in terms of additive contributions of frequency and consistency subject to an asymptotic activation function (only the top of which is shown).

back-propagation. 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.

# 2 Impaired Oral Reading in Surface Dyslexia

Although Plaut et al. (1996) 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 (also see Harm, 1998; Hillis & Caramazza, 1991; Van Orden & Goldinger, 1994), and that individuals may differ in the relative competence of each pathway (Plaut, 1997; 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-by-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 an abnormal pattern of reading performance, known as *surface dyslexia* (see Patterson, Coltheart, & Marshall, 1985). In its purest, so-called *fluent* form (e.g., MP, Behrmann & Bub, 1992; Bub, Cancelliere, & Kertesz, 1985; KT, McCarthy & Warrington, 1986; HTR, Shallice et al., 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 that shown by normal subjects in their naming latencies. 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 *regularization* 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 regularized (Graham, Hodges, & Patterson, 1994; Hillis & Caramazza, 1991). Moreover, 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, Marin, & Saffran, 1979) or dementia of the Alzheimer's type (Balota & Ferraro, 1993; Patterson, Graham, & Hodges, 1994).

The framework for lexical processing depicted in Figure 1 (and the implied computational principles) provides a natural formulation of how contributions from both the semantic and phonological pathways might be integrated in oral reading. At an abstract level, given that phonological units simply sum their inputs from the two pathways, the influence of the semantic pathway can be included in a straightforward manner by adding an additional term, $S^{[t]}$, to the summed input in Equation 1. Furthermore, if we assume that this term increases with imageability, this accounts for the three-way interaction of frequency, consistency, and imageability found by Strain et al., 1995. When formulated explicitly in connectionist terms, however, this integration has far-reaching implications for the nature of *learning* in the two pathways. During training, to the extent that the contribution of one pathway reduces the overall error, the other will experience less pressure to learn. Specifically, if the semantic pathway contributes significantly to the pronunciation of words, then the phonological pathway need not learn to pronounce all of the words by itself. Rather, this pathway will tend to learn best those words high in frequency and/or consistency (i.e., those items with large positive terms in Equation 1); on its own it may never master low-frequency exception words completely. Of course, in skilled readers, the *combination* of the semantic and phonological pathways will be fully competent. But brain damage that reduced or eliminated the semantic pathway would lay bare the latent inadequacies of the phonological pathway, giving rise to surface dyslexia.

In further simulations, Plaut et al. (1996; Plaut, 1997) 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 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 oral reading by providing the output (phoneme) units of the the phonological pathway with external input that pushed them towards the correct pronunciation of each word during training. Semantic damage, then, was modeled 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 removed and, moreover, that individual differences in the severity of surface dyslexia can arise, not only from differences in the amount of semantic damage, but also from *premorbid* differences in the division of labor between the semantic and phonological pathways. This division of labor—and the overall competence 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, relative experience in reading aloud versus silently, the computational resources (e.g., numbers of units and connections) devoted to each pathway, and the reader's more general skill levels in visual pattern recognition and in spoken word comprehension and production. Thus, the few patients exhibiting mild to moderate semantic impairments without concomitant regularization errors (early WLP, Schwartz et al., 1979; MB, Raymer & Berndt, 1994; DRN, Cipolotti & Warrington, 1995; DC, Lambon Ralph, Ellis, & Franklin, 1995) may have, for various reasons, reading systems with relatively weak reliance on the semantic pathway (see Plaut, 1997, for relevant simulations and discussion).

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 behavior 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. Unfortunately, while emphasizing the importance of semantics, the Plaut et al. simulations offer little insight into the specific nature of semantic representations and processes; the simulations of surface dyslexia in particular are limited by the lack of an actual implementation of the semantic pathway (see, however, Harm, 1998). Moreover, without such an implementation, Plaut et al. were also unable to remedy the limitations of the Seidenberg and McClelland (1989) model in performing lexical decision.

In fact, given that the relationship between a (monomorphemic) word and its meaning is essentially arbitrary, it might seem that an implementation of the semantic pathway would require the use of word-specific representations. However, implementations of the semantic pathway using distributed representations have been pursued on a smaller scale in the context of modeling semantic and associative priming in lexical decision (Plaut, 1995b), impaired reading via meaning in deep dyslexic patients (Plaut & Shallice, 1993), and remediation of semantics by retraining after damage (Plaut, 1996). The next three sections take up the issues in each of these domains in turn.

## 3 Semantic and Associative Priming in Lexical Decision

In a variety of lexical tasks, including naming and lexical decision, subjects are faster and more accurate to process a word, such as BUTTER, when it is preceded by a semantically related word like BREAD relative to an unrelated word like HOUSE (e.g., Meyer & Schvaneveldt, 1971, see Neely, 1991, for a review). This semantic priming effect is influenced by a number of stimulus factors, including perceptual factors like visual quality (greater for visually degraded stimuli; see, e.g., Becker & Killion, 1977; Meyer, Schvaneveldt, & Ruddy, 1975), lexical factors like word frequency (greater for low-frequency targets; see, e.g., Becker, 1979), and semantic factors like category dominance (greater for high-dominance exemplars; see, e.g., Lorch, Balota, & Stamm, 1986). Priming also varies with the stimulus onset asynchrony (SOA) between prime and target and is subject to both facilitation and inhibition effects (e.g., Neely, 1977). Such findings are taken by many theorists as reflecting fundamental properties of the organization of semantic knowledge.

To provide an account for these findings, Plaut (1995b) trained a distributed attractor network on an artificial version of the task of deriving the meanings of written words (i.e., mapping orthography to semantics in Figure 1). As is standard in distributed network models (e.g., Kawamoto, 1988; Masson, 1991; 1995; McRae, de Sa, & Seidenberg, 1993; Sharkey & Sharkey, 1992), semantic relatedness among words was reflected in the degree of overlap of their semantic features. These models typically employ only a single, symmetric manipulation—pattern overlap—to encode word relatedness, so there is no opportunity for different types of relations among words to behave differently. In particular, one can distinguish an *associative* relation among words (e.g., as measured by free association norms; Postman & Keppel, 1970) from a purely *semantic* relation (i.e., having similar meanings, such as category co-ordinates). These two types of relations have been shown to give rise to different empirical effects in a number of contexts (e.g., Becker, 1980; Glosser & Friedman, 1991; Moss & Marslen-Wilson, 1993; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Seidenberg, Waters, Sanders, & Langer, 1984b). In Plaut's (1995b) simulation, semantic relatedness was encoded in terms of degree of pattern overlap (as in most distributed models) but an association from one word to another was encoded in a different manner: in the likelihood that the one follows the other during training (see Moss, Hare, Day, & Tyler, 1994, for a similar approach).

Semantic patterns were constructed to form categories by generating 8 random prototype patterns (e.g., *bird*), such that each feature had a probability of 0.1 of being active. Sixteen category exemplars were then generated from each prototype by randomly changing some of its features; fewer for high-dominance exemplars (e.g., *robin*) than for low-dominance exemplars (e.g., *goose*). The resulting 128 semantic representations were randomly assigned orthographic representations consisting of patterns of activity over 20 orthographic units. These patterns were generated randomly such that each unit had a probability of 0.1 of being active, with the constraint that every pattern had at least two active units, and all pairs of patterns differed in the activities of at least two units. No attempt was made to model orthographic relatedness among words; the orthographic patterns simply guaranteed that the written forms of words were fairly sparse and were discriminable from each other. The critical property of the task was that, although there were systematic relationships among word meanings, there was no systematic relationship between the written form of a word and its meaning.

During training, the network started from the final activity pattern produced by the previous word in processing the next word. Often the next word chosen was the associate of the previous word. On the remaining trials, the probability that words were selected for training depended on their assigned frequency, such that high-frequency words were twice as likely to be trained as low-frequency words. To encourage robust performance, each orthographic pattern was corrupted slightly with Gaussian noise (*SD*=0.05) when presented for training. The network was trained with a continuous version of back-propagation through time (Pearlmutter, 1989) to activate the word's semantic features as quickly as possible when presented with its orthography.

After training, the network was tested for priming effects by presenting a prime word for some specified duration, then replacing it with the target word and allowing the network to settle to the appropriate semantic representation for the target. The prime could be semantically related, associatively related, or unrelated to the target. The RT of the network to the target was defined as the time it took the network to settle to the point where no semantic unit changed it state by more than 0.001. The difference in RT values for unrelated versus related primes constitutes a measure of (semantic or associative) priming.

When trained and tested in this manner, the network exhibited two empirical effects that have posed problems for other distributed network theories of priming (e.g., Kawamoto, 1988; Masson, 1991; 1995; McRae et al., 1993; Sharkey & Sharkey, 1992): much stronger associative priming than semantic priming (e.g., Becker, 1980; Shelton & Martin, 1992), and significant associative priming across an intervening unrelated item (e.g., Joordens & Besner, 1992; McNamara, 1994).[5] It also reproduced the empirical findings of greater priming for low-frequency targets, degraded targets, and high-dominance category exemplars.

Although not reported by Plaut (1995b), the network's performance on words provides a reliable basis for performing lexical decision (LD). A natural way to perform LD is on the basis of some measure of the *familiarity* of the stimulus (Balota & Chumbley, 1984). A commonly used measure of familiarity in distributed networks is the negative of the *energy*, $\sum_{i<j} s_i s_j w_{ij}$ (Hopfield, 1982), which reflects the degree to which unit states satisfy the soft constraints imposed by the weights. A number of researchers (e.g., Besner & Joordens, 1995; Borowsky & Masson, 1996; Masson & Borowsky, 1995; Rueckl, 1995, also see Masson, this volume) have proposed recently that it may be possible to perform LD on the basis of differences in the energy of 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 synaptic strengths 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 is different from rest:

$$S_j = s_j \log_2(s_j) + (1 - s_j) \log_2(1 - s_j) - \log_2(0.5) \quad (2)$$

The stress of a unit is 0 when its state is 0.5 and approaches 1 as its state approaches either 0.0 or 1.0. The target semantic patterns for words are binary vectors (i.e., consist of 1s and 0s) 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 orthographic features with words that have conflicting semantic features. As a result, nonwords typically fail to drive semantic units as strongly, producing semantic patterns with much lower average stress. Thus, the average stress of semantic units, here termed simply *semantic stress*, forms a reliable basis for performing LD.[6]

To demonstrate the adequacy of this approach, 128 nonwords were created by generating new orthographic patterns in the same manner as the trained patterns, ensuring

[5] Consistent with Plaut's (1995b) account, Masson (1995 and this volume) considered the possibility that the intervening word might be processed only partially, leaving residual semantic activation from BREAD to influence BUTTER. Using a Hopfield (1982) network, Masson simulated the small priming effect across unrelated words in a naming task by basing the network's response on the activity of phonological units which were updated more frequently than semantic units. However, the simulations used a very small vocabulary (only three pairs of semantically related items), and no independent justification was provided for why phonological and semantic units should behave differently.

[6] It is assumed that LD responses are actually generated by a stochastic decision process (e.g., Ratcliff, 1978; Usher & McClelland, 1995) that computes stress by integrating over semantic unit states and that adopts a decision criterion such that stress values farther from the criterion are responded to more quickly and accurately. In a given experimental context, a specific criterion is chosen that allows fast responding with acceptable error rates depending on the composition of the word and nonword stimuli.

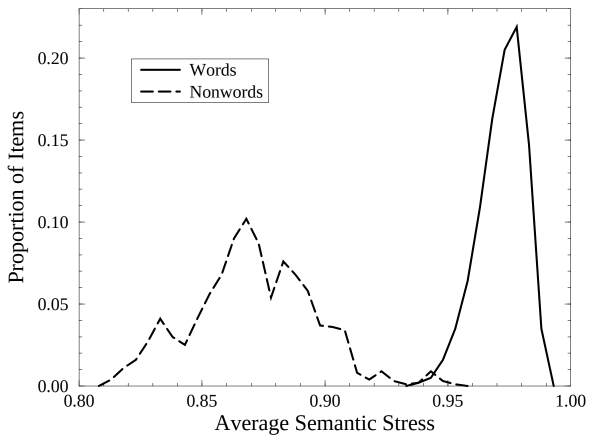


Figure 3: The distributions of average stress of semantic units for words and nonwords exhibited by the Plaut (1995b) network.

that each nonword differs from every word by at least two units (i.e., the same constraint that applies between any two words). These patterns function as nonwords in that they have the same statistical structure as the orthographic patterns for words but were never presented to the network during training. The performance of the network was tested on each word and nonword as target when preceded by each unrelated word as prime (over a range of prime durations). Using a response criterion by which a "yes" response was given if the semantic stress after settling exceeded 0.945, the network made only 0.70% misses and 0.38% false-alarms in LD, yielding a $d'$ of 5.12 (see Figure 3). Thus, a distributed network that maps orthography to semantics can account for the ability of skilled readers to reliably distinguish word from word-like nonwords, based on a measure of the degree to which semantic units receive strong input that drives them to near-binary states.

More recently, Plaut (1997) has successfully applied this approach to model LD on actual word and nonword stimuli. A feedforward network was trained to map from the orthographic representations of the 2998 monosyllabic words in the Plaut et al. (1996) corpus to their phonological representations and to artificially created semantic representations like those described above. After 1300 epochs of training, the network accurately derived the phonology and semantics of each word when presented with its orthography. It was then tested for its ability to distinguish these words from 591 pronounceable nonwords (Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994) on the basis of the levels of semantic stress produced by each type of stimuli. When a "yes" criterion of 0.955 was adopted, the network produced only 0.90% (27/2998) misses and 1.52% (9/582) false-alarms, corresponding to a $d'$ of 4.53. In a second test involving 64 pseudohomophones (PH; e.g., JOAK) and closely matched nonpseudohomophone (nonPH) control nonwords (e.g., HOAK; Seidenberg, Petersen, MacDonald, & Plaut, 1996), the network produced reliably higher semantic stress values—and thus poorer discrimination from words—for the pseudohomophones (means: PH = 0.9246, nonPH = 0.9184; paired t[63] = 2.408, p = .019). Thus, the network exhibited accurate performance overall as well as the empirical finding of a pseudohomophone disadvantage in LD (see Besner, this volume; Coltheart, Davelaar, Jonasson, & Besner, 1977; McCann & Besner, 1987).

The Plaut (1995b) and Plaut (1997) simulations focus on normal skilled performance of lexical decision. Other simulations of the operation of the semantic pathway have attempted to address patterns of impaired performance among certain types of brain-damaged patients, and how such impairments might be remediated.

## 4 Impaired Reading via Meaning in Deep and Phonological Dyslexia

As one might expect from the name, patients with *deep dyslexia* (see Coltheart, Patterson, & Marshall, 1980) have reading impairments that are in many ways opposite to those with surface dyslexia, in that they appear to read almost entirely via semantics. Deep dyslexic patients are thought to have severe (perhaps complete) damage of the phonological pathway, as evidenced by their inability to read even the simplest of pronounceable nonwords (but see Buchanan, Hildebrandt, & MacKinnon, 1994a; 1994b; 1996; this volume). They also have impairments in reading words that suggest additional partial damage to the semantic pathway. In particular, the hallmark symptom of deep dyslexia is the occurrence of *semantic* errors in oral reading (e.g., reading CAT as "dog"). Strangely, these semantic errors co-occur with a peculiar combination of other symptoms. Central among these are other errors that involve visual similarity: pure *visual* errors (e.g., CAT ⇒ "cot"), mixed *visual-and-semantic* errors (e.g., CAT ⇒ "rat"), and even mediated *visual-then-semantic* errors (e.g., SYMPATHY ⇒ "orchestra", presumably via *symphony*). Furthermore, the likelihood that a word is read correctly depends on its part-of-speech (nouns > adjectives > verbs > function words) and its concreteness or imageability (concrete, imageable words > abstract, less imageable words). Finally, differences across patients in written and spoken comprehension, and in the distribution of error types, suggests that the secondary damage to the semantic pathway may occur before, within, or after semantics (Shallice & Warrington, 1980).

Deep dyslexia is closely related to another type of acquired dyslexia—so-called *phonological dyslexia* (Beauvois & Derouesné, 1979). The defining characteristic of phonological dyslexic patients is that they have a selective impairment in reading nonwords compared with reading words. Although such patients do not produce above-chance rates of semantic errors, they can be quite similar to deep dyslexic patients in other respects. In fact, Glosser and Friedman (1990, also see Buchanan, Hildebrandt, & MacKinnon, this volume) argued that deep and phonological dyslexic patients fall on a continuum of severity of impairment, with deep dyslexia at the most severe end. Moreover, Friedman (1996, also see Klein, Behrmann, & Doctor, 1994) has argued that the symptoms in deep dyslexia resolve in a particular order over the course of recovery, reflecting the continuum of impairment. The occurrence of semantic errors is the first symptom to resolve, constituting a somewhat arbitrary transition from deep to phonological dyslexia). The concreteness effect is the next symptom to resolve, followed by the part-of-speech effect, then the visual and morphological errors, and only lastly, the impaired nonword reading. A similar pattern of recovery has been documented in deep dysphasic patients, who make semantic errors in repetition (see Martin, Dell, & Schwartz, 1994; Martin & Saffran, 1992; Martin, Saffran, & Dell, 1996).

Hinton and Shallice (1991) reproduced the co-occurrence of visual, semantic, and mixed visual-and-semantic errors in deep dyslexia by damaging a connectionist network that mapped orthography to semantics. During training, the network learned to form *attractors* for 40 word meanings across five categories, such that patterns of semantic features that were similar to a known word meaning were pulled to that exact meaning over the course of settling. When the network was damaged by removing some units or connections, it no longer settled normally; the initial semantic activity caused by an input would occasionally fall within a neighboring attractor basin, giving rise to an error response. These errors were often semantically related to the stimulus because words with similar meanings correspond to nearby attractors in semantic space. The damaged network also produced visual errors due to its inherent bias towards similarity: visually similar words tend to produce similar initial semantic patterns, which can lead to a visual error if the basins are distorted by damage (see Figure 4).

Plaut and Shallice (1993) extended these initial findings in a number of ways. They established the generality of the co-occurrence of error types across a wide range of simulations, showing that it does not depend on specific characteristics of the network architecture, the learning procedure, or the way responses are generated from semantic activity. A particularly relevant simulation in this regard involved an implementation of the full semantic pathway—mapping orthography to phonology

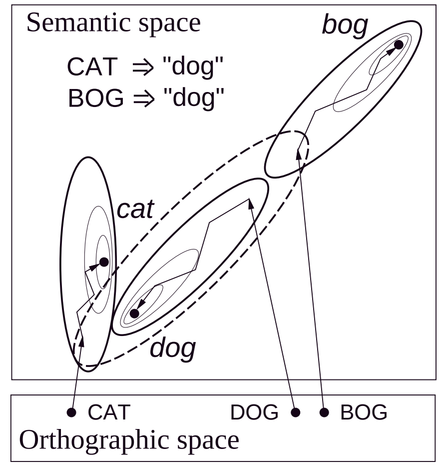


Figure 4: How damage to attractors (dashed oval) can give rise to both semantic and visual errors.

via semantics—using a deterministic Boltzmann Machine (Hinton, 1989; Peterson & Anderson, 1987). Lesions throughout the network gave rise to both visual and semantic errors, with lesions prior to semantics producing a bias towards visual errors and lesions after semantics producing a bias towards semantic errors. Thus, the network replicated both the qualitative similarity and quantitative differences among deep dyslexic patients. The network also exhibited a number of other characteristics of deep dyslexia not considered by Hinton and Shallice (1991), including the occurrence of visual-then-semantic errors, greater confidence in visual as compared with semantic errors, and relatively preserved lexical decision with impaired naming.

Plaut and Shallice carried out further simulations to address the influences of concreteness on the reading performance of deep dyslexic patients. Another full implementation of the semantic pathway, shown in Figure 5, was trained to pronounce a new set of words consisting of both concrete and abstract words. Concrete words were assigned far more semantic features than were abstract words, under the assumption that the semantic representations of concrete words are less dependent on the contexts in which they occur (Saffran, Bogyo, Schwartz, & Marin, 1980; Schwanenflugel, 1991). As a result, the network developed stronger attractors for concrete than abstract words during training, giving rise to better performance in reading concrete words under most types of damage, as observed in deep dyslexia (see Figure 6a). Surprisingly, severe damage to connections implementing the attractors at the semantic level produced the opposite pattern, in which the network read *abstract* words better than concrete words (see Figure 6b). This pattern of performance is reminiscent of CAV, the single, enigmatic

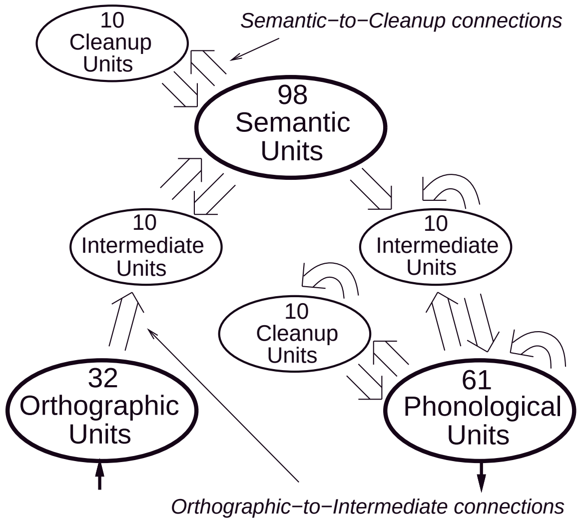


Figure 5: The architecture used by Plaut and Shallice (1993) to model the effects of concreteness in deep dyslexia. The network constitutes a full implementation of the semantic pathway of Figure 1 with the addition of extra "cleanup" units that allow the network to learn stronger semantic and phonological attractors.

patient with *concrete word dyslexia* (Warrington, 1981). The double dissociation between reading concrete versus abstract words in patients is often interpreted as implying that there are separate modules within the cognitive system for concrete and abstract words. The current simulation demonstrates that such a radical interpretation is unnecessary: the double dissociation can arise from damage to different parts of a distributed network, in which parts process both types of items but develop somewhat different functional specializations through learning (see Plaut, 1995a, for further results and discussion).

The Plaut and Shallice (1993) simulations of deep dyslexia provide strong support for characterizing the operation of the semantic pathway, and lexical semantic processing more generally, in terms of a distributed network that learns to form attractors for patterns of semantic features that correspond to word meanings. It should be pointed out, however, that it is possible to model similar phenomena using word-specific representations. For example, Dell (1986; 1988) used a connectionist network with localist units to model semantic and phonological influences in speech production errors, and Martin et al. (1994) replicated aspects of *deep dysphasia* (Howard & Franklin, 1988; Katz & Goodglass, 1990; Martin & Saffran, 1990), including semantic and phonological errors in word repetition, by introducing abnormally rapid decay of lexical activation in the Dell model. The advantage of the distributed approach in the current context is that the prop-

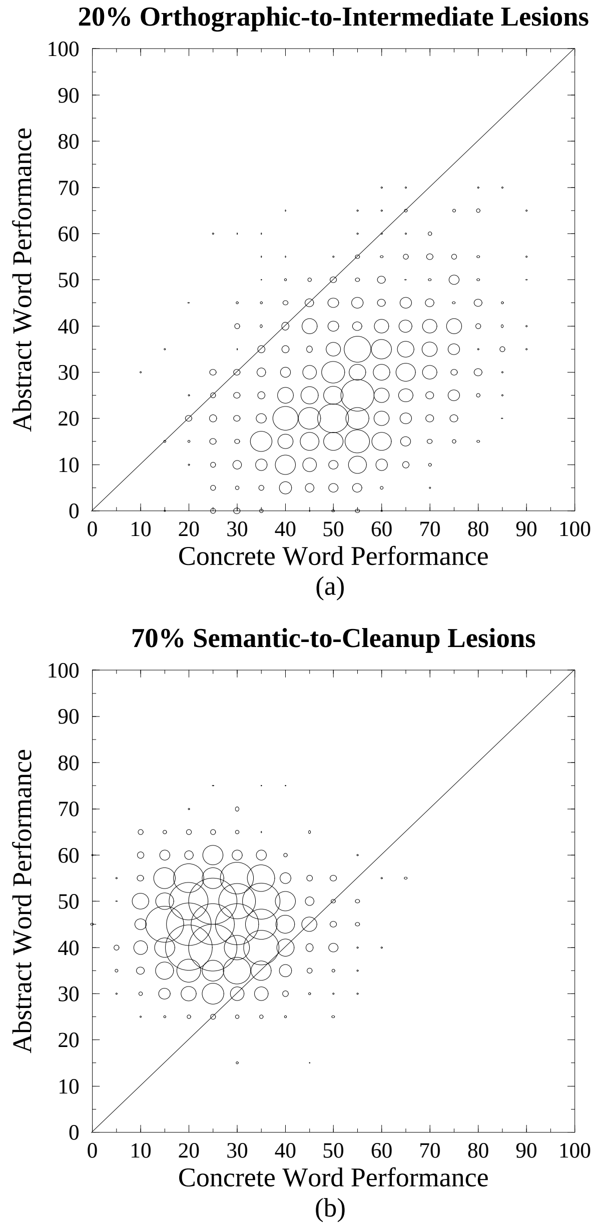


Figure 6: Percent correct performance on concrete versus abstract words of the Plaut and Shallice (1993) simulation after (a) 1000 lesions of 20% of orthographic-to-intermediate connections and (b) 1000 lesions of 70% of semantic-to-cleanup connections, as depicted in Figure 5. The radius of each circle is proportional to the number of lesions yielding the performance levels indicated by the position of the circle. The diagonal lines correspond to equal levels of performance on concrete and abstract words.

erties of normal and impaired semantic processing arise out of the same computational principles that operate in the rest of the lexical system.

## 5 Rehabilitating Reading via Meaning

An important but often neglected motivation for theoretical analyses of normal cognitive processing and its breakdown following brain damage in individual patients is that such analyses may lead to the design of more effective therapy to remediate cognitive impairments (Howard & Hatfield, 1987). Attempts at cognitive rehabilitation of the mapping between orthography and semantics (e.g., Behrmann, 1987; Coltheart & Byng, 1989; Scott & Byng, 1989; Weekes & Coltheart, 1996) have resulted in considerable improvement in performance on treated words and significant generalization to untreated but related words, although there is little understanding of the underlying mechanisms by which this occurs. Furthermore, the degree and breadth of recovery and generalization can vary considerably across patients: Some patients show generalization in some semantic categories but not others (e.g., CH; Behrmann & Lieberthal, 1989); some learn the treated items well but show no generalization to untreated items (e.g., PS; Hillis, 1993); still others have difficulty learning the treated items themselves. As Hillis (1993) points out, what is needed is a theory of rehabilitation that provides a detailed specification of the impaired cognitive system, how it changes in response to treatment, and what factors are relevant to the efficacy of the treatment.

With the goal of contributing to such a theory, Plaut (1996) investigated the degree of recovery and generalization produced when networks that read via meaning are retrained after damage. In one experiment, a replication of the Hinton and Shallice (1991) network was trained until it was fully accurate on all 40 words. It was then subjected to damage either near orthography or within semantics, and retrained on half of the words. Performance was measured both for those treated words and for the untreated words. For comparison, performance of the network when retrained on all 40 words was also measured. Plaut found that retraining produced rapid improvement on treated words and substantial generalization to untreated words only after lesions within semantics; when retraining after lesions near orthography, treated improvement was erratic and there was no generalization to untreated words (see Figures 7a and b). This difference is due to the relative degree of consistency in the mapping performed at different levels of the network. Figure 8 presents a graphical depiction of this effect using vectors (arrows) to represent weight changes. Within semantics, similar words require similar interactions, so that the weight changes caused by retraining on some words will tend also to improve performance on other, related words (i.e., the optimal weight changes for words are mutually consistent). By contrast, similar orthographic patterns typically must generate very different semantic patterns. As a result, when retraining after lesions near orthography, the weight changes for treated items are unrelated to those that would improve the untreated items, and there is no generalization. These finding provide a basis for understanding the mechanisms of recovery and generalization in patients, and may help explain the observed variability in their recovery.

A theory of rehabilitation should provide guidance in selecting items for treatment so as to maximize generalized recovery. In a second experiment, Plaut (1996) used an artificial version of the task of mapping orthography to semantics to investigate whether generalization was greater when retraining on high- versus low-dominance category exemplars. Somewhat surprisingly, although retraining on high-dominance exemplars produced greater recovery on treated items, retraining on low-dominance exemplars produced greater generalization to untreated items. These findings can be understood in terms of the relative adequacy with which the sets of high- versus low-dominance exemplars approximate the range of semantic similarity among all of the words. In the simulation, high-dominance words accurately estimate the central tendency of a category, but provide little information about the ways in which category members can *vary*. By contrast, each low-dominance word indicates many more ways in which members can differ from the prototype and yet still belong to the category. Thus, collectively, the semantic representations of low-dominance words cover more of the features needed by the entire set of words than do the representations of high-dominance words. At the same time, the average affects of retraining on low-dominance words provides a reasonable estimate of the central tendency of the category, yielding generalization to high-dominance words (as found in human category learning by, e.g., Posner & Keele, 1968).

In a final simulation, Plaut (1996) used the *failure* of the network in replicating the error pattern of recovering deep dyslexic patients to constrain the underlying theory of normal and impaired word reading. Plaut measured the changes in the distribution of error types brought about by retraining an orthography-to-semantics network after damage. Rather than semantic errors being the first to drop out, visual and unrelated errors were eliminated earliest. Semantic and mixed visual-and-semantic errors were eliminated only at the very end of retraining. Thus, the changes in the pattern of errors produced by the network in recovery to near normal levels of correct performance failed to reproduce the transition from deep to phonological dyslexia observed in patients (Friedman, 1996; Klein et al., 1994). This discrepancy between the

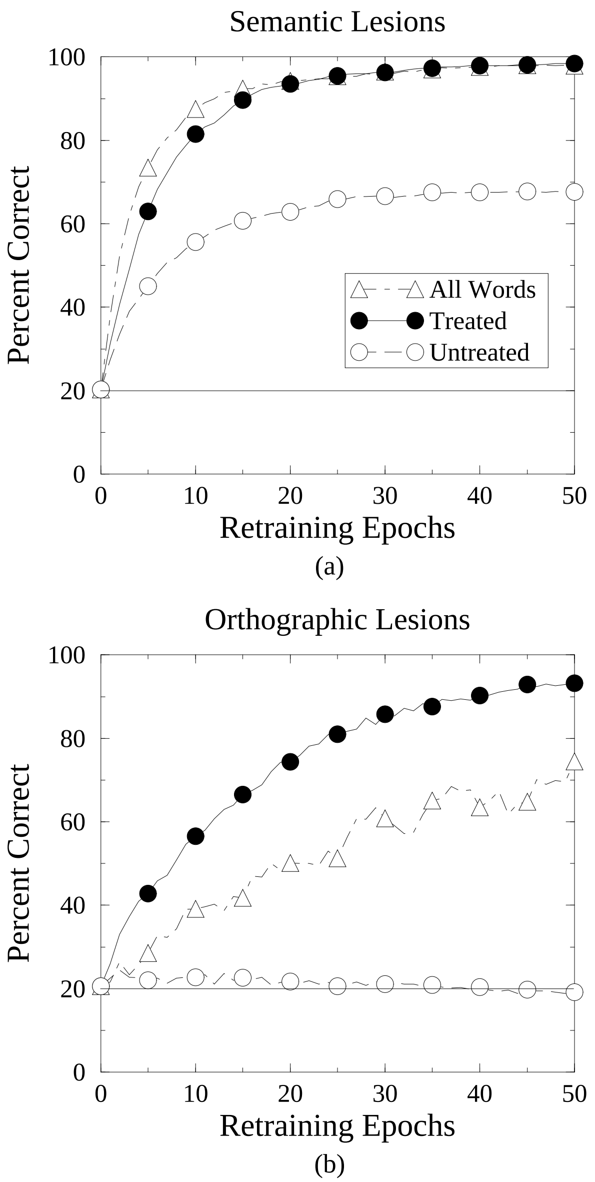


Figure 7: Correct performance in pronouncing treated and untreated items when retraining the Plaut (1996) network that maps orthography to semantics after (a) lesions within semantics (i.e., 50% of cleanup-to-semantics connections), and (b) lesions near orthography (i.e., 30% of orthographic-to-intermediate connections; see Figure 5). Performance when retraining on all 40 words is also shown for each condition.

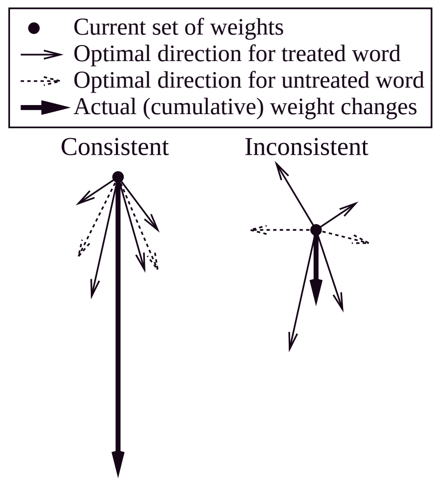


Figure 8: Depiction of the effect of consistent versus inconsistent weight changes on the extent of recovery and generalization in relearning. In each condition, the small solid arrows represent directions of weight change induced by treated words; the large solid arrow is the (vector) sum of these smaller arrows, representing the actual weight changes administered to the network. The length of this vector reflects the speed of relearning the treated words. The dotted arrows represent directions of weight change that would be optimal for untreated words if they were trained—to the extent that these point in the same direction as the actual weight change vector, retraining on the treated words will also improve performance on the untreated words.

behavior of the network and that of patients can be understood if recovery in the patients involves more than relearning in the semantic route alone. In particular, the findings suggest that, within the current approach, the transition from deep to phonological dyslexia must also involve some improvement in the operation of the phonological pathway (or in phonology itself). Such improvement would produce a greater reduction in semantic errors relative to other types of error because even partial correct phonological information about the stimulus would be sufficient to rule out most semantic errors (Newcombe & Marshall, 1980).

It must be kept in mind that the Plaut (1996) findings relate to patient therapy only in the most general way, given that the version of the task of mapping orthography to semantics it performs is much simpler than the actual task performed by patients. Nonetheless, the principles that emerge as central to understanding the nature of relearning and generalization in the networks may provide the foundations for understanding the nature of recovery in patients.

# 6 Conclusion

The traditional way of thinking about the mechanisms subserving word reading (and other lexical tasks) involves stipulating rather complicated and domain-specific structures and processes. Thus, there are representations that apply only to specific words, or to words but not to nonwords, or to concrete words but not abstract words, etc., and there are separate sets of rules or pathways that process words but not nonwords, or only regular words but not exception words, etc.

The current chapter has attempted to articulate and support an alternative view of lexical knowledge and processing: that it develops through the operation of general learning principles as applied to 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 among and between the relevant types of information—orthographic, phonological, and semantic. The structural divisions within the system—which are critical in accounting for specific patterns of acquired dyslexia—arise from the neuroanatomic localization of input and output modalities, not from differences in the content of representations (see Farah, 1994; Farah & McClelland, 1991, for similar arguments).

The simulations described in this chapter illustrate how connectionist computational principles—distributed representation, structure-sensitive learning, and interactivity—can provide insight into central empirical phenomena in normal skilled reading, its breakdown due to brain damage, and its remediation following damage. This is not to say that the 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.

# 7 Acknowledgments

This work was supported by NIH/NIMH Grant MH47566 and the McDonnell-Pew Program in Cognitive Neuroscience Grant T89–01245–016. I'd like to thank Jay McClelland, Karalyn Patterson, Mark Seidenberg, and Tim Shallice for their contributions to the work described in this chapter, and Marlene Behrmann, Derek Besner, Max Coltheart, Joe Devlin, Geoff Hinton, and Eamon Strain for helpful discussions and comments. Correspondence should be addressed to David C. Plaut, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213–3890, plaut@cmu.edu

# References

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

Balota, D., & Ferraro, R. (1993). A dissociation of frequency and regularity effects in pronunciation performance across yong adults, older adults, and individuals with senile dementia of the Alzheimer type. *Journal of Memory and Language*, *32*, 573–592.

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.

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

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 252–259.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory and Cognition*, *8*, 493–512.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 389–401.

Behrmann, M. (1987). The rites of righting writing: Homophone remediation in acquired dysgraphia. *Cognitive Neuropsychology*, *4*, 365–384.

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

Behrmann, M., & Lieberthal, T. (1989). Category-specific treatment of a lexical semantic deficit: A single case study of global aphasia. *British Journal of Communication Disorders*, *24*, 281–299.

Besner, D. (this volume). Basic processes in reading: Multiple routines in localist and connectionist models. In R. M. Klein, & P. A. McMullen (Eds.), *Converging methods for understanding reading and dyslexia*. Cambridge, MA: MIT Press.

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). Hillsdale, NJ: Erlbaum.

Buchanan, L., Hildebrandt, N., & MacKinnon, G. E. (1994a). Implicit phonological processing in deep dyslexia. *Brain and Language*, *47*, 435–437.

Buchanan, L., Hildebrandt, N., & MacKinnon, G. E. (1994b). Phonological processing of nonwords by a deep dyslexic patient: A rowse is implicitly a rose. *Journal of Neurolinguistics*, *8*, 163–181.

Buchanan, L., Hildebrandt, N., & MacKinnon, G. E. (1996). Phonological processing of nonwords in deep dyslexia: Typical and independent? *Journal of Neurolinguistics*, *9*, 113–133.

Buchanan, L., Hildebrandt, N., & MacKinnon, G. E. (this volume). Effects of phonology on deep dyslexia. In R. M. Klein, & P. A. McMullen (Eds.), *Converging methods for understanding reading and dyslexia*. Cambridge, MA: MIT Press.

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: Erlbaum.

Coltheart, M., & Byng, S. (1989). A treatment for surface dyslexia. In X. Seron, & G. Deloche (Eds.), *Cognitive approaches in neuropsychological rehabilitation* (pp. 159–174). Hillsdale, NJ: Erlbaum.

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.

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: Erlbaum.

Coltheart, M., Patterson, K., & Marshall, J. C. (Eds.). (1980). *Deep dyslexia*. London: Routledge & Kegan Paul.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124–142.

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.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, *6*, 205–254.

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.

Friedman, R. B. (1996). Recovery from deep alexia to phonological alexia. *Brain and Language*, *52*, 114–128.

Glosser, G., & Friedman, R. B. (1990). The continuum of deep/phonological alexia. *Cortex*, *26*, 343–359.

Glosser, G., & Friedman, R. B. (1991). Lexical but not semantic priming in Alzheimer's disease. *Psychology and Aging*, *6*, 522–527.

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.

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

Harm, M. W. (1998). *Division of labor in a computational model of visual word recognition*. PhD thesis, Department of Computer Science, University of Southern California, Los Angeles, CA.

Hillis, A. E. (1993). The role of models of language processing in rehabilitation of language impairments. *Aphasiology*, *7*, 5–26.

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

Hinton, G. E. (1989). Deterministic Boltzmann learning performs steepest descent in weight-space. *Neural Computation*, *1*, 143–150.

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

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceed-*

*ings of the National Academy of Science, USA*, *79*, 2554–2558.

Howard, D., & Franklin, S. (1988). *Missing the meaning?* Cambridge, MA: MIT Press.

Howard, D., & Hatfield, F. M. (1987). *Aphasia therapy*. Hillsdale, NJ: Erlbaum.

Joordens, S., & Besner, D. (1992). Priming effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 483–491.

Katz, R. B., & Goodglass, H. (1990). Deep dysphasia: Analysis of a rare form of repetition disorder. *Brain and Language*, *39*, 153–185.

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.

Klein, D., Behrmann, M., & Doctor, E. (1994). The evolution of deep dyslexia: Evidence for the spontaneous recovery of the semantic reading route. *Cognitive Neuropsychology*, *11*, 579–611.

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

Lorch, R. F., Balota, D., & Stamm, E. (1986). Locus of inhibition effects in the priming of lexical decisions: Pre- or post-lexical access? *Memory and Cognition*, *14*, 95–103.

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

Martin, N., Dell, G. S., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, *47*, 609–660.

Martin, N., & Saffran, E. M. (1990). Repetition and verbal STM in transcortical sensory aphasia: A case study. *Brain and Language*, *39*, 254–288.

Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language*, *43*, 240–274.

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal-STM capacity and lexical errors in repetition. *Brain and Language*, *52*, 83–113.

Masson, M. E. J. (1991). A distributed memory model of context effects in word identfication. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 233–263). Hillsdale, NJ: Erlbaum.

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. (this volume). Interactive processes in word identification: A computational approach. In R. M. Klein, & P. A. McMullen (Eds.), *Converging methods for understanding reading and dyslexia*. Cambridge, MA: MIT Press.

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. (1987). Reading pseudohomophones: Implications for models of pronunciation and the locus of the word-frequency effects in word naming. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 14–24.

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.

McNamara, T. P. (1994). Theories of priming II: Types of primes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 507–520.

McRae, K., de Sa, V., & Seidenberg, M. S. (1993). Modeling property interactions in accessing conceptual memory. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 729–734). Hillsdale, NJ: Erlbaum.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

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.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt, & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press.

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 & Kegan Paul.

Moss, H. E., Hare, M. L., Day, P., & Tyler, L. K. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science*, *6*, 413–427.

Moss, H. E., & Marslen-Wilson, W. D. (1993). Access to word meanings during spoken language comprehension: Effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1254–1276.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 863–883.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Newcombe, F., & Marshall, J. C. (1980). Transcoding and lexical stabilization in deep dyslexia. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 176–188). London: Routledge & Kegan Paul.

Norris, D. (1994). A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1212–1232.

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., Coltheart, M., & Marshall, J. C. (Eds.). (1985). *Surface dyslexia*. Hillsdale, NJ: Erlbaum.

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

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

Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation*, *1*, 263–269.

Peterson, C., & Anderson, J. R. (1987). A mean field theory learning algorithm for neural nets. *Complex Systems*, *1*, 995–1019.

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: Erlbaum.

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

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of naming and lexical decision. *Language and Cognitive Processes*, *12*, 767–808.

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: Erlbaum.

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.

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

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.

Postman, L., & Keppel, G. (1970). *Norms of word associations*. New York: Academic Press.

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.

Reggia, J. A., Marsland, P. M., & Berndt, R. S. (1988). Competitive dynamics in a dual-route connectionist model of print-to-sound transformation. *Complex Systems*, *2*, 509–547.

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.

Saffran, E. M., Bogyo, L. C., Schwartz, M. F., & Marin, O. S. M. (1980). Does deep dyslexia reflect right-hemisphere reading? In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 381–406). London: Routledge & Kegan Paul.

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings*. Hillsdale, NJ: Erlbaum.

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.

Scott, C., & Byng, S. (1989). Computer assisted remediation of a homophone comprehension disorder in surface dyslexia. *Aphasiology*, *3*, 301–320.

Seidenberg, M. S. (1992). Beyond orthographic depth: Equitable division of labor. 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., 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.

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

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984a). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behaviour*, *23*, 383–404.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984b). Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition*, *12*, 315–328.

Shallice, T., & McCarthy, R. (1985). Phonological reading: From patterns of impairment to possible procedures. In K. Patterson, M. Coltheart, & J. C. Marshall (Eds.), *Surface dyslexia* (pp. 361–398). Hillsdale, NJ: Erlbaum.

Shallice, T., & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 119–145). London: Routledge & Kegan Paul.

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

Sharkey, A. J., & Sharkey, N. E. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language*, *31*, 543–572.

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1191–1210.

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.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.

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.

Warrington, E. K. (1981). Concrete word dyslexia. *British Journal of Psychology*, *72*, 175–196.

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

Weekes, B., & Coltheart, M. (1996). Surface dyslexia and surface dysgraphia: Treatment studies and their theoretical implications. *Cognitive Neuropsychology*, *13*, 277–315.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist “dual-process” model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.