A Model of Prefrontal-Hippocampal Interactions in Strategic Recall

Jean C. Lim, lim@curie.psychology.mcmaster.ca
Suzanna Becker, becker@mcmaster.ca
Department of Psychology, McMaster University, 1280 Main St. West, Hamilton, ON, Canada.

May 24, 2000

Retrieval of episodic memories may be aided by the prefrontal cortex either by its providing contextual, temporal source cues or by serving an executive role - strategically organizing information into chunks, categorizing, and separating lists (Stuss, 1986). The hippocampal region is also believed to play a role in encoding episodic memories (Zola-Morgan et. al., 1990). Modelling human performance on free recall tasks that involve strategic organization of items is difficult because temporal memory of events, delayed rewards and learning in the absence of external reinforcement are required. For example, in the California Verbal Learning Test (CVLT) the task is to study a list of 16 words (four words each from four different semantic categories) and recall the list over five repeated trials (Delis et. al., 1987). Young healthy subjects typically use a semantic clustering strategy to recall the list. Elderly and frontal lobe damaged patients fail to subjectively organize such words and show poorer recall performance (Hultsch, 1975).

To simulate the hippocampal component of our model, we used a Hopfield network (Hopfield, 1982), because of its rapid learning, pattern association, recall and recognition capabilities. For our prefrontal module, we used a recurrent network trained with a reinforcement learning rule similar to that proposed by Barto and Sutton (1986) which can detect correlations between traces of past inputs and changes in outputs. For our list-learning task, a positive reward signal was used to strengthen relevant prefrontal weights during study. During retrieval, recall of non-studied items resulted in an internally generated negative signal.

The hippocampal module consisted of 400 recurrent, symmetrically connected units with no self-feedback connections. Bidirectional and symmetrical weights connected each unit of the Hopfield network to two layers: (1) the prefrontal cortex – a layer of 10 units; and (2) an input/output layer of 52 units – localist representations of the vocabulary words.

Weights to each localist word unit were pre-trained with a Hebbian update rule. This established the network's pre-experimental vocabulary of 52 word patterns. Sixteen of these words, drawn from four different semantic categories (four words from each), were the study list words. Of the remaining 36 words, 20 were semantically similar to the study list, eight were drawn from two new semantic categories, and eight were sparse random vectors. The representation of a word consisted of a distributed vector of 400 semantic features, with semantically related words having more highly correlated feature vectors. The simulation consisted of 5 study and recall trials. During study, the network was trained on the 16 list word patterns. During retrieval, the prefrontal activations served as cues to recall the list words. Learning in both phases took place in the connections between the prefrontal and hippocampal modules. A recalled word could be "correct", a perseveration (repetition) or an intrusion (non-list word). Clustering performance was determined by comparing the observed semantic clustering score to the expected clustering score.

We modelled a frontal lesioned network by freezing its prefrontal weights at zero. Calculated ratios of observed to expected cluster scores showed that the lesioned network did not cluster above chance whereas the normal network did. The lesioned model also produced, in order of frequency, more perseverative errors, similar intrusions and random intrusion errors than the normal model. Simulation results suggest that an elderly or frontal-lobe damaged subject, modelled by a hippocampal system operating independently of the prefrontal layer, is capable of pattern recognition and recall when given external guidance and cues. To perform more complex tasks such as free recall and provide context-rich source cues, an additional layer, represented in our model as the prefrontal cortex, is needed. With temporally predictive reinforcement learning, the prefrontal module was able to detect semantic clustering of word patterns, and use this to generate retrieval cues to recall the list items optimally.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). The California Verbal Learning Test. San Antonio, Tx: Psychological Corporation.

Hopfield, J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences.* 79:2554-2558.

Hultsch, D. (1975). Adult age differences in retrieval: Trace-dependent and cue-dependent forgetting. *Developmental Psychology*, 12, 83-84.

Stuss, D., Benson, F. (1986) The Frontal Lobes. Raven Press. New York.

Sutton, S., Barto, A. (1981) Toward a Modern Theory of Adaptive Networks: Expectation and Prediction. *Psychological Review* 88:2:135-170.

Zola-Morgan, S., Squire, L. (1990). The neuropsychology of memory: parallel findings in humans and nonhuman primates. In A. Diamond (Ed.), *The Development of Learning.* New York: New York Academy of Sciences.