Effects of Word Abstractness in a Connectionist Model of Deep Dyslexia*

David C. Plaut
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
dcp@cs.cmu.edu

Tim Shallice
Department of Psychology
University College
London, England WC1E 6BT
ucjtsts@ucl.ac.uk

To appear in the Proceedings of the 13th Annual Meeting of the Cognitive Science Society, Chicago, IL, August 1991.

Abstract

Deep dyslexics are patients with neurological damage who exhibit a variety of symptoms in oral reading, including semantic, visual and morphological effects in their errors, a part-of-speech effect, and better performance on concrete than abstract words. Extending work by Hinton & Shallice (1991), we develop a recurrent connectionist network that pronounces both concrete and abstract words via their semantics, defined so that abstract words have fewer semantic features. The behavior of this network under a variety of "lesions" reproduces the main effects of abstractness on deep dyslexic reading: better correct performance for concrete words, a tendency for error responses to be more concrete than stimuli, and a higher proportion of visual errors in response to abstract words. Surprisingly, severe damage within the semantic system yields better performance on *abstract* words, reminiscent of CAV, the single, enigmatic patient with "concrete word dyslexia."

Introduction

Extensive work within cognitive neuropsychology suggests that there are (at least) two separable processing routes for pronouncing a written word: a "semantic" route that recognizes the word and accesses its pronunciation from its meaning, and a "phonological" route that obtains the pronunciation based on spelling-to-sound correspondences. Strong evidence for the separability of these routes comes from the existence of two complementary sets of neurological patients (Marshall & Newcombe, 1973). "Surface" or "semantic" dyslexics (Patterson et al., 1985) have an impaired semantic route and so must rely primarily on spelling-to-sound correspondences in oral reading. In contrast, "deep" dyslexics (Coltheart et al., 1980) have lost the ability to derive phonology directly from print and so can pronounce words only via their meaning. These latter patients shows a number of characteristics. Five types of errors typically occur: semantic (e.g. CHEER $\Rightarrow$ "laugh"), visual (SWORD $\Rightarrow$ "words"), morphological (GROWN $\Rightarrow$ "growing", also called "derivational"), visual-and-semantic (PAPER $\Rightarrow$ "page") and visual-then-semantic (SYMPATHY $\Rightarrow$ "orchestra"). In addition, there is a pronounced part-of-speech effect, with the ordering of correct performance being roughly: nouns $>$ adjectives $>$ verbs $>$ function words. Furthermore, concrete, highly imageable words are read much better than abstract, less imageable words. Finally, pronounceable non-words cannot be read. Of these effects, the morphological errors and part-of-speech effects may well be secondary to other characteristics (see Funnell, 1987), but any account of the disorder needs to explain all the other apparently independent symptoms.

In the conclusion of their review article, "Deep Dyslexia since 1980," Coltheart, Patterson & Marshall (1987) argue that deep dyslexia presents cognitive neuropsychology with a major challenge. They raise two main issues specific to the domain of reading. First, they argue that standard "box-and-arrow" information-processing accounts of deep dyslexia (e.g. Morton & Patterson, 1980) provide no explanation for why such a variety of symptoms should virtually always occur in patients who make semantic errors. Second, they point out that the standard explanations

*Most of the research presented in this paper was carried out while the authors were visiting researchers in the Departments of Psychology and Computer Science at the University of Toronto under the generous support and guidance of Geoff Hinton, whom we believe deserves to be a co-author of this paper but would not be persuaded to be included. We also wish to thank Marlene Behrmann and Angela Hickman for their help. All of the simulations described in this paper were run on a Silicon Graphics Iris-4D/240S using the Xerion simulator developed by Tony Plate. This research was supported by grant 87-2-36 from the Alfred P. Sloan Foundation.

for semantic errors and for effects of abstractness involve *different* impairments along the semantic route.

> The loss of semantic information for abstract words that explained visual errors in oral reading cannot readily explain semantic errors in oral reading, since semantic errors typically occur on moderately concrete words.... The deficit in the semantic routine that gives a pretty account of semantic errors is, rather, an abnormal sloppiness in the procedure of addressing a phonological output code from a set of semantic features. .... Must we now postulate several different semantic-routine impairments in deep dyslexia, and if so, why do we not observe patients who have one but not the other: in particular, patients who make semantic errors but do not have difficulty with abstract words? [Coltheart et al., 1987, pp. 421-422]

Hinton & Shallice (1991, hereafter H&S) put forward a connectionist account that addresses the first issue—why semantic, visual and mixed visual-and-semantic errors co-occur. Based on previous work by Hinton & Sejnowski (1986) with Boltzmann Machines, they trained a recurrent back-propagation network to map from the orthography of 40 three- or four-letter words to a simplified representation of their semantics, described in terms of 68 predetermined semantic features. They then systematically lesioned the network, by removing proportions of units or connections, or by adding noise to the weights, and found that the damaged network occasionally settled into a pattern of semantic activity that satisfied the response criteria for a word other than the one presented. These errors were more often semantically and/or visually similar to presented stimuli than would be expected by chance. While the network showed a greater tendency to produce visual errors with damage near the input layer and semantic errors with damage near the output layer, both types of error occurred for almost all sites of damage.

H&S explain the co-occurrence of visual and semantic errors in terms of the effects of damage in a network that builds attractors in mapping between two arbitrarily related domains (see Figure 1). The network can generate completely different meanings from visually similar words (e.g. CAT and COT) by constructing large basins of attraction around each familiar meaning, such that any initial semantic pattern within the basin will move to that meaning. Visually similar words are free to generate similar initial semantic patterns as long as they each fall somewhere within the appropriate basin of attraction. Damage within the semantic system distorts these basins, occasionally causing the normal initial semantic pattern of a word to be "captured" within the basin of a visually similar word. Essentially, the layout of attractor basins must be sensitive to both visual and semantic similarity, and so these metrics are reflected in the types of errors that occur as a result of damage.

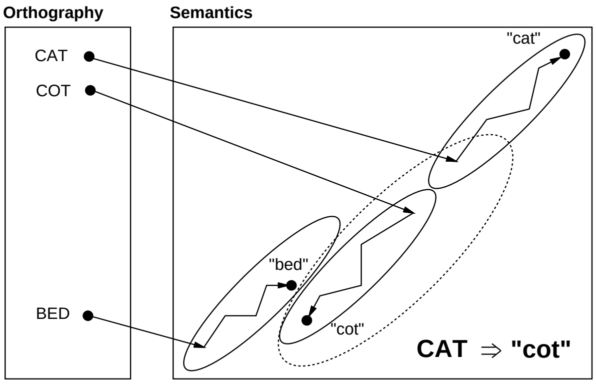

Figure 1: How damage to semantic attractors can cause visual errors. The solid ovals depict the normal basins of attraction; the dotted one depicts a basin after semantic damage.

The nature of attractors in the H&S model addresses the first challenge raised by Coltheart, Patterson & Marshall—why various error types virtually always co-occur in deep dyslexia.The current research investigates whether essentially the same account can be extended to successfully address the second difficulty they raise, relating to the abstractness of the stimuli. The next section describes in more detail the effects of abstractness on the reading behavior of deep dyslexics. Following this, we define a semantic representation capable of describing both concrete and abstract words, and develop a network that maps from orthography to phonology via semantics. We then subject the network to a variety of lesions and compare its impaired performance with that of deep dyslexics.

Effects of abstractness in deep dyslexia

The effect of the abstractness of the stimulus on deep dyslexic reading has been investigated in a number of ways. The most basic is its effect on the probability that a word will be read correctly. Coltheart et al. (1987) claim that all patients who make semantic errors find concrete words easier to read than abstract ones. In many patients a very large difference is observed: 73% vs. 14% for KF (Shallice & Warrington, 1980), 67% vs. 13% for PW and 70% vs. 10% for DE (Patterson & Marcel, 1977).

A more subtle effect is the way that the concreteness of a word can affect the probability of the occurrence of visual errors. In the three patients in which the relative concreteness of the stimulus and response for visual errors has been investigated, the response was consistently more concrete for two of them (88% for GR (Barry & Richardson, 1988); 73% for KF) and showed a similar trend in the third ($p < .06$ for PS (Shallice & Coughlan, 1980)). The same effect is also apparent in the corpora of visual errors made by PW and DE (see Coltheart et al., 1980, Appendix 2) and by FM (Gordon et al., 1987)). In addition, stimuli with a low concreteness rating produce more visual errors than stimuli with high concreteness in the two patients in which it has been examined (KF and PS). Also, stimuli producing visual errors were less concrete than those producing semantic errors for PD (Coltheart, 1980) and FM (but not for GR). Thus a semantic variable—concreteness—clearly influences the nature of *visual* errors.

There is a single known exception to the advantage for concrete words shown by deep dyslexics: patient CAV with "concrete word dyslexia" (Warrington, 1981). CAV failed to read words like MILK and TREE but succeeded on 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. However, the incidence of visual errors was approximately equal for words above and below the median in concreteness. While CAV made no more semantic errors than might be expected by chance, he appeared to be relying at least in part on the semantic route because his performance improved when given a word's semantic category. CAV is clearly a very unusual patient, but any account of the relation between visual errors and concreteness can hardly ignore him.

A semantic representation for concrete and abstract words

The type of semantic feature representation used by H&S is quite similar to that frequently employed in psychological theorizing on semantic memory (e.g. Smith & Medin, 1981). More complex, frame-like representations can be implemented using this approach if units can represent a conjunction of a role and a property of what fills it (Hinton, 1981). More critically for the present purpose, there is a natural extension to the problem of the effects of imageability. Jones (1985) has argued that words vary greatly in the ease with which predicates about them can be generated, and that this measure reflects a psychologically important property of semantic representation. For example, more predicates can be generated for basic-level words than for subordinate or superordinate words (Rosch et al., 1976). Jones showed that there is a very high correlation (0.88) between a measure of ease-of-predication and imageability, 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 imageability and part-of-speech in deep dyslexia can be accounted for by assuming that the semantic route is sensitive to ease-of-predication. Within the present framework, the natural way to realize this distinction is by representing the semantics of concrete and abstract words in terms of different numbers of features.

TART	TACT	GRIN	GAIN	FLAN	PLAN	REED	NEED
TENT	RENT	LOCK	LACK	HIND	HINT	LOON	LOAN
FACE	FACT	ROPE	ROLE	WAVE	WAGE	CASE	EASE
DEER	DEED	HARE	HIRE	FLEA	PLEA	FLAG	FLAW
COAT	COST	LASS	LOSS	STAR	STAY	POST	PAST

Figure 2: The 40 words used in the simulation.

To examine the effect of concreteness on visual errors, a set of 20 abstract and 20 concrete words were chosen such that each pair of words differed by a single letter (see Figure 2). We represented the semantics of each of these words in terms of 98 semantic features.[1] Sixty-seven of these are taken from the H&S semantic features for concrete words (e.g. *main-shape-3d*, *found-woods*, *living*). The 31 additional features (e.g. *has-duration*, *relates-location*, *quality-difficulty*) are required to make distinctions among abstract words, but occasionally apply to concrete words as well. Overall, concrete and abstract words differ systematically in their semantic representations: concrete words have an average of 18.2 features while abstract words have an average of only 4.7 features. We do not claim that this type of representation adequately captures the richness and subtlety of the true meanings of any of these words. Rather, we claim that it captures important qualitative distinctions about the relationships *between* word meanings—namely, that similar words (e.g. LACK and LOSS) have similar representations, and that there is a systematic difference between the semantics of concrete and abstract words reflecting their relative ease of predication.

Mapping from orthography to phonology via semantics

A network that maps from orthography to phonology via semantics was developed incrementally. An "input" network, analogous to the H&S model, was trained to map from orthography to semantics. A similarly structured "output" network was trained separately to map from semantics to phonology. These two networks were then combined into the complete network, shown in Figure 3.

The task of the input network is to generate the semantics of each word from its orthography. Orthography is represented in terms of 4 groups of 8 features, with a separate group for each letter in a word. The set of features was designed to ensure that visually similar letters (e.g. E and F) have similar representations, while keeping the number

[1] See Plaut & Shallice (1991) for precise details of the orthographic, semantic, and phonological representations of words, as well as for how the network architecture and training procedure were motivated.

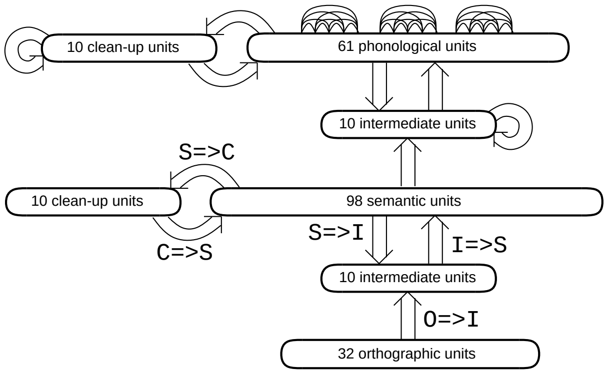

Figure 3: The network for mapping from orthography to phonology via semantics. Arrows represent unidirectional sets of connections between groups of units—sets that will be lesioned are labeled by the initials of the source and destination unit groups (e.g. $\texttt{O} \Rightarrow \texttt{I}$ for orthographic-to-intermediate connections).

of features to a minimum. The architecture of the input network is shown in the bottom half of Figure 3. It was trained with an iterative version of the back-propagation learning procedure, known as "back-propagation through time" (Rumelhart et al., 1986), to activate the appropriate semantic units for a word when presented with the word's orthography corrupted by independent gaussian noise with mean 0.0 and standard deviation $\sigma = 0.1$. The purpose of training on noisy input is to encourage the development of strong semantic attractors by enforcing a particular kind of generalization: inputs that are *near* known patterns must give identical responses. The network was required to activate each semantic unit to within 0.1 of its correct state over the last 3 of 8 iterations to ensure it had developed stable, accurate fixedpoints for each word. The network satisfied these criteria reliably after 4700 sweeps through the training set.

Some procedure is needed for converting the pattern of semantic activity produced by the input network into an explicit response. H&S use an *external* procedure, comparing the semantic activity produced by the network with the correct semantics of all known words, selecting the closest-matching word as long as the match is sufficiently good (the *proximity* criterion) and sufficiently better than any other match (the *gap* criterion). However, H&S provide no evidence that these criteria adequately approximate the input requirements of a network that can generate actual phonological output. An even more severe problem is that the criteria are based on the semantic representations themselves. Any differences found in performance on concrete and abstract words might simply be due to an inherent bias in the response criteria. However, by developing an output network that pronounces concrete and abstract words equally well under normal operation, any systematic differences observed under damage must be due to properties of the network itself and not some external interpretation procedure.

Phonology is represented in terms of 7 "slots," each consisting of a group of position-specific, mutually-exclusive phoneme units (including one for the "null" phoneme). There are three slots for the initial (onset) consonant cluster, one slot for the vowel, and three slots for the final (coda) consonant cluster. The task of the output network is to generate the phonological representation of each word from its semantic representation. The architecture of this network is shown in the top half of Figure 3. In addition to the major sets of connections, phoneme units in the same consonant (or vowel) cluster are fully interconnected. This connectivity allows units within a slot to develop a "winner-take-all" strategy while still cooperating with units in other slots within the same cluster. The clean-up units provide for coordination and competition between clusters.

The output network was trained in a way that maximizes the strength of the attractors it develops—no attempt was made to simulate the development or mode of operation of the human speech production system. Specifically, the "direct" pathway (from semantics to phonology) was trained to produce the correct phonemes of each word during the last 2 of 5 iterations when presented with its semantics corrupted by noise ($\sigma = 0.1$). After about 3000 sweeps through the training set, the activity of each phoneme unit was accurate to within 0.2 of its correct value for each word. At this point, intra-phoneme connections and the clean-up pathway were added and the amount of noise was increased to 0.2. In this way the clean-up pathway learned to compensate for the limitations of the direct pathway when pressed by severely corrupted input. The network was trained to produce the correct phonemes over the last 3 of 8 iterations to within 0.1 of their correct values. The amount of noise prevented the network from achieving this criterion consistently, and after 18,000 training sweeps performance had ceased to improve. However, the network easily satisfied the criterion for every word given uncorrupted input.

Finally, the output network was attached to the input network and given about 100 sweeps of additional training with the weights of the input network held fixed. This ensured that the output network could generate the correct pronunciation of each word over the last 3 of 14 iterations with semantics generated by the input network rather than being clamped.

The effects of lesions

After training, the complete network successfully derives the semantics and phonology of each word when presented with its orthography. We model the neurological damage of deep dyslexic patients by removing a proportion of the connections between groups of units in the network. As in patients, this damage impairs the ability of the network to derive the correct pronunciations of words. In order to directly compare the behavior of the damaged network with the reading responses of patients, we used the following procedure to interpret the corrupted output of the network as an oral response. Given the pattern of activity over phoneme units produced by the stimulus, we determined the most likely binary output vector for each slot, interpreting unit states as independent probabilities. If each of these vectors had exactly one phoneme active and probability greater than 0.6, the set of active phonemes constituted the response of the network (which might be correct or an error). Otherwise, the network was considered to have made an omission—in fact, patients frequently produce no response to a word, or respond, "I don't know." This procedure is closely related to the maximum-likelihood interpretation of the cross-entropy error function that was used to train the network (Hinton, 1989). In contrast to the response criteria that H&S applied to semantics, it does not rely on any knowledge of what the network has been trained on—it only considers the *form* of the output representation. In particular, it cannot distinguish concrete from abstract words.

Each of the 5 main sets of connections in the input network was subjected to "lesions" of a wide range of severity, in which a proportion of the connections were chosen at random and removed. Fifty instances of each location and severity of lesion were carried out, and correct, omission, and error responses were accumulated. Figure 4 shows the overall correct performance of the network as a function of lesion severity. Considering correct responses to concrete and abstract words separately, there is a significant advantage for concrete words (55% correct) over abstract words (49% correct, $F(1, 2205) = 489.3, p < .001$). The relative difference in correct performance between these two sets is shown in Figure 5. Two main results are apparent from the figure. The first is that the advantage for concrete over abstract words arises almost entirely from lesions to the direct pathway, where the majority (88%) of errors are produced. The second, unexpected result is that severe lesions of the clean-up pathway produce the reverse advantage—abstract words are responded to more accurately than concrete words ($F(1, 49) > 22$, $p < .001$ for each of S$\Rightarrow$C(0.5,0.7) and C$\Rightarrow$S(0.5,0.7)). This result is consistent with what is known about the concrete word dyslexic, CAV (Warrington, 1981). His reading disorder was quite severe initially, and he also showed an advantage for abstract words in picture-word matching tasks and with auditory presentation, suggesting modality-independent impairment at the level of the semantic system.

Error responses were categorized in terms of their visual

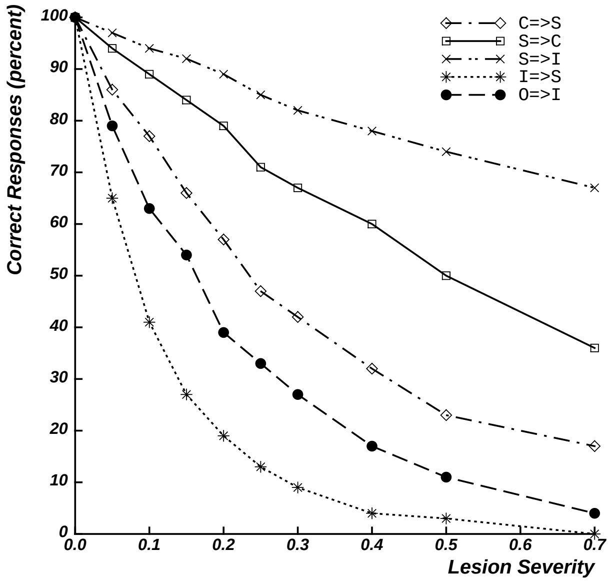

Figure 4: Correct response rates as a function of severity of lesions to the 5 main sets of connections in the input network.

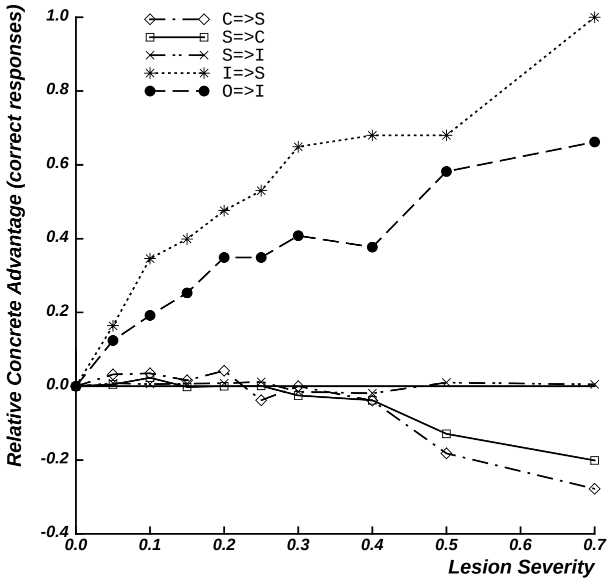

Figure 5: The relative difference in correct performance between concrete and abstract words, $(C - A)/(C + A)$, where C and A are the number of correct responses to concrete and abstract words, respectively. Positive values reflect a concrete advantage.

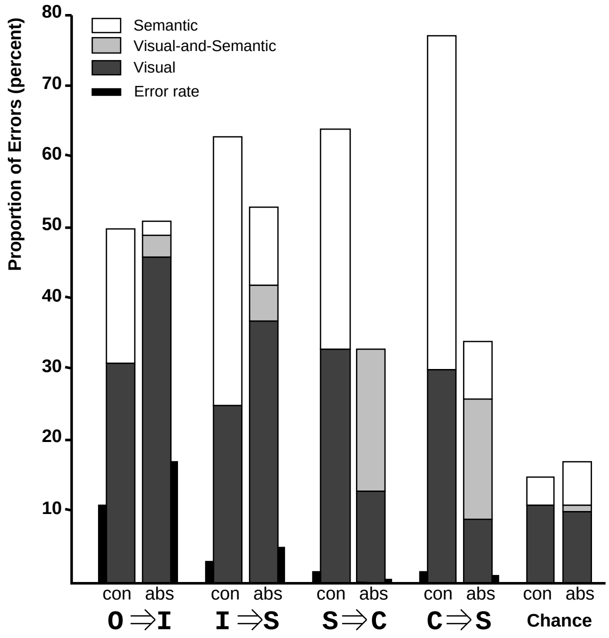

Figure 6: Overall proportions of error types for concrete (con) and abstract (abs) words for each lesion location (except S⇒I lesions which produce so few errors). Remaining errors (summing to 100%) are unrelated to the stimulus.

and semantic similarity to the stimulus. Words were considered visually similar if they overlapped in two or more letters, and semantically similar if their semantic representations overlapped by at least 84% for concrete words and 95% for abstract words.[2] Figure 6 shows, for concrete and abstract words separately, the overall error rates and the proportions of error types for each lesion location (and for "chance" error responses chosen randomly from the word set). Overall, the network replicates (on a different word set) the H&S finding of mixtures of error types for lesions throughout the network, including purely visual errors for lesions entirely within the semantic clean-up system. Interestingly, a number of the unclassified errors are actually of the visual-then-semantic type found in deep dyslexia (e.g. PLAN ⇒ (flan) ⇒ "tart"). When this type of error occurs, the semantic activity tends to match the intermediate word only moderately well.

A comparison of error types for concrete and abstract words revealed that the proportion of errors which are visual is higher for abstract words ($F(1, 2205) = 92.24$, $p < .001$), while the proportion of errors which are semantic is higher for concrete words ($F(1, 2205) = 228.8$, $p < .001$). This effect is most clearly shown in Figure 6 for lesions of the direct pathway (which produce the majority of errors). As a measure of the "abstractness" of the errors produced by a lesion, we used the number of errors to abstract words minus the number of errors to concrete words. Applying this measure to visual and semantic errors separately revealed that visual errors are more abstract (mean 0.1644) than semantic errors (mean -0.1458, $F(1, 2205) = 249.6$, $p < .001$). Finally, for each pair of visually similar words of contrasting types (e.g. TART and TACT), we compared how often each word produced the other as an error. Overall, abstract words are more likely to produce the paired visually similar concrete word as an error than *vice versa* (Wilcoxon signed-ranks test, $n = 900$, $Z = 3.68$, $p < .001$). However, severe lesions of the clean-up pathway produce the opposite effect ($n = 80$, $Z = 1.98$, $p < .025$).

These effects can be understood in the following way. As abstract words have fewer semantic features, they are less effective than concrete words at engaging the semantic clean-up mechanism and must rely more heavily on the direct pathway. Concrete words are read better under lesions to this pathway because of the stronger semantic clean-up they receive. Abstract words are more likely to produce visual errors as the influence of visual similarity is strongest in the direct pathway. Slight or moderate damage to the clean-up pathway impairs what little support abstract words receive from this system, but also impairs concrete words, producing no relative difference. Under severe damage to this pathway, the processing of most concrete words is impaired but many abstract words can be read solely by the direct pathway, producing an advantage of abstract over concrete words in correct performance.

Discussion

> No architecture which anyone has proposed contains anything remotely approaching a component damage to which would produce all the nine symptoms we are considering. [Coltheart et al., 1987, p. 417]

The symptoms of deep dyslexia that Coltheart, Patterson & Marshall are referring to are: (1) semantic errors, (2) visual errors, (3) function word substitutions, (4) morphological/derivational errors, (5) advantage for concrete over abstract words, (6) advantage for content over function words, (7) inability to read non-words, (8) inability to access phonology from print for words or in non-naming tasks, and (9) impaired writing and spelling. This last symptom is beyond the scope of the present simulation, which is concerned solely with reading. Symptoms 7 and 8 are commonly attributed to the loss of a phonological route employing spelling-to-sound correspondences. The current research demonstrates how symptoms 1, 2, and 5 can arise from unitary lesions to a network trained to map

[2]The definition of semantic similarity is more complicated because of the systematic differences between concrete and abstract semantics and because the semantic representations are not organized into categories as in the H&S simulations. Note that two typical unrelated words have roughly 67% overlap if both are concrete and 91% if both are abstract.

from orthography to phonology via semantics, in which abstract words have far fewer semantic features. From the arguments of Jones (1985), Funnell (1987) and others, it seems likely that the morphological/derivational errors reduce to a special case of mixed visual-and-semantic errors, and part-of-speech effects can be accounted for in terms of the ease-of-predication variable on which our simulation was based. Thus the effects we have demonstrated may also account for the remaining symptoms (3, 4, 6). Most critically, the present simulation provides an explanation for the puzzling cross-domain interactions that occur between the abstractness of stimuli/responses and the occurrence of visual errors. The explanation has some similarities to those previously offered for the interaction (e.g. Morton & Patterson, 1980; Shallice & Warrington, 1980) but these were essentially *ad hoc* verbal extrapolations from cascade notions unrelated to other aspects of the syndrome, without even a principled account of the abstract/concrete difference. The present account is supported by a simulation, is linked to explanations of other aspects of the syndrome, and offers the possibility of also addressing concrete word dyslexia.

Are the differences obtained for performance on abstract and concrete words simply due to differences we built in? The answer is yes, but not *simply*. The way the contrast was realized—in terms of a difference in number of features—was independently motivated, and the effects obtained were complex and not *transparent* from the characteristics of the semantic representation. In addition, we ensured that normal performance on the two types of words was equivalent, so that the contrast is reflected only in the performance under damage, over which we had no direct control. Taken together, the replication of the diverse set of symptoms of deep dyslexia through unitary lesions of a network that pronounces words via their meanings strongly suggests that the computational principles underlying the network's behavior may shed light on normal and impaired reading mechanisms in humans.

References

Barry, C. & Richardson, J. T. E. (1988). Accounts of oral reading in deep dyslexia. In Whitaker, H. A., editor, *Phonological Processing and Brain Mechanisms*. Springer-Verlag, New York.

Coltheart, M. (1980). Deep dyslexia: A right-hemisphere hypothesis. In Coltheart, M., Patterson, K. E., & Marshall, J. C., editors, *Deep Dyslexia*, chapter 16, pages 326–380. Routledge, London.

Coltheart, M., Patterson, K., & Marshall, J. C. (1987). Deep dyslexia since 1980. In Coltheart, M., Patterson, K. E., & Marshall, J. C., editors, *Deep Dyslexia*, chapter 18, pages 407–451. Routledge, London.

Coltheart, M., Patterson, K. E., & Marshall, J. C. (1980). *Deep Dyslexia*. Routledge, London.

Funnell, E. (1987). Morphological errors in acquired dyslexia: A case of mistaken identity. *Quarterly Journal of Experimental Psychology*, 39A:497–539.

Gordon, B., Goodman-Schulman, R., & Caramazza, A. (1987). Separating the stages of reading errors. Technical Report 28, Cognitive Neuropsychology Laboratory, Johns Hopkins University, Baltimore, MD.

Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In Hinton, G. E. & Anderson, J. A., editors, *Parallel Models of Associative Memory*, pages 161–188. Lawrence Erlbaum Associates, Hillsdale, NJ.

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

Hinton, G. E. & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann Machines. In Rumelhart, D. E., McClelland, J. L., & the PDP research group, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, pages 282–317. MIT Press, Cambridge, MA.

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

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

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

Morton, J. & Patterson, K. (1980). A new attempt at an interpretation, or, and attempt at a new interpretation. In Coltheart, M., Patterson, K. E., & Marshall, J. C., editors, *Deep Dyslexia*, chapter 4, pages 91–118. Routledge, London.

Patterson, K. E., Coltheart, M., & Marshall, J. C. (1985). *Surface Dyslexia*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Patterson, K. E. & Marcel, A. J. (1977). Aphasia, dyslexia and the phonological coding of written words. *Quarterly Journal of Experimental Psychology*, 29:307–318.

Plaut, D. C. & Shallice, T. (1991). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*. Submitted.

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 representations by back-propagating errors. *Nature*, 323(9):533–536.

Shallice, T. & Coughlan, A. K. (1980). Modality specific word comprehension deficits in deep dyslexia. *Journal of Neurology, Neurosurgery and Psychiatry*, 43:866–872.

Shallice, T. & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In Coltheart, M., Patterson, K. E., & Marshall, J. C., editors, *Deep Dyslexia*. Routledge, London.

Smith, E. E. & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press, Cambridge, MA.

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