Title
Towards a Multiple Components Model of Human Memory: A Hippocampal-Cortical Memory Model of Encoding Specificity

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 23(23)

ISSN
1069-7977

Authors
Kwok, Kenneth
McClelland, James L.

Publication Date
2001

Peer reviewed

Towards a Multiple Components Model of Human Memory: A Hippocampal-Cortical Memory Model of Encoding Specificity

Kenneth Kwok (kenkwok@cmu.edu)
James L. McClelland (jlm@cnbc.cmu.edu)
Department of Psychology, Carnegie Mellon University and the
Center for the Neural Basis of Cognition
115 Mellon Institute, 4400 Fifth Avenue,
Pittsburgh, PA 15213, USA

Abstract

While many researchers studying memory today subscribe to a multiple memory systems view of human memory, most existing models of memory do not explicitly embody multiple interacting memory systems in their accounts of human memory performance. The present work is aimed at developing a connectionist model of human memory which: (1) takes into account known principles about the neural basis of human learning and memory; and, (2) is an instantiation of a multiple components memory system along the lines of a *components of processing* framework of memory (Roediger, Buckner, & McDermott, 1999).

The *Hippocampal-Cortical Memory Model* (McClelland, McNaughton & O'Reilly, 1995) comprises two memory components: a hippocampal component which supports rapid learning, and a cortical component which learns more slowly in order to develop integrative representations of the statistical characteristics of the environment across many learning episodes. It provides a suitable starting point for our endeavor because it addresses two important aspects of human memory: (1) the ability of humans to rapidly learn new arbitrary associations; and (2) the fact that memory performance is affected by pre-existing knowledge.

The memory phenomenon that we address here with the model is that of encoding specificity. An early statement of the *encoding specificity principle* (ESP) by Tulving & Thomson (1973) asserts that: *"…specific encoding operations performed on what is perceived determine what is stored, and what is stored determines what retrieval cues are effective in providing access to what is stored"*. This principle was used to explain a pattern of results from a paired-associate memory task in which target words were studied with cues which were weak associates, and memory for the target words was subsequently tested either by cued recall using novel cues which were strong pre-existing associates of the target words, or by recognition. Two key aspects of the results accounted for by the ESP were: (1) the relative ineffectiveness of extralist cues (strong associates of the target words) compared to intralist cues (weak associates) in facilitating recall of the target words; and, (2) recognition failure of words that had been recalled to extralist cues. The crux of this explanation lies in a belief that successful recall and recognition depends on the episodic trace of the event being sufficiently similar to the properties of the retrieval information.

We simulated the ESP experiments by presenting stimuli representing *context*, *relation*, *cue* and *target words* to our model in the form of patterns of activity. After the model had learnt the appropriate associations during a training phase, recall was simulated by presenting the model with cue and context patterns, with the network filling in the target word and relation patterns. Recognition was simulated by presenting the cue and target word patterns, with the network filling in the context and relation patterns.

Our model was able to reproduce Tulving and Thomson's pattern of results in simulations. However, unlike their original explanation which places the burden of successful recall and recognition on an episodic memory system, successful performance of our model in this task relies on both pre-experimental knowledge of word associations in the cortical system and newly learnt associations in the hippocampal system. Hence, in line with a *components of processing* framework, encoding specificity can be understood as a phenomenon that is neither purely episodic nor semantic but relies on both forms of memory.

Furthermore, as a mechanistic instantiation of the ESP, our model is able to make explicit what is meant by statements such as "the trace of the event is sufficiently similar to the properties of the retrieval information" – in a connectionist framework, this is simply pattern completion on a suitable retrieval cue. This in turn provides inroads for further exploration of the parameters and mechanisms which make up a multiple components memory model, and ultimately for characterizing the roles played by different components of the memory system and their interactions, in subserving human memory performance. Further work with the model is being directed at these issues.

References

McClelland J.L., McNaughton B. & O'Reilly R.C. (1995). Why there are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Success and Failures of Connectionist Models of Learning and Memory. *Psy. Rev.* 102, 419-457.

Tulving E., & Thomson D. (1973) Encoding specificity and retrieval processes in episodic memory. *Psy. Rev.* 80(5), 352-373

Roediger,H.L., Buckner,R., & McDermott,K. (1999). Components of Processing. In J. Foster & Jelicic M. (Eds.), Memory: Systems, Process or Function? (pp. 31-65). Oxford University Press.