Complimentary Learning System (CLS)

- A synthesis of a wide variety of findings concerning hippocampal function — particularly neuroscience and behavioral memory phenomena (McClelland, McNaughton, & O'Reilly, 1995)

- The paper provides *computational insight* into why the hippocampus exists in terms of a well identified limitation found in artificial neural network systems from AI/Machine learning

- Highly influential theory developed across multiple papers, explains processes of encoding vs. recall, recognition memory, consolidation, interactions between hippocampus and neocortex

Rumelhard, McClelland & Connectionism

- While symbolic AI systems were useful metaphors there was a lack of biological plausibility to the kind of computations they perform

- The brain isn't much like a traditional von Neumann computer architecture (i.e., the one in you laptop), but instead is instantiated in neural hardware

- CONNECTIONISM is a branch of cognitive science devoted to the study of how interconnected, neural-like assemblies might give rise to human thought

- Key idea of DISTRIBUTED REPRESENTATION whereby multiple unit contribute to the representation of a thought or idea

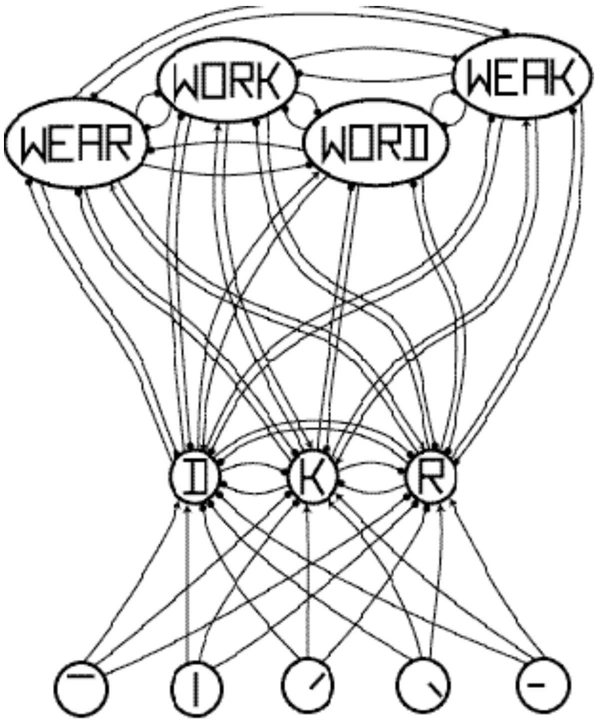


Figure 1. Interactive Activation Network Model (after McClelland and Rumelhart, 1981).

Major problem: Catastrophic interference

Context B (Form A-B)

locomotive-dishtowel

table-street

carpet-idea

Context C (Form A-C)

locomotive-banana

table-basket

carpet-pencil

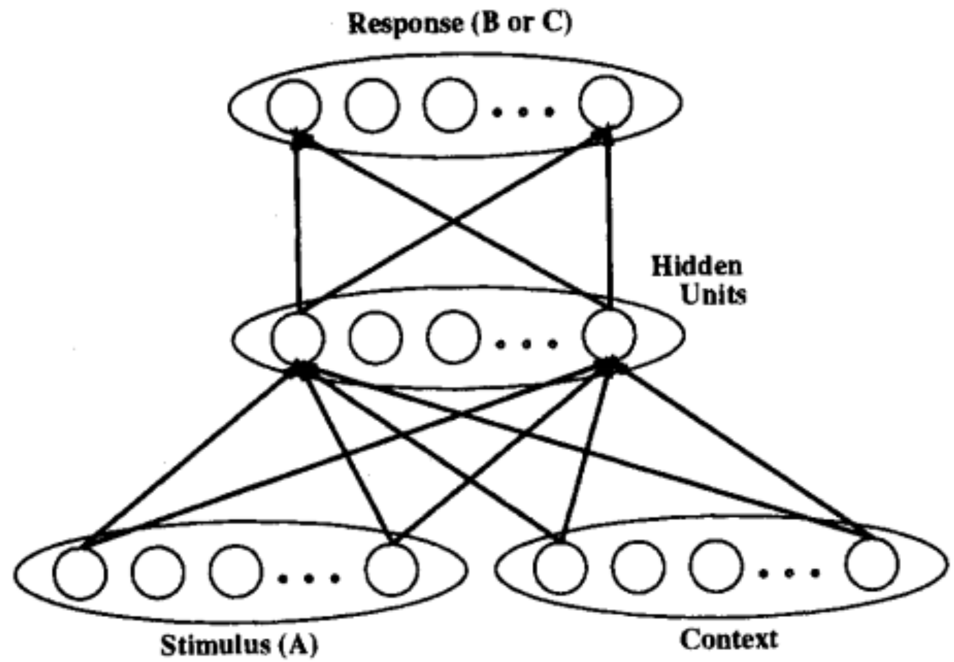


Figure 9. Our depiction of the network used by McCloskey and Cohen (1989) to demonstrate catastrophic interference in back-propagation networks. All output units receive connections from all hidden units, and all hidden units receive inputs from both sets of input units.

Major problem: Catastrophic interference

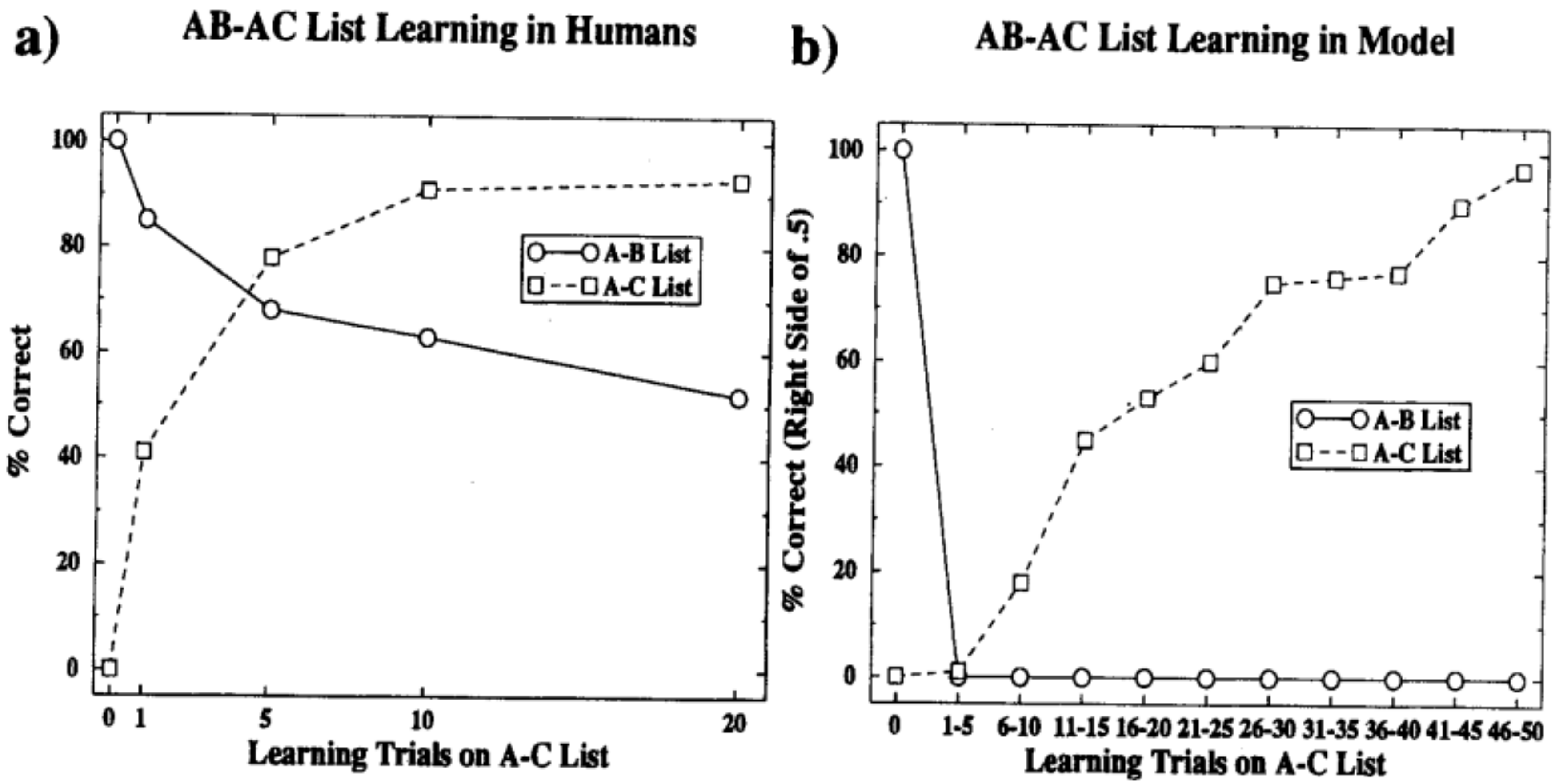


Figure 10. Panel a: Experimental data showing mild interference in humans in the *AB–AC* paradigm (Barnes & Underwood, 1959). Panel b: Simulation results demonstrating catastrophic interference. *Note.* From "Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem," pp. 125 and 129, by M. McCloskey and N. J. Cohen, in *The Psychology of Learning and Motivation,* edited by G. H. Bower, 1989, New York: Academic Press. Copyright 1989 by Academic Press, Inc. Reprinted with permission.

One solution: Interleaved training

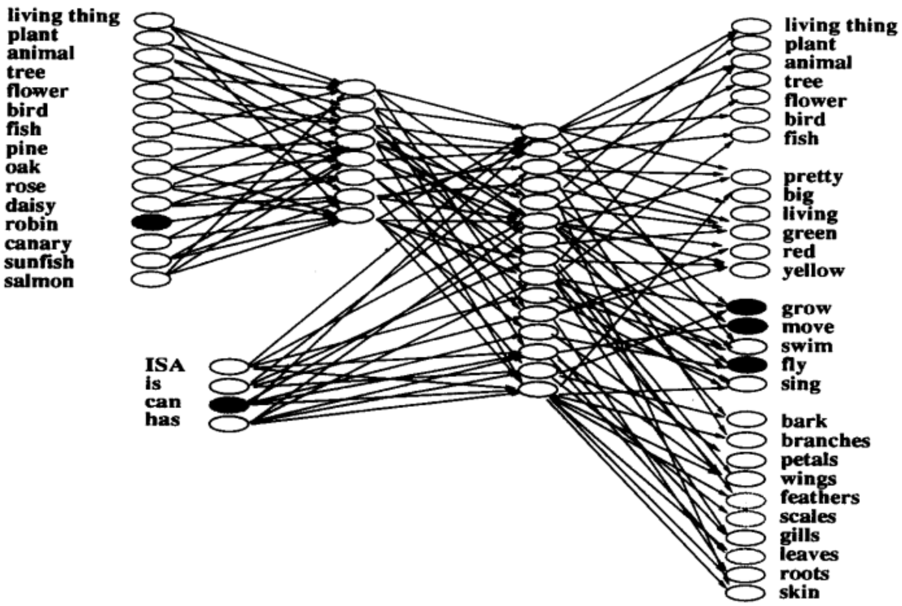


Figure 5. Depiction of the connectionist network used by Rumelhart (1990) to learn propositions about the concepts shown in Figure 4. The entire set of units used in the actual network is shown. Inputs are presented on the left, and activation propagates from left to right. In instances in which connections are indicated, every unit in the pool on the left (sending) side projects to every unit on the right (receiving) side. An input consists of a concept–relation pair; the input *robin can* is illustrated by a darkening of the active input units. The network is trained to turn on all of those output units that represent correct completions of the input pattern. In this case, the correct units to activate are *grow, move,* and *fly;* the units for these outputs are darkened as well. Subsequent analysis focuses on the concept representation units, the group of eight units to the right of the concept input units. Based on the network used by Rumelhart and Todd

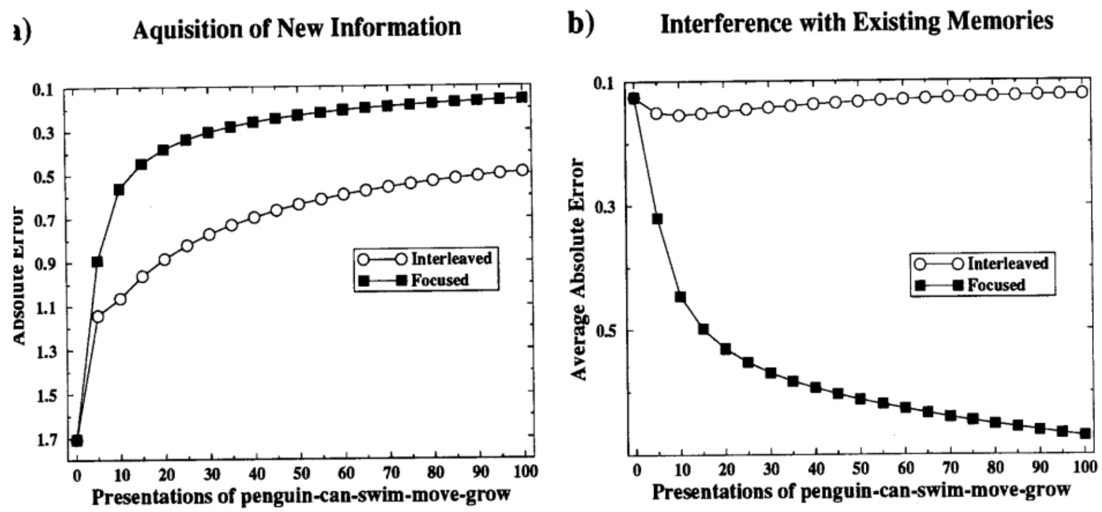


Figure 11. Effects of focused and interleaved learning on the acquisition of new knowledge and on interference with existing knowledge. Simulations were carried out with Rumelhart's (1990) network; the connection weights resulting from the initial 500 epochs of training with the base corpus were used. The performance measure, absolute error, is defined as the sum across output units of the absolute value of the difference between the correct response for each pattern and the actual response. The measure reaches its optimal value of 0 when the output exactly matches the target. Better performance corresponds to lower error, and the axis is inverted for better visual correspondence to standard memory performance curves. In the analysis of interference with other memories, the performance measure is the average of the absolute error over all 15 of the cases in the initial training corpus involving the *can* relation. The scales of each graph are different and are set to encompass the range of values spanned in each case. The interference is much greater for some items than for others and falls predominantly on those output units in which the correct answer for the preexisting memory differs from the correct answer for the penguin.

Key points

- Learning the complex structure of a domain (e.g., semantic knowledge) occurs across life and probably takes the form of gradual, interleaved learning

- Learning new information rapidly in a network that has already learned can lead to catastrophic interference

- Incorporation of new material without interference can occur if the new material is incorporated gradually, interleaved with ongoing exposure to examples of the domain that has already been learned.

Key points

- Learning the complex structure of a domain (e.g., semantic knowledge) occurs across life and probably takes the form of gradual, interleaved learning
- Learning new information rapidly in a network that has already learned can lead to catastrophic interference
- Incorporation of new material without interference can occur if the new material is incorporated gradually, interleaved with ongoing exposure to examples of the domain that has already been learned.

so…

- Hippocampal system might provide a "medium for the initial storage of memories in a form that avoids interference"
- Might facilitate slow interleaving of new information with old to enable abstraction of structure (e.g., semantic memory and consolidation)

In other words

Fast and slow learning are at odds (in neural networks type systems, not generally) and thus biological system built on such principles might need more than one type of learning/ memory

Sparse coding

- As discussed, overall activity in hippocampus seems low compared to nearby areas like EC
- Suggestive of a sparse code which is helpful for *pattern separation*
- But how do you get something out of it?

Pattern separation
high capacity, no overlap

Invertibility
how do you recover what was encoded? sparse=information loss

Stability
memories need to remain of extended periods

Some options

5x5x9 = 405 (112 on - 28%)

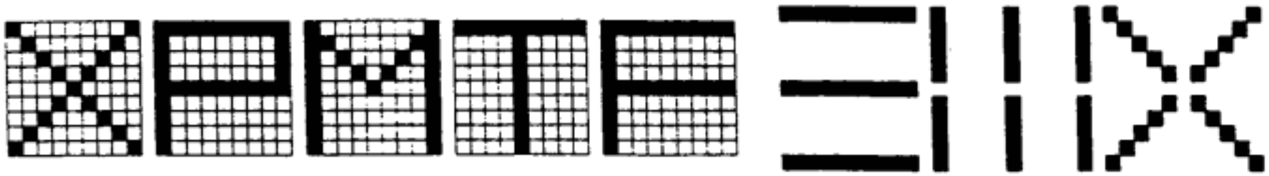


FIGURE 1. Five letters on a 5 × 9 × 9 pixel array, and a set of strokes that can provide a more efficient encoding.

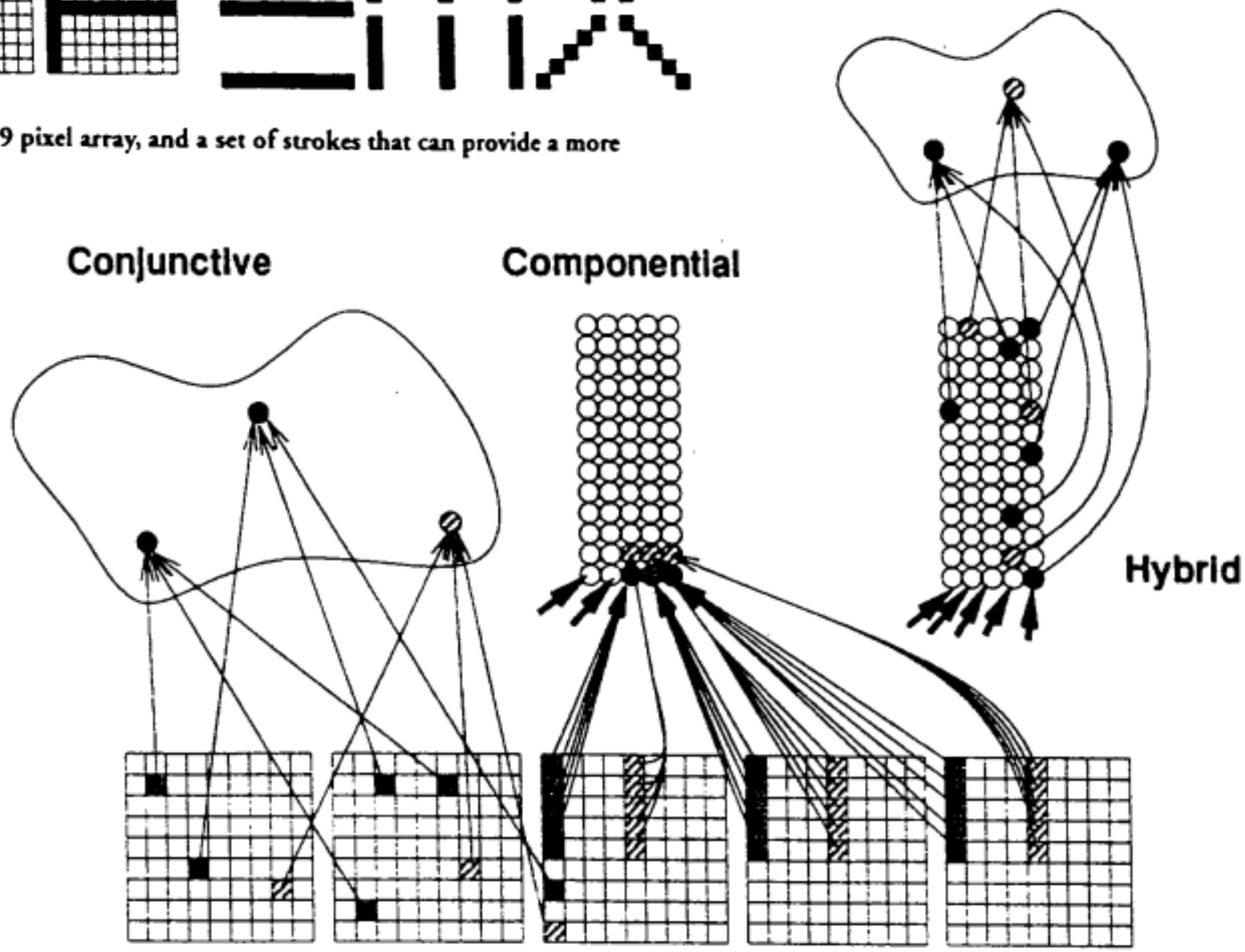


FIGURE 2. Two coding schemes and their hybrid combination.

Some options

5x5x9 = 405 (112 on - 28%)

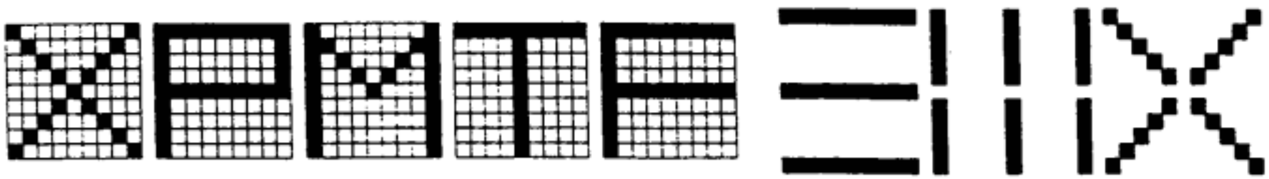


FIGURE 1. Five letters on a 5 × 9 × 9 pixel array, and a set of strokes that can provide a more efficient encoding.

Random, conjunctive coding

Random triplets -
10 million neurons
2% active to encode letters
Very low overlap

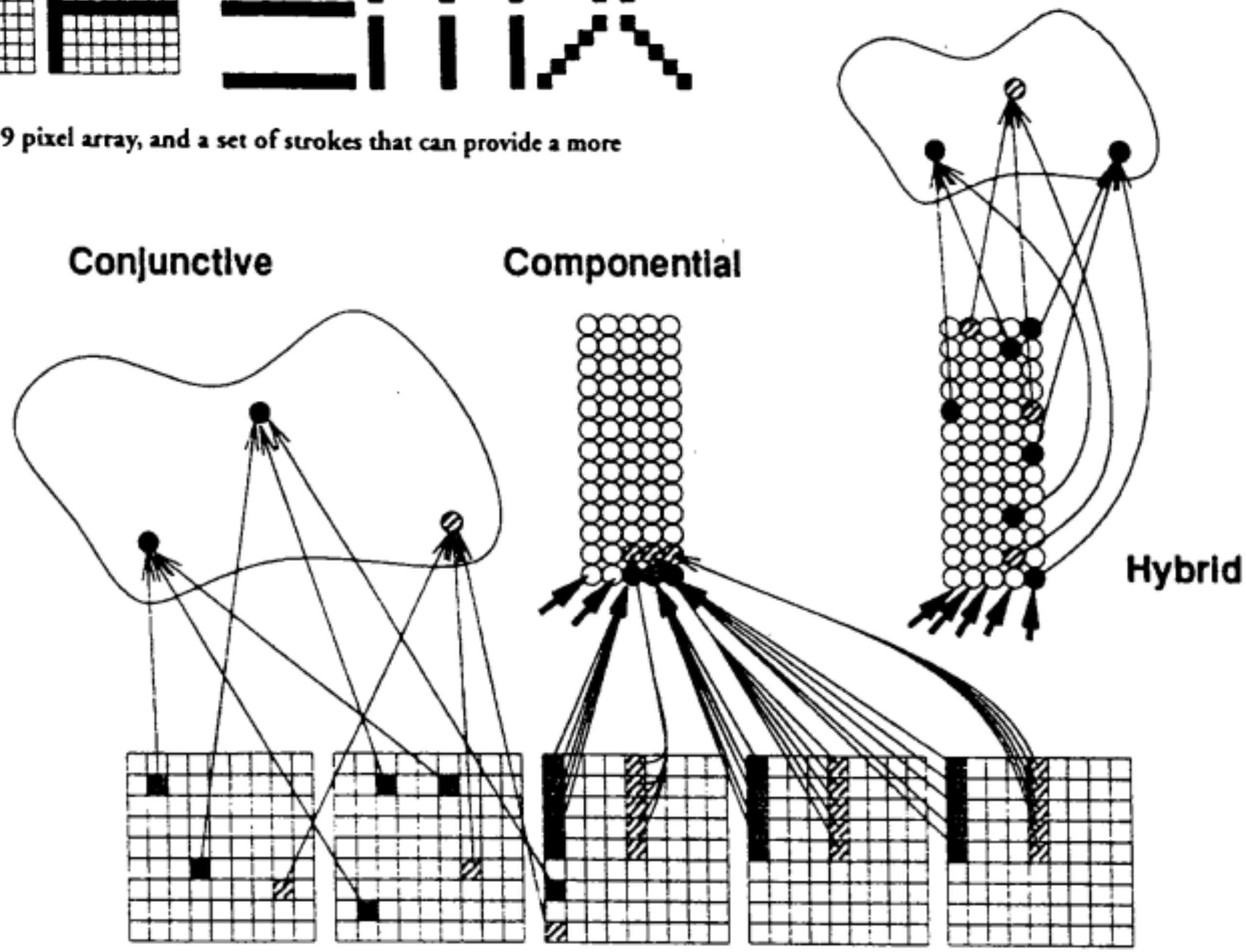


FIGURE 2. Two coding schemes and their hybrid combination.

Some options

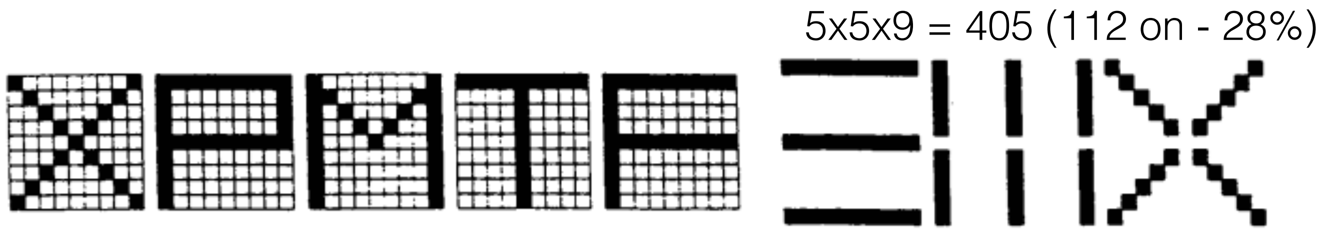


FIGURE 1. Five letters on a 5 × 9 × 9 pixel array, and a set of strokes that can provide a more efficient encoding.

Componential coding

Recurring/frequent patterns are "chunked" into useful components

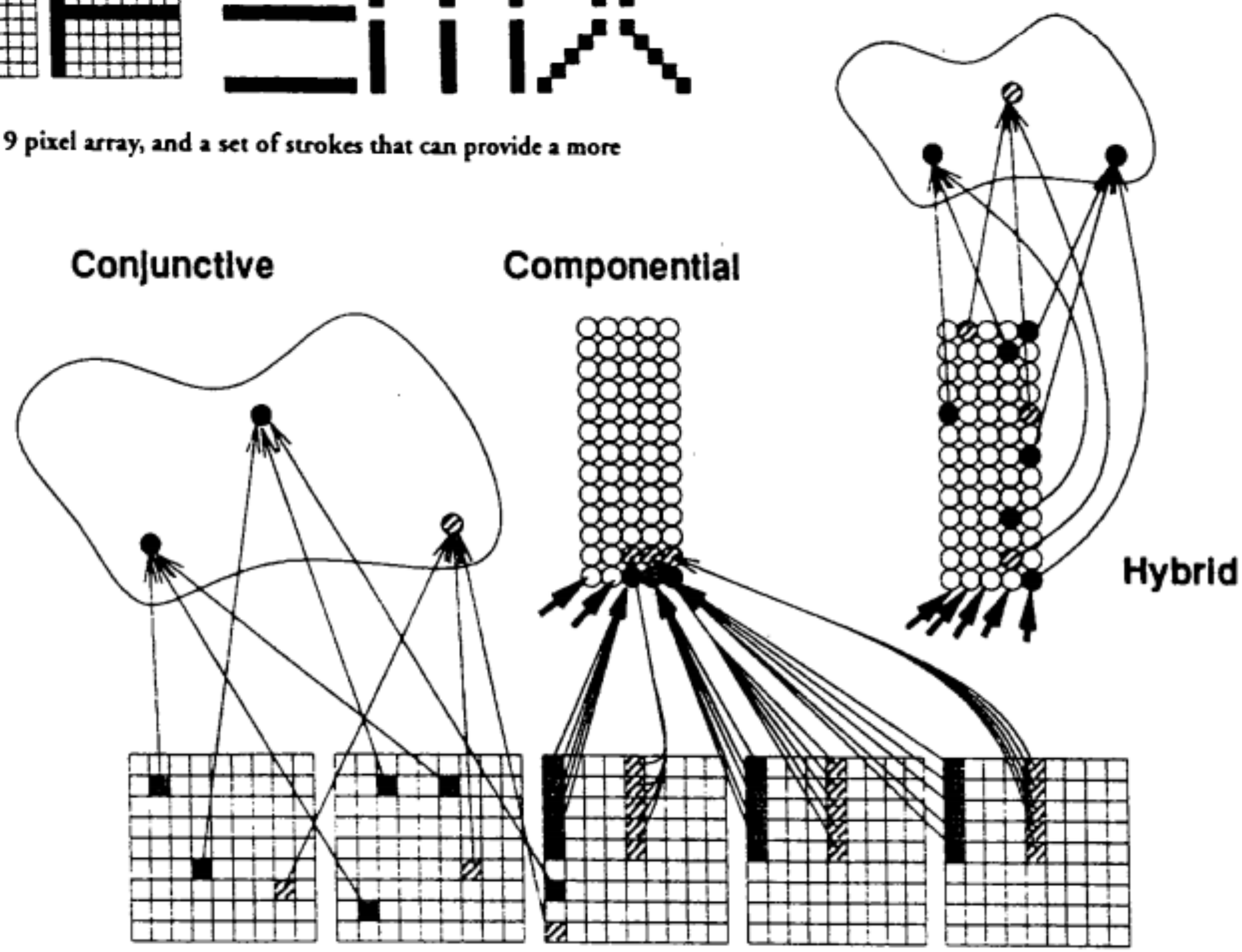


FIGURE 2. Two coding schemes and their hybrid combination.

Some options

5x5x9 = 405 (112 on - 28%)

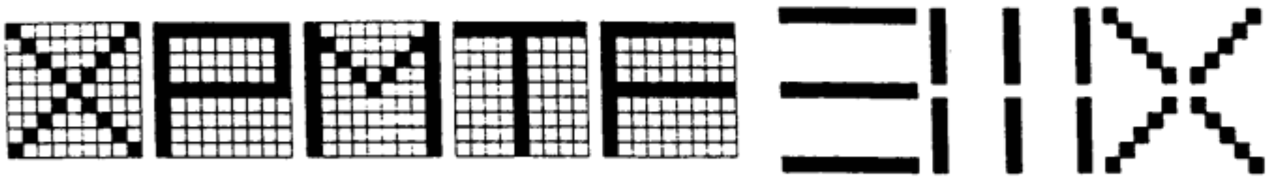


FIGURE 1. Five letters on a 5 × 9 × 9 pixel array, and a set of strokes that can provide a more efficient encoding.

Hybrid

Random conjunctive coding of components.

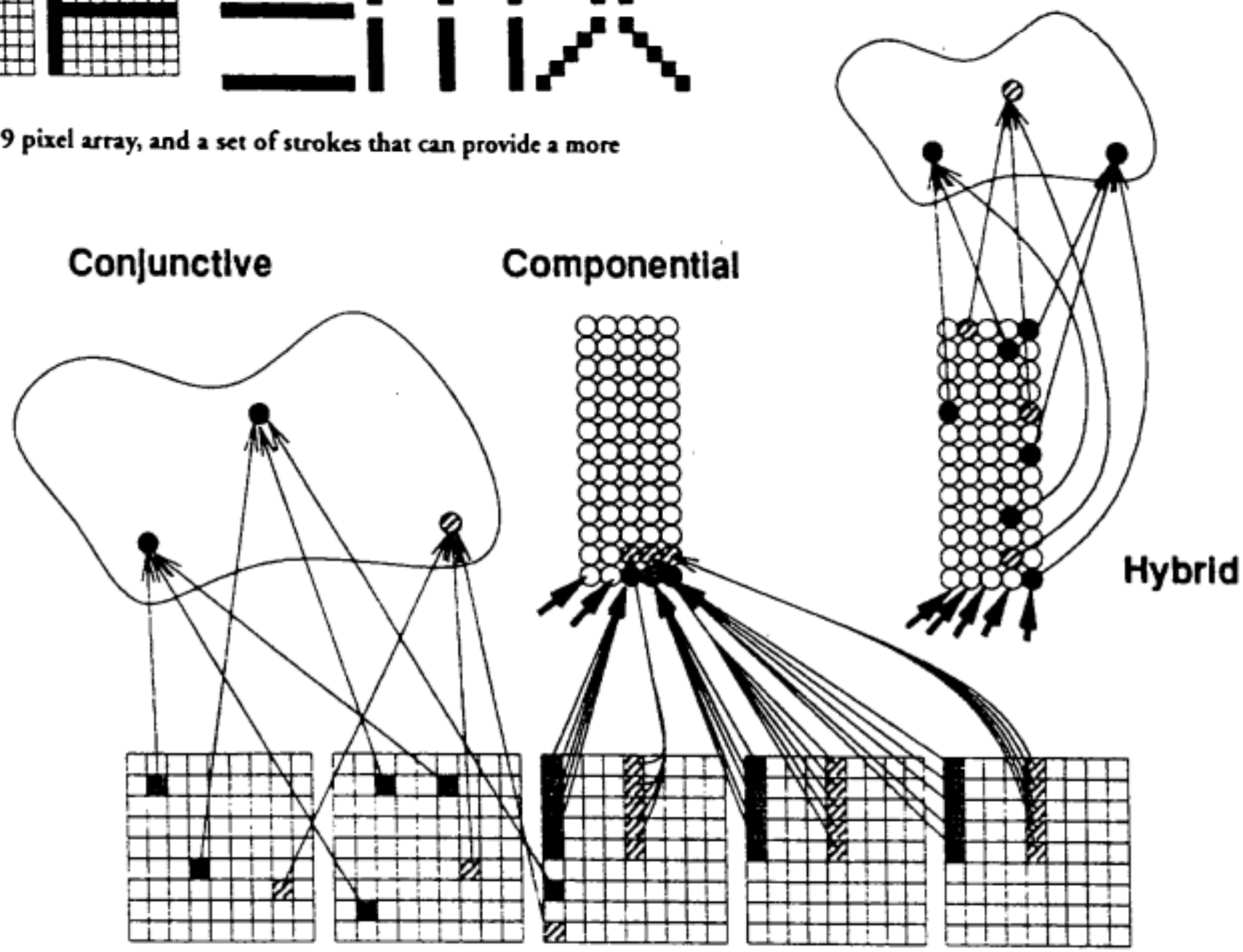
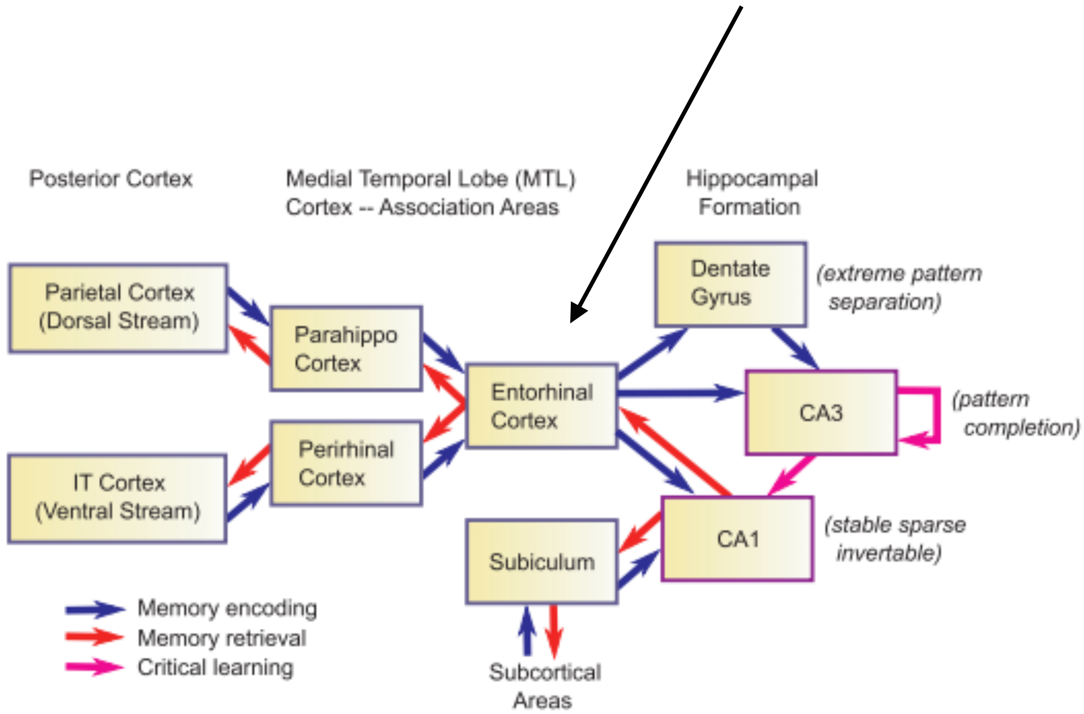


FIGURE 2. Two coding schemes and their hybrid combination.

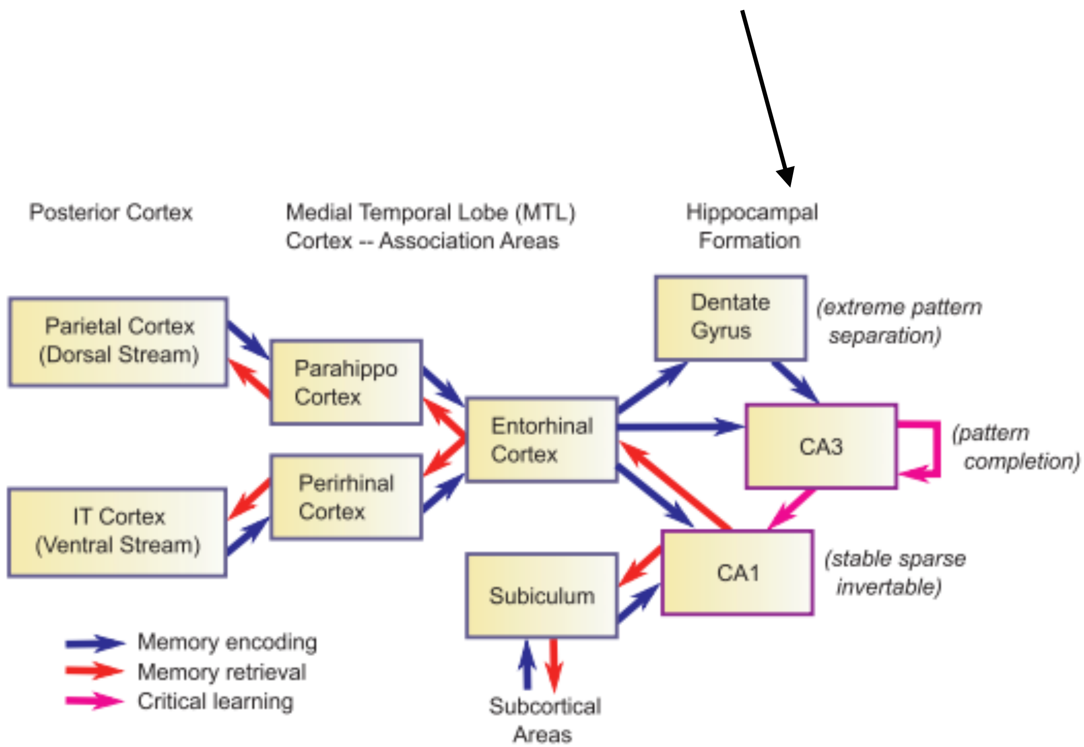
Computational roles

Compressed, componential representation

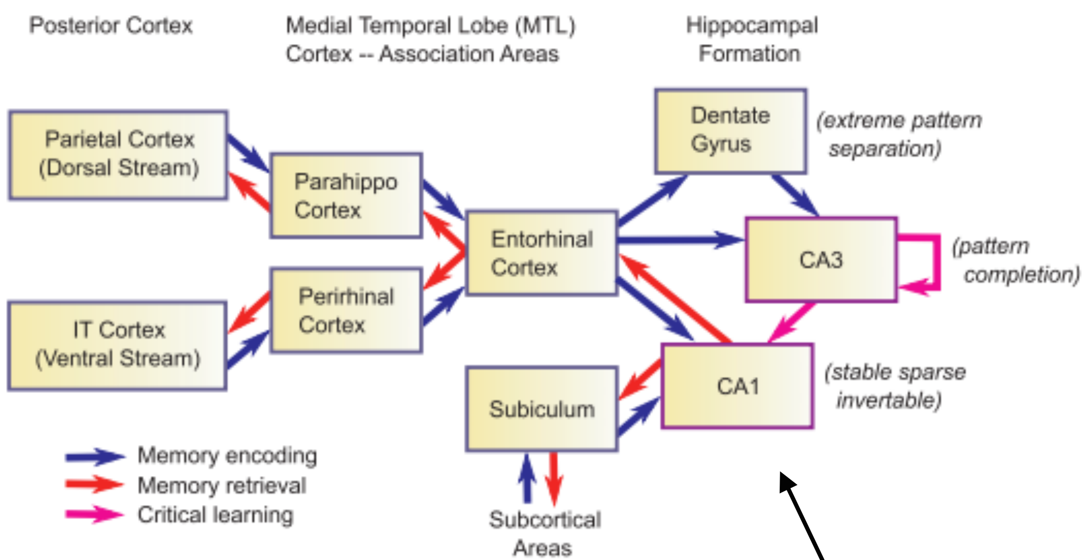


Computational roles

Random conjunctive coding



Computational roles



Invertible decoder??