Are Attractors Necessary in a Connectionist Model of Deep Dyslexia?

Lori A. Blanc and J. Devin McAuley and Janet Wiles
Schools of Information Technology and Psychology
University of Queensland
Brisbane, Queensland 4072 Australia
janetw@it.uq.edu.au

Hinton and Shallice (1991) proposed a recurrent connectionist model of deep dyslexia in order to provide a unified account of the co-occurrence of visual, semantic, and mixed errors characteristic of deep dyslexic patients. They argued that the deep-dyslexic error pattern was a natural consequence of lesioning a connectionist model that maps orthography to semantics using attractors for word meanings. This study was extended by Plaut and Shallice (1993) who showed that a variety of recurrent network architectures and learning algorithms give rise to the same characteristic pattern of errors. Based on this work, Plaut and Shallice (P&S) identified four necessary properties of a connectionist model of deep dyslexia: (1) distributed orthographic and semantic representations; (2) gradient-descent learning; (3) attractors for word meanings; and (4) greater richness of concrete versus abstract semantics.

Although attractors for word meanings were a basic assumption of the H&S and P&S models, H&S acknowledged that some of the phenomena can be observed in a feed-forward network. Since no results of a feed-forward model were reported by either H&S or P&S, it remains unclear whether attractors are necessary to model the characteristics of deep dyslexia. This possibility is investigated in the present study using a 3-layer feed-forward network.

Background

Deep dyslexia, one of several acquired dyslexias (Marshall and Newcombe, 1973) is a selective deficit in reading which occurs following localised brain damage. Deep dyslexia is characterised by semantic confusions (e.g., reading "peach" as "apricot"). However, deep dyslexic patients frequently exhibit other types of reading errors, including visual errors (e.g., reading "peach" as "beach"), and mixed visual-and-semantic errors, (e.g., reading "dog" as "hog"), which are difficult to explain using dual route models of reading. One possible explanation is via connectionism, which supports the possibility that reading processes are distributed throughout a single system. In this paradigm, one lesion can produce multiple error types.

A Feed-Forward Model

The feed-forward model explored in this study consisted of three-layers (orthographic, intermediate, and semantic). The task of the network was to generate the semantic representation of a word when presented with its orthography, using the H&S data set. The data set consisted of 40 three and four letter words subdivided into five semantic categories. The orthographic representation of a word consisted of a distributed vector of orthographic features while the word's meaning was represented by a distributed vector of semantic features.

The network was trained using backpropagation to map orthography to semantics for each of the 40 words in the data set. The trained network was systematically lesioned by either deleting a proportion of the weights in the network or by adding random noise to the weights. The lesioned network was then tested on each word in the data set recording correct and incorrect responses, according to the criteria established by H&S. Errors were classified as either semantic, visual, mixed or other. Words with one letter in common in the same position within the word were considered to be visually similar, while words within the same category were considered to be semantically similar.

Results and Discussion

As in the H&S and P&S studies, the likelihood of the observed error types occurring by chance was assessed by comparing the observed error rates with the expected chance rates in the data set. Consistent with the H&S and P&S findings, simulations of the feed-forward network showed a co-occurrence of visual, semantic and mixed errors independent of lesion location at higher than chance levels. The results suggest that vector similarity and damage to the mapping process are sufficient to explain why the deep-dyslexic error pattern occurs. We conclude that attractors are not the fundamental cause of the deep-dyslexic error pattern.

References

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

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

Plaut, D. C. and Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10(5):377–500.