In J. Reggia, R. Berndt, & E. Ruppin (Eds.), *Neural modeling of cognitive and brain disorders* (pp. 157–176). New York: World Scientific, 1996.

CONNECTIONIST MODELING OF THE BREAKDOWN AND RECOVERY OF READING VIA MEANING

DAVID C. PLAUT
Department of Psychology, Carnegie Mellon University
and the Center for the Neural Basis of Cognition
Pittsburgh, PA 15213-3890, plaut@cmu.edu

At least two processing routes in the brain are involved in pronouncing written words: a *semantic* route that derives the pronunciation via meaning, and a *phonological* route that derives it via spelling-sound correspondences. Simulations involving partial damage to an isolated semantic route (Plaut & Shallice, 1993) provide a comprehensive account of the rather peculiar combination of symptoms exhibited by patients with *deep dyslexia*, including the occurrence of semantic errors (e.g., reading RIVER as "ocean"), their co-occurrence with visual errors, and influences of imageability or concreteness on correct and error performance. Furthermore, when a version of the model is retrained after damage (Plaut, 1996), the degree and variability of its recovery and generalization are qualitatively similar to the results of some cognitive rehabilitation studies. The results challenge traditional assumptions about the nature of the mechanisms subserving word reading, and illustrate the value of explicit computational simulations of normal and impaired cognitive processes. They also suggest that connectionist modeling can provide a framework for generating specific hypotheses about strategies for rehabilitation.

Cognitive neuropsychology attempts to relate the patterns of impaired and preserved abilities of brain-injured patients to models of normal cognitive functioning, with the goals of explaining the behavior of the patients in terms of the effects of damage in the model, and of informing the model based on the observed behavior of patients (Coltheart, 1985; Ellis & Young, 1988; Shallice, 1988). A major motivation for many researchers is that a more detailed analysis of the normal mechanism, and the way it is impaired in particular patients, should lead to the design of more effective therapy to remediate these impairments (Howard & Hatfield, 1987; Riddoch & Humphreys, 1994; Seron & Deloche, 1989). Moreover, the patterns of recovery exhibited by patients place additional constraints on models of normal and impaired cognitive processing. The purpose of this chapter is to illustrate in a particular domain—reading via meaning—how computational principles from connectionist or parallel distributed processing (PDP) research can provide insight into the nature of normal cognitive processes, how they can break down following brain damage, how they can recover, and how to design therapy to maximize this recovery.

Perhaps the most detailed attempts at relating the behavior of damaged connectionist networks to that of brain-injured patients has been in the domain of acquired reading disorders (see, e.g., Hinton & Shallice, 1991; Mozer & Behrmann, 1990; Patterson, Seidenberg, & McClelland, 1989; Plaut, McClelland, Seidenberg, & Pat-

terson, 1996; Plaut & Shallice, 1993). This is in part because investigations of reading in both cognitive psychology and neuropsychology (Coltheart, 1987) have produced a rich and often counterintuitive set of empirical findings.

Prior to the late 1960's, the major distinction among acquired dyslexic patients was simply whether the reading deficit was accompanied by a deficit in writing—alexia with agraphia—or whether it occurred in isolation—alexia without agraphia, or *pure* alexia (Dejerine, 1892). Little attempt was made to distinguish among different types of reading deficits until Marshall and Newcombe (1966; 1973) identified a number of separate types of acquired dyslexia based on the typical patterns of errors that patients made in reading aloud. In particular, *surface* dyslexia involved phonological confusions in the procedure by which words are sounded-out based on typical spelling-sound correspondences (e.g., SEW →"sue"), whereas *deep* dyslexia involved semantic confusions, in which words were often misread as semantically related words (e.g., DINNER →"food").

Marshall and Newcombe (1973) explained the existence of these distinct types of dyslexia in terms of damage to a "dual-route" model of normal reading (also see Coltheart, 1978; 1985; Coltheart, Curtis, Atkins, & Haller, 1993; Meyer, Schvaneveldt, & Ruddy, 1974; Morton & Patterson, 1980; Paap & Noel, 1991). In Marshall and Newcombe's model, written words can be pronounced through either of two pathways. The first is a *phonological* pathway that translates from spelling to sound using grapheme-phoneme correspondence (GPC) rules. This pathway enables people to read word-like nonsense letter strings (e.g., MAVE) as well as so-called *regular* words that obey standard spelling-sound correspondences (e.g., GAVE). The second way of pronouncing words is via a *semantic* pathway in which a word is first recognized and assigned a meaning which is then used to access its pronunciation. The semantic pathway enables people to read so-called *exception* words that violate the standard GPC rules (e.g. HAVE). According to Marshall and Newcombe, surface dyslexic patients have a selective impairment of the semantic pathway, such that their errors reflect the isolated operation of the phonological pathway. Conversely, deep dyslexia reflects the isolated—and, according to more recent theories (see Shallice, 1988), partially impaired—operation of the semantic pathway following severe damage to the phonological pathway.

Considerable further research has examined the characteristics of both surface and deep dyslexia in more detail. The chapter by Patterson, Plaut, McClelland, Seidenberg, Behrmann, and Hodges (this volume) articulates and provides empirical support for an account of surface dyslexia based on connectionist implementations of the phonological pathway (Plaut et al., 1996). The current chapter focuses on deep dyslexia and, more generally, on the breakdown and recovery of the operation of the semantic pathway.

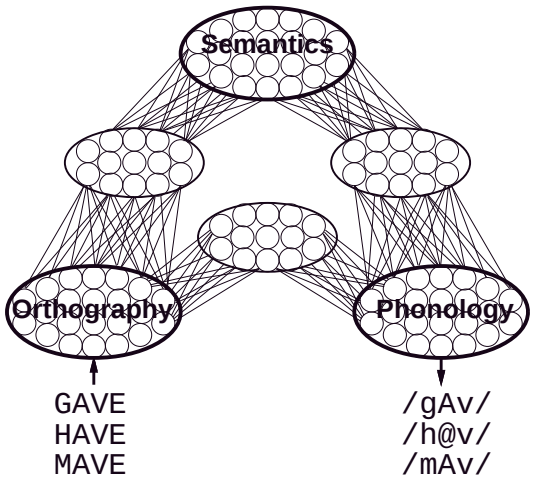


Figure 1: A connectionist framework for lexical processing. Adapted from Seidenberg and McClelland (1989).

1 A Connectionist Framework for Lexical Processing

Seidenberg and McClelland (1989) presented a connectionist framework for lexical processing that has some similarities with—but also some important differences from—standard dual-route theory. Within the framework, depicted in Figure 1, orthographic, phonological, and semantic information is represented as distributed patterns of activity over groups of simple neuron-like processing units. Within each domain, similar words are represented by similar patterns of activity. Transformations among domains are accomplished via cooperative and competitive interactions among units, including intermediate or *hidden* units that mediate between the orthography, phonology, and semantics. 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 weights on connections between them which collectively encode the system's knowledge and are learned through exposure to written words, spoken words, and their meanings.

Although this framework 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 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 Van Orden, Pennington, & Stone, 1990). Second, although the system is composed of a phonological and a semantic pathway, these pathways operate according to very different principles and have very different functional properties than the analogous pathways in traditional dual-route models (e.g. Coltheart et al., 1993). In particular, the phonological pathway does not apply GPC rules that succeed only for regular words and nonwords. Rather, it learns to map orthography to phonology for all types of stimuli (including exception words) based on a sensitivity to spelling-sound *consistency* (Glushko, 1979). The degree of mastery achieved by the phonological route for items it finds most difficult—low-frequency exception words—will generally not be perfect and will depend on a number of factors, including the strength of contribution from the semantic during learning (see Patterson et al., this volume, and Plaut et al., 1996).

2 Impaired Reading Via Meaning in Deep Dyslexia

As suggested earlier, patients with deep dyslexia (see Coltheart, Patterson, & Marshall, 1980) seem to have a severe impairment of the phonological pathway. This is indicated, in part, by the fact that they are virtually unable to read pronounceable nonwords (e.g., MAVE). They also have impairments in reading words that suggest additional partial damage to the semantic pathway. In particular, deep dyslexics make *semantic* errors in oral reading (e.g., reading CAT as "dog"), along with pure *visual* errors (e.g., CAT $\rightarrow$"cot"), mixed *visual-and-semantic* errors (e.g., CAT $\rightarrow$"rat"), and even mediated *visual-then-semantic* errors (e.g., SYMPATHY $\rightarrow$"orchestra", presumably via *symphony*). 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). Performance on additional tests, such as auditory comprehension and picture-word matching, suggests that the secondary damage to the semantic pathway may occur before, within, or after semantics (Shallice & Warrington, 1980).

Hinton and Shallice (1991) reproduced the co-occurrence of semantic and visual errors in deep dyslexia by damaging a network that mapped orthography to semantics (see Figure 2a). During training, the network learned to form attractors for word meanings (using a separate set of semantic "clean-up" units), 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. Following damage, the semantic activity caused by an input would occasionally fall within the attractor basin of a neighboring (related) word, giving rise to a semantic error. Visual errors also occurred due to the network's 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 2b).

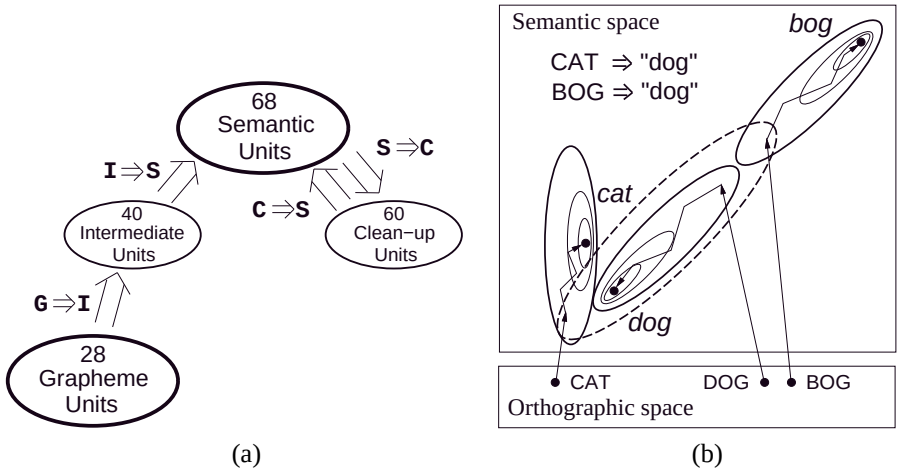
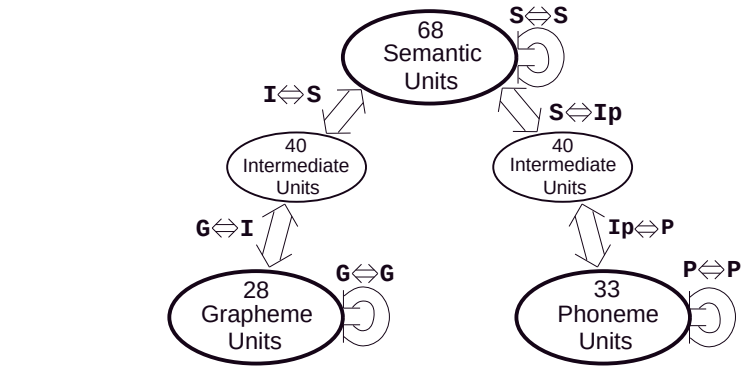


Figure 2: (a) A depiction of the network used by Hinton and Shallice (1991) to model deep dyslexia, corresponding to a portion of the semantic pathway of the Seidenberg and McClelland (1989) framework in Figure 1. Sets of connections are labeled by the initials of the connected groups of units (e.g., G⇒I for connections from the Grapheme units to the Intermediate units). (b) How damage to attractors (dashed oval) can cause both semantic and visual errors. The current patterns of activation in orthography and semantics are each represented as a point in a multi-dimensional *state space* in which the activation of each unit is coded along a separate dimension (for convenience, only two dimensions are depicted). Adapted from Hinton and Shallice (1991) and Plaut and Shallice (1993).

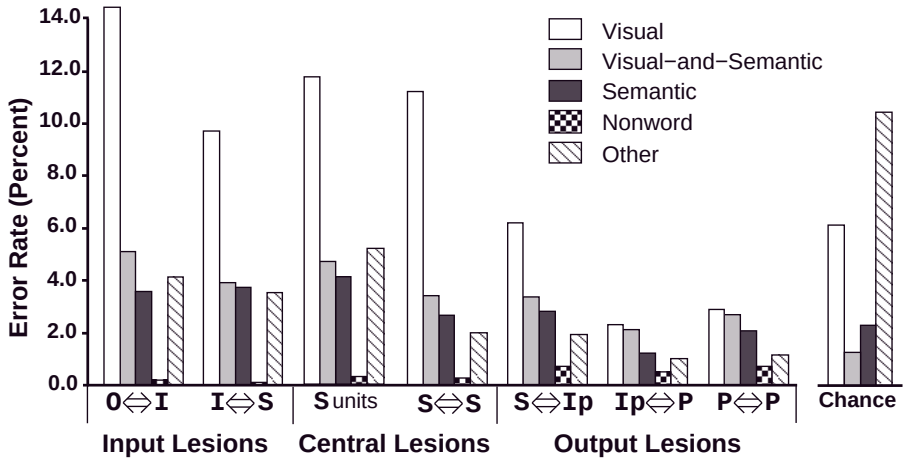
Plaut and Shallice (1993) extended these initial findings in a number of ways. They established the generality of the co-occurrence of error types across a wide range of simulations, showing that it does not depend on specific characteristics of the network architecture, the learning procedure, or the 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 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 (relative to the "chance" distribution; see Figure 3). 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 (1993, also see Plaut, 1995) carried out further simulations to address the influences of concreteness on the reading performance of deep dyslexic patients. As previously mentioned, deep dyslexic patients perform better at reading concrete, high-imageable words compared with abstract, low-imageable words. Strangely, the effects of concreteness—a semantic variable—interact with visual similarity in errors, such that abstract words are more likely than concrete words to produce visual errors, and the resulting responses tend to be more concrete than the stimulus (e.g., SCANDAL →"sandals" Barry & Richardson, 1988). These effects could not be addressed using the original Hinton and Shallice word set because it contains only concrete nouns.

Accordingly, Plaut and Shallice (1993) designed a version of the task of reading via meaning that would allow the effects of concreteness and visual similarity to be investigated directly. Twenty pairs of four-letter words were chosen such that one member of the pair was concrete, the other was abstract, and the two differed by only a single letter (e.g., ROPE and ROLE). The critical difference between the concrete and abstract words related to their semantic representations. Plaut and Shallice's approach to capturing this distinction was based in part on Jones' (1985) demonstration that words vary greatly in the ease with which predicates about them can be generated. For example, more predicates can be generated for basic-level words than for subordinate or superordinate words (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Jones showed that there is a very high correlation (0.88) between ease-of-predication ratings and imageability (which also correlates highly with concreteness), and that the relative difficulty of parts-of-speech in deep dyslexia maps perfectly onto their ordered mean ease-of-predication scores. He argued that the effects of both image-



(a)



(b)

Figure 3: (a) The architecture of the deterministic Boltzmann Machine (DBM) implemented by Plaut and Shallice (1993). (b) Error rates produced by lesions to each main set of connections in the DBM network shown in (a). "Chance" is the distribution of error types if responses were chosen randomly from the word set. Its absolute height is set arbitrarily—only the relative rates are informative. Results are averaged over lesion densities which produced an overall correct response rate between approximately 20% and 80%. Adapted from Plaut and Shallice (1993).

ability and part-of-speech in deep dyslexia can be accounted for by assuming that the semantic pathway is sensitive to ease-of-predication. Plaut and Shallice instantiated this distinction by assigning concrete words more semantic features (predicates) than abstract words: an average of 18.2 versus 4.7 out of 98 possible features, respectively. The first 67 of the semantic features were based on those used by Hinton and Shallice (1991) and applied only to the concrete words; The remaining 31 features (e.g., *has-duration, relates-location, quality-difficulty*) applied primarily to abstract words but occasionally to concrete words as well. The ordering of the features and, in particular, the separation of concrete and abstract features, were irrelevant to the simulation.

Note that it would be misleading to interpret the assignment of more features to concrete words as a literal claim about semantic representations, given that abstract words can certainly make rich and substantial contributions to meaning. Rather, a more appropriate interpretation of the manipulation relates to the degree of *variability* across contexts in the semantics generated by different types of words. As Saffran, Bogyo, Schwartz, and Marin (1980, p. 400; see also Schwanenflugel, 1991) have pointed out,

> A concrete word—a reference term like "rose"—has a core meaning little altered by context (a rose *is* a rose) .... The meanings of abstract words, on the other hand, tend to be more dependent on the contexts in which they are embedded.

A similar contrast appears to hold among different parts-of-speech—for example, between nouns and verbs (Gentner, 1981). Thus, the use of fixed semantic representations containing fewer features for abstract words should really be considered an approximation of a more realistic simulation in which abstract words have fewer semantic features that are activated consistently across a variety of contexts. In fact, if a connectionist network were trained to generate pronunciations from such variable semantic representations, it would come to rely on just those few features that are consistently predictive of the correct response (see McClelland & Rumelhart, 1985, for illustrations of this property). The Plaut and Shallice semantic representations can be thought of as containing only these predictive features.

Plaut and Shallice (1993) trained a network with back-propagation through time (Rumelhart, Hinton, & Williams, 1986) to map orthography to phonology via these semantic representations. To enable the network to learn semantic attractors, the architecture included semantic "clean-up" units much like those in the Hinton and Shallice (1991) network (shown in Figure 2a). Because abstract words have far fewer semantic features, they are less able to engage this semantic clean-up mechanism effectively to form strong attractors during training. These words must therefore rely more heavily on the direct mapping from orthography to semantics, where visual influences are strongest. As a result, lesions to this pathway reproduce the effects of

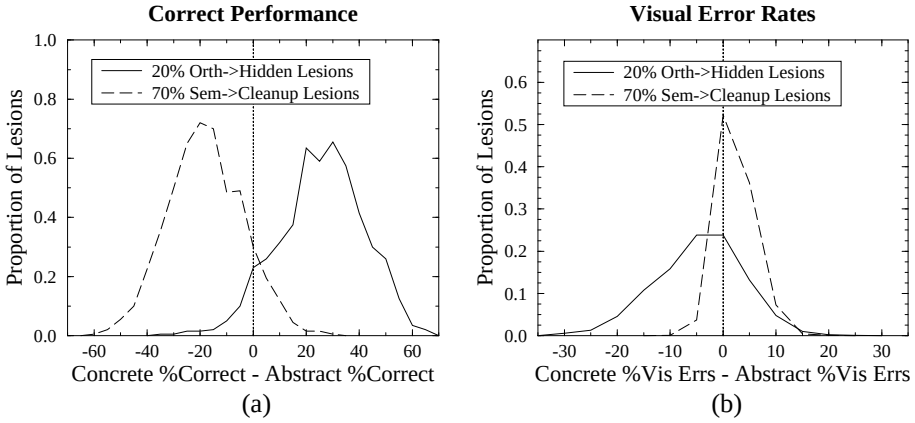


Figure 4: Differences between concrete and abstract words in (a) correct performance, and (b) visual error rates, after 1000 moderate lesions to the "direct" pathway from orthography to semantics (solid lines) in the Plaut and Shallice (1993) concrete/abstract network, and after 1000 severe lesions to the "clean-up" pathway that implemented the semantic attractors (dashed lines). Plotted values are differences (concrete−abstract); positive values (towards the right) reflect greater correct performance or visual error rates for concrete as compared with abstract words. The solid lines correspond to the pattern of deep dyslexia; the dashed lines correspond to the pattern of concrete word dyslexia. Adapted from Plaut (1995).

concreteness and their interaction with visual errors found in deep dyslexia: better correct performance for concrete over abstract words, a tendency for error responses to be more concrete than stimuli, and a higher proportion of visual errors in response to abstract compared with concrete words (see the solid lines in Figure 4).

Surprisingly, severe lesions to connections within the clean-up mechanism that implemented the semantic attractors produced the *opposite* effect: abstract words were now read better than concrete words, and concrete words produced more visual errors than did the abstract words (see the dashed lines in Figure 4). This reversal arises because, under this type of lesion, the processing of most concrete words is impaired but many abstract words can be read solely by the direct pathway. In fact, there is a single known exception to the advantage for concrete words shown by deep dyslexic patients: patient CAV with *concrete word dyslexia* (Warrington, 1981). CAV failed to read concrete words like MILK and TREE but succeeded at highly abstract words such as APPLAUSE, EVIDENCE, and INFERIOR. Overall, abstract words were more likely to be correctly read than concrete (55% vs. 36%). In complementary fashion, 63% of his visual error responses were more abstract than the stimulus. Furthermore, the hypothesis of severe damage to semantic attractors is consistent with other aspects of his performance: CAV's reading disorder was quite severe initially, and he also showed an advantage for abstract words in picture-word matching with

auditory presentation, suggesting severe modality-independent damage at the level of the semantic system.

The double dissociation between reading concrete versus abstract words in patients was interpreted by Warrington and others (e.g., Morton & Patterson, 1980) as implying that concrete and abstract semantics are represented separately in the brain. The Plaut and Shallice (1993) 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, 1995, for further results and discussion).

Overall, the Plaut and Shallice (1993) simulations of deep dyslexia (and of the single, enigmatic case of concrete word dyslexia) provide strong support for characterizing the operation of the semantic pathway, and lexical semantic processing more generally, in terms of a distributed network like that in Figure 1, which learns to form attractors for word meanings. It should be pointed out, however, that it is possible to model analogous phenomena by using localist word units to implement the semantic attractors (see Martin, Dell, & Schwartz, 1994; Martin, Saffran, & Dell, 1996, and the chapter by Dell, Schwartz, Martin, Saffran, & Gagnon, this volume). The advantage of the fully distributed approach in the current context is that the properties of normal and impaired semantic processing arise out of the same computational principles that operate in the rest of the lexical system.

3 Rehabilitating Reading Via Meaning

A computationally explicit theory of normal and impaired cognitive processing should aid in attempts to remediate the impairments (Howard & Hatfield, 1987). In a complementary fashion, accounting for the patterns of recovery exhibited by brain-damaged patients undergoing specific treatment can provide a stringent test of cognitive theories. However, relatively few remediation studies have been based directly on cognitive analyses, and while these have been relatively successful, the specific contribution of the cognitive model—typically a box-and-arrow diagram—is often unclear (see Riddoch & Humphreys, 1994; Margolin, 1992; Seron & Deloche, 1989).

Coltheart and Byng (1989) undertook a series of remediation studies with a surface dyslexic patient, EE, with left temporal-parietal damage due to a fall. On the basis of a number of preliminary tests, they determined that EE had a specific deficit in deriving semantics from orthography. In one study, they gave EE 485 high-frequency words for oral reading and the 54 words he misread were divided in half randomly into treated and untreated sets. For words in the treated set, EE studied cards of the written words augmented with mnemonics for their meanings. As a result, his reading performance on the treated words improved from 44% to 100% correct. Surprisingly,

the untreated words also improved, from 44% to 85% correct; that is, the improvement on untreated words was 73% as much as on treated words. This generalized improvement was specific to the intervention because EE's performance on the words was stable both before and after therapy. Two other studies with EE produced broadly similar results. Overall, Coltheart and Byng found excellent recovery of treated items and substantial generalization to untreated items (also see Weekes & Coltheart, in press).

Unfortunately, such promising results are not always found in rehabilitation studies, even those with very similar types of patients. Scott and Byng (1989) treated a surface dyslexic patient for homophone confusions in reading (e.g., TAIL/TALE) and produced improvement on treated items and, to a lesser extent, untreated items, but found no generalization to his writing of the same items (also see Behrmann, 1987). Behrmann and Lieberthal (1989) trained a globally aphasic patient with semantic impairments on a semantic category sorting task. They found improvement on untreated items only within some categories and minimal generalization to items in untreated categories. Finally, Hillis (1993) carried out an extensive rehabilitation program with a patient who had both orthographic and semantic impairments. The patient was able to learn trained tasks (e.g., lexical decision, naming) but showed virtually no generalization to untrained tasks.

Why some patients improve while others do not is not entirely clear. Furthermore, even in those patients who do improve and show generalization, the cause of this generalization—in terms of changes to the underlying cognitive mechanism induced by treatment—is unknown. An explanation of these findings should account not only for the occurrence of generalization in some patients and conditions, but also for its absence in others. 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.

Early connectionist research (Hinton & Plaut, 1987; Hinton & Sejnowski, 1986) demonstrated that simple networks trained on unstructured tasks can, when retrained after damage, exhibit rapid recovery on treated items and generalization to untreated items. Plaut (1996) extended these findings to apply directly to understanding the basis and variability of recovery in patients, and to provide a platform for testing hypotheses on how to select items for treatment to maximize generalized recovery.

In one simulation, a replication of the Hinton and Shallice (1991) network shown in Figure 2a was subjected to damage either near orthography (to the grapheme-to-intermediate connections) or within semantics (to the semantics-to-cleanup connections) and retrained on half of the words. This retraining produced rapid improvement on treated words and substantial generalization to untreated words only after lesions within semantics; when retraining after lesions near orthography, improvement on

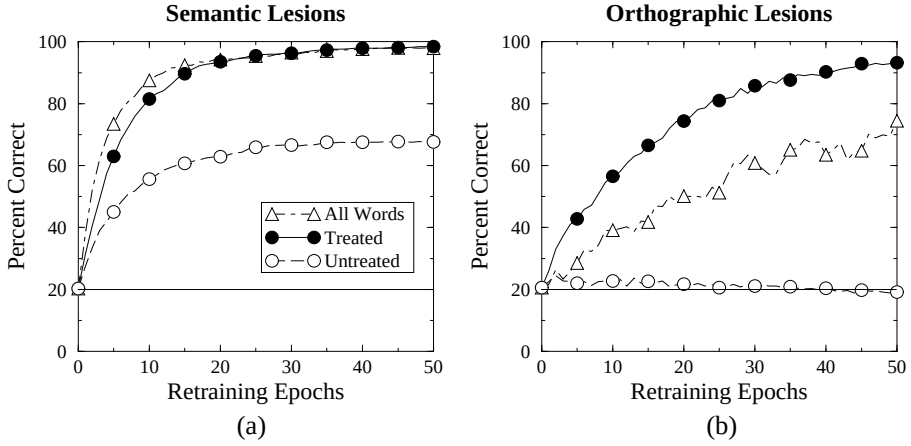


Figure 5: Improvement on treated and untreated items when retraining a network that maps orthography to semantics after (a) lesions within semantics (to the semantics-to-cleanup connections shown in Figure 2a), and after (b) lesions near orthography (to the grapheme-to-intermediate connections). Adapted from Plaut (1996).

treated words was erratic and there was no generalization to untreated words (see Figures 5a and b). This difference was due to the relative degree of *consistency* in the mapping performed at different levels of the network. 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.

In a second simulation, Plaut (1996) used an artificial version of the task of mapping orthography to semantics to investigate whether generalization was greater when retraining on typical versus atypical category exemplars (e.g., ROBIN vs. GOOSE). Somewhat surprisingly, although retraining on typical exemplars produced greater recovery on treated items, retraining on atypical exemplars produced greater generalization to untreated items (see Figure 6). These findings make sense given the adequacy with which sets of typical versus atypical exemplars approximate the range of semantic similarity among all of the words. Semantically typical words accurately estimate the central tendency of a category, but provide little information about the ways in

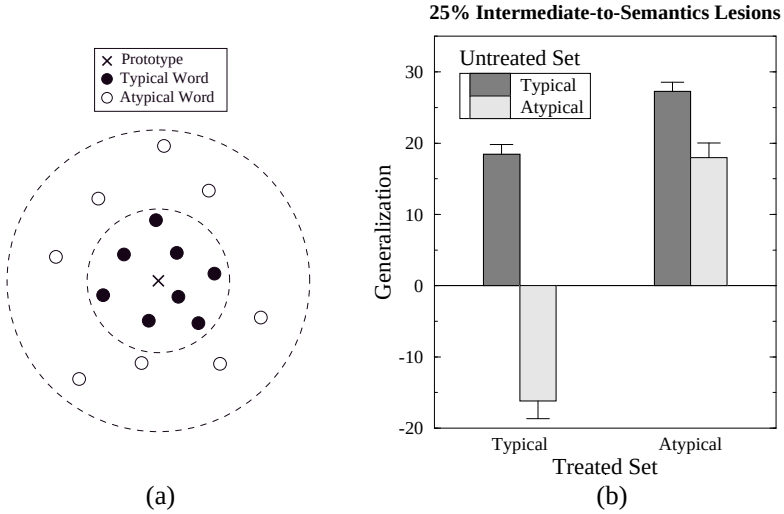


Figure 6: (a) A depiction of the relationship in semantic space between the prototype of a category and typical versus atypical exemplars in that category. (b) Generalization from retraining after lesions of 25% of the intermediate-to-semantics connections, as a function of the semantic typicality of the treated and untreated sets. Adapted from Plaut (1996).

which category members can *vary*. By contrast, each atypical 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 atypical words cover more of the features needed by the entire set of words than do the representations of more typical words. At the same time, the average effects of retraining on atypical words provides a reasonable estimate of the central tendency of the category, yielding generalization to typical words (as found in human category learning by, e.g., Posner & Keele, 1968). In this way, the simulation generated a novel prediction about how to select items for treatment so as to maximize generalized recovery.

In a final simulation, Plaut (1996) used the *failure* of the network to replicate the error pattern of recovering deep dyslexic patients to constrain the underlying theory: that the improvement of these patients must be due, at least in part, to some recovery of function in the (unimplemented) phonological pathway. The relevant empirical finding is that deep dyslexia can resolve into 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 make semantic errors, they can be quite similar to deep dyslexic patients in other respects. In fact, Glosser and Friedman (1990, also see Newcombe & Marshall, 1980) 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, & Dell, 1996, and Dell et al., this volume).

Plaut (1996) 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. This discrepancy between the 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).

One clear indicator of the operation of the phonological route is the ability to read pronounceable nonwords, as such items cannot be read via semantics. Thus, the above explanation is supported by the observation that, for many deep/phonological patients, as their rates of semantic errors dropped to near zero, their nonword reading performance improved. For example, on initial testing, patient GR (Glosser & Friedman, 1990) made 11% semantic errors and read correctly 5% (1/20) of nonwords. Seven months later, he made no purely semantic errors (although 3% were visual-and-semantic); concurrently, his nonword reading had improved to 44% (22/50). Similar results have been found with a number of other patients (e.g., DV, Glosser & Friedman, 1990; EG, Laine, Niemi, & Marttila, 1990; but see Plaut, 1996, for discussion of a possible exception: RL, Klein et al., 1994). Thus, while the behavior of the network on its own fails to account for the resolution of deep to phonological dyslexia, its performance is consistent with a more general account in which the phonological route also contributes to the nature of the recovery in these patients.

In summary, the Plaut (1996) simulations demonstrate that the investigation of relearning after damage in connectionist networks can provide insight into the basis and variability of recovery of the impairments of brain-damaged patients, can generate interesting hypotheses on how to design therapy to remediate these impairments, and can contribute valuable theoretical constraints on our understanding of normal and impaired cognitive processing.

4 Conclusions

The current work adopts a perspective on lexical processing in which distributed representations of orthographic, phonological, and semantic information interact and mutually constrain each other in the process of settling on the best interpretation and response for a given input. The success of connectionist implementations of the phonological pathway—from orthography to phonology—in modeling normal and impaired word reading (Plaut et al., 1996; Seidenberg & McClelland, 1989) stems in large part from the fact that connectionist networks are biased to give similar responses to similar inputs. For this very reason, though, mappings within the semantic pathway—from orthography to semantics to phonology—pose a particular challenge to such networks as there is no systematic relationship between the surface forms of (monomorphemic) words and their meanings.

Critically, even though very different problems are being solved within the phonological and semantic pathways, the same computational principles are effective for both. In particular, learning and processing in distributed connectionist networks are sensitive to the *frequency* with which particular items are presented for training, their *similarity* to other items within each domain, and the *consistency* of their mappings between domains with those of other items. Nonetheless, in meeting the different demands of the two pathways, the principles give rise to rather different functional properties. When applied within the phonological pathway, connectionist networks embodying these principles give rise to the empirical pattern of interaction between word frequency and spelling-sound consistency observed in the naming latencies of skilled readers and in the naming accuracy of surface dyslexic readers (see Patterson et al., this volume). When applied within the semantic pathway, networks learn to form *attractors* for familiar word meanings; under damage, these attractors give rise to semantic errors and their co-occurrence with visual errors, and effects of imageability/concreteness, as exhibited by deep dyslexic patients. Moreover, the degree of generalized recovery following retraining of the damaged semantic pathway depends on the degree of consistency of the optimal weight changes for treated and untreated items—which depend on the specific location of damage within the network and the items selected for treatment.

Taken together, the replication of the diverse set of empirical findings in both

normal and impaired reading by networks embodying a common set of computational principles provides strong evidence that the same principles apply within the normal and impaired human reading system.

Acknowledgments

I would like to acknowledge Tim Shallice for his contributions to the research reported in this chapter. The research was supported by grants from the McDonnell-Pew Program in Cognitive Neuroscience (T89–01245–016), the National Science Foundation (ASC–9109215), and the National Institute of Mental Health (MH47566).

References

Barry, C., & Richardson, J. T. E. (1988). Accounts of oral reading in deep dyslexia. In H. A. Whitaker (Ed.), *Phonological processing and brain mechanisms* (pp. 118–171). New York: Springer-Verlag.

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

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

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.

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.

Coltheart, M. (Ed.). (1987). *Attention and performance XII: The psychology of reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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: Lawrence Erlbaum Associates.

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., Patterson, K., & Marshall, J. C. (Eds.). (1980). *Deep dyslexia*. London: Routledge & Kegan Paul, 2 edition.

Dejerine, J. (1892). Contribution à l'étude anatomoclinique et clinique des differentes variétés de cécité verbale. *Mémoires del la Société de Biologie*, *4*, 61–90.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (this volume). A connectionist model of naming errors in aphasia. In J. Reggia, R. Berndt, & E. Ruppin (Eds.), *Neural modeling of cognitive and brain disorders*. New York: World Scientific.

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, *4*, 161–178.

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

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.

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

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

Hinton, G. E., & Plaut, D. C. (1987). Using fast weights to deblur old memories. In *Proceedings of the 9th Annual Conference of the Cognitive Science Society* (pp. 177–186). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann Machines. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 282–317). Cambridge, MA: MIT Press.

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

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

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language*, *24*, 1–19.

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.

Laine, M., Niemi, J., & Marttila, R. (1990). Changing error patterns during reading recovery: A case study. *Journal of Neurolinguistics*, *5*, 75–81.

Margolin, D. (Ed.). (1992). *Cognitive neuropsychology in clinical practice*. Oxford: Oxford University Press.

Marshall, J. C., & Newcombe, F. (1966). Syntactic and semantic errors in paralexia.

Neuropsychologia, *4*, 169–176.

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., & 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.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.

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 & Kegan Paul.

Mozer, M. C., & Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, *2*, 96–123.

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.

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

Patterson, K., Seidenberg, M. S., & McClelland, J. L. (1989). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neuroscience* (pp. 131–181). London: Oxford University Press.

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

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

17, 291–321.

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., 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.

Riddoch, M. J., & Humphreys, G. W. (Eds.). (1994). *Cognitive neuropsychology and cognitive rehabilitation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). 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. Volume 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

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: Lawrence Erlbaum Associates.

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

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

Seron, X., & Deloche, G. (Eds.). (1989). *Cognitive approaches in neuropsychological rehabilitation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

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.

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.
Weekes, B., & Coltheart, M. (in press). Surface dyslexia and surface dysgraphia: Treatment studies and their theoretical implications. *Cognitive Neuropsychology*.