

SANDIA REPORT

SAND2009-6164

Unlimited Release

Printed October 2009

Modeling Aspects of Human Memory for Scientific Study

Michael Bernard, J. Dan Morrow, Shawn Taylor, Stephen Verzi, Craig Vineyard, Thomas Caudell, Neal Cohen, Howard Eichenbaum, Mark McDaniel, & Patrick Watson

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.



Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: <http://www.osti.gov/bridge>

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



Modeling Aspects of Human Memory for Scientific Study

Michael Bernard¹, J. Dan Morrow¹, Shawn Taylor², Stephen Verzi³ & Craig Vineyard¹

¹Cognitive & Optical Military Systems

³Cognitive System Research & Applications

²Science & Engineering Information Systems

Sandia National Laboratories

P.O. Box 5800

Albuquerque, New Mexico 87185-MS1188

Thomas Caudell¹, Neal Cohen², Howard Eichenbaum³, Mark McDaniel⁴, & Patrick Watson²

³Boston University, ²University of Illinois—Champaign-Urbana Beckman Institute,

¹University of New Mexico, & ⁴Washington University

Abstract

Working with leading experts in the field of cognitive neuroscience and computational intelligence, SNL has developed a computational architecture that represents neurocognitive mechanisms associated with how humans remember experiences in their past. The architecture represents how knowledge is organized and updated through information from individual experiences (episodes) via the cortical-hippocampal declarative memory system. We compared the simulated behavioral characteristics with those of humans measured under well established experimental standards, controlling for unmodeled aspects of human processing, such as perception. We used this knowledge to create robust simulations of & human memory behaviors that should help move the scientific community closer to understanding how humans remember information. These behaviors were experimentally validated against actual human subjects, which was published. An important outcome of the validation process will be the joining of specific experimental testing procedures from the field of neuroscience with computational representations from the field of cognitive modeling and simulation.

ACKNOWLEDGMENTS

We would like to thank Tom Caudell, Neal Cohen, Howard Eichenbaum, Mark McDaniel, and Patrick Watson for their invaluable contribution to both the development and the testing of the SNL computational model of episodic memory and recall.

CONTENTS

1. Introduction.....	9
1.1. Overview of the Problem and Idea	9
2. The Declarative Memory System	11
2.1 What is recollection?.....	11
2.2 The anatomy of memory	19
2.3 Towards a functional organization of a cortical-hippocampal memory system.....	21
3. Current Computational Representation.....	24
3.2. General description of architecture.....	24
3.1.1 The hippocampus	26
4. Assessment of The Model	35
4.1 Temporal and sequential memory and recall of objects	35
4.1.1 Experimental results.....	38
4.2. Associating object/scenes pairs.....	41
4.2.1 Results	42
4.2.2 Repeated experiment	43
4.3 Co-occurrence of shared scenes with novel objects	43
4.3.1 Qualitative comparison results	44
4.3.2 Quantitative comparison results	46
5. General Discussion and Conclusions.....	50
5.1 Model limitations	51
6. References.....	53
Appendix A: Layout and Connection Structure of The Model.....	63
Appendix B: The Input Set Presented to The Model.....	64
Distribution.....	65

FIGURES

Figure 1. The hippocampal and parahippocampal regions within the brain.....	12
Figure 2. The Hippocampal Structures of the Brain.....	22
Figure 3. A conceptual view of the computational architecture.....	25
Figure 4. An example of an Adaptive Resonance Theory (ART) module.....	26
Figure 5. The modeled representation of the hippocampal system.....	27
Figure 6. Step-by-step process of episodic/semantic activation.....	32
Figure 7. The bottom-up/top-down flow of information.....	33
Figure 8. Time series of temporal integrator outputs.....	36
Figure 9. ART with temporal integrator.....	37
Figure 10. Example temporal integrator outputs.....	37
Figure 11. Memory and recall input sequence.....	39
Figure 12. A sample memory and recall output sequence.....	39
Figure 13. Hannula image scheme.....	41
Figure 14. Experiment outputs.....	42
Figure 15. Face and scene pairs.....	44
Figure 16. Person 'A' with house 'A'.....	45
Figure 17. Person 'B' with house 'A'.....	45
Figure 18. Person 'C' with house 'C' produces distinct mapping of activation.....	46
Figure 19. Model comparison to Preston study.....	47
Figure 20. Faces paired with different contexts.....	48
Figure 21. Experimental and model image of CA3 hippocampal activation.....	49

NOMENCLATURE

ART	Adaptive Resonance Theory
CA1	Cornu Ammonis area 1
CA3	Cornu Ammonis area 3
DG	Dentate Gyrus
EC	Entorhinal Cortex
DOE	Department of Energy
IT	InferoTemporal Cortex
LAPART	LATERALLY Primed Adaptive Resonance Theory
MTL	Medial Temporal Lobe
ROC	Receiver Operating Characteristic
SNL	Sandia National Laboratories
SOM	Self-Organizing Map-Adaptive Resonance Theory
SOMART	Self-Organizing Map-Adaptive Resonance Theory
TIART	Temporally Integrated Adaptive Resonance Theory
WTA	Winner-Take-All

1. INTRODUCTION

Cognitive neuroscience research has found that the hippocampus plays a central role in forming and temporarily storing representations of personal experiences. These representations are later migrated to widespread areas of the cerebral cortex, which are then permanently stored. The focus of the Laboratory Directed Research and Development (LDRD) work was to produce a computational model that: (1) represents the fundamental features of hippocampus-dependent relational processing and (2) tests this representation against human memory by comparing the performance of normal humans subjects and people lacking normal hippocampal function (amnesic subjects) with the performance of the full model and the model without a functional “hippocampus” on the same memory tasks. Success of the model was measured in terms of similar performance compared to that of humans with and without the contribution of hippocampal processing, as a function of experimental parameters and controls. The intent was to develop a falsifiable model (i.e., one where we can find and understand its limitations and characterize them properly), where the trend of results from human experimentation match the trends of simulation. A second focus, modeling aspects of human reasoning, was dropped after the first fiscal year due to cutbacks in the overall funding of this project. As such, this technical report will only discuss the modeling of declarative memory.

1.1. Overview of the Problem and Idea

Memory is usually thought of as a passive record of past events and acquired factual knowledge. But our adaptive application of memory is to make plans for our future actions. Therefore, our conscious lives are dominated by interactions between retrospective memory, the capacity for recollection of general knowledge and one’s personal history of previous actions and their outcomes, and prospective memory, our intentional application of knowledge, and history in directing ongoing decisions and behavior. This project is currently modeling of how the brain accomplishes retrospective recollection and memory. Our capacity for recollection is known to be supported by a system composed of several cortical association areas interacting with structures in the medial temporal lobe, and in particular, the hippocampus. There is a general consensus that the cortex is the repository of detailed representations of perceptions and thoughts and that the hippocampus supports the ability to bind together cortical representations and, when cued by part of a previous representation, to reactivate the full set of cortical representations that compose a recollective, declarative (explicit) memory.

To date, computational models have not neurocognitively represented episodic recollection memory within an embodied, simulation environment. This creates several limitations regarding the plausibility of current models. First, current approaches do not dynamically collect “what,” “where,” and “when” perceptual information to produce an episodic memory trace. Second, current approaches typically create a false distinction between semantic and event-based, episodic memory. While semantic memory has a different phenomenology than episodic memory, there is strong evidence they are part of the same system (McKoon et al., 1986).

Research has found that the hippocampus plays a central role in forming and temporarily storing representations of personal experiences. These representations are later migrated to widespread areas of the cerebral cortex, which are then permanently stored.

To address the need for more plausibility model Sandia National Laboratories (SNL) has (1) produced a computational model that represents the fundamental features of hippocampus-dependent relational processing and (2) tested this representation against human memory by comparing the performance of normal humans subjects and people lacking normal hippocampal function with the performance of the full model and the model without a functional “hippocampus” on the same memory tasks. Success of the model was measured in terms of similar performance compared to that of humans with and without the contribution of hippocampal processing, as a function of experimental parameters and controls.

This effort is extending the current Sandia Cognitive Framework by incorporating a representation of memory processing, focusing on hippocampus and neocortical systems described in current complimentary learning systems theory (i.e., cortical-hippocampal theory of declarative memory, Eichenbaum, 2007). The model also specifies how hippocampal and cortical representations interact at multiple levels of abstraction to support the interleaving of new information within the cerebral cortex. For example, we integrated the perceptual features of relational memory processing into our computational model. This project produced two main products: (1) a neuro-cognitive computational architecture that represents episodic memory and (2) major review paper(s) submitted for publication. This work extended current computational models (for example, McClelland et al.; *Psych Rev.* 1995) wherein a pre-existing knowledge structure in cortical areas is challenged to incorporate new information within an existing network. To accomplish this goal we collaborated with leading experts from academia. Specifically, the external research team consisted of: (1) Howard Eichenbaum, professor of psychology and neuroscience and Director of the Center for Memory and Brain at Boston University, (2) Neal Cohen, professor of psychology and neuroscience at the Beckman Institute of the University of Illinois, (3) Thomas Caudell, professor and Director of the Center for High Performance Computing at University of New Mexico, and (3) Mark McDaniel, professor of psychology at Washington University.

2. THE DECLARATIVE MEMORY SYSTEM

Like the Roman god Janus, memory looks both into the past and the future. Memory is usually thought of as a passive record of past events and acquired factual knowledge. But our adaptive application of memory is to make plans for our future actions. Therefore, our conscious lives are dominated by interactions between *retrospective memory*, the capacity for recollection of general knowledge and one's personal history of previous actions and their outcomes, and *prospective memory*, our intentional application of knowledge and history in directing ongoing decisions and behavior. The discussion will begin by outlining the experimental evidence on the cognitive and neural mechanisms of recollection, and then consider retrospective memory from experimental studies in cognitive science. The paper will then outline a formal model and its implementation in software.

2.1 What is recollection?

We have all been in the situation where we meet someone who seems highly familiar but we cannot recall who they are or why we know them. Sometimes, we just give up and say, "Don't I know you?" Alternatively, when a clue or sufficient mental searching helps us retrieve a wealth of information all at once, including the name, where we met before, and the circumstances of the meeting. Considerable current research on recollection has focused the distinction between a vivid recollection the lesser condition of a sense of familiarity with a particular person or object. Familiarity comes rapidly and reflects the strength match between a cue and a stored memory template. It is an isolated ability to identify a person or object as previously experienced. Recollection is typically slower and measured by the number of qualitative associations retrieved and the organization of the memory retrieved. Thus, recollections typically include not only the item sought in memory but also three other kinds of additional information:

- (1) a spatial and temporal context of the experience in which the item was previously encountered
- (2) a replay of the sequence of events that compose an entire episode with that item
- (3) and remembering additional related experiences with the item.

Furthermore, one brain area, the hippocampus, is critically involved in each of these aspects of recollection. Yonelinas et al. (2002) ROC analysis on recognition memory performance to show that mild hypoxia that causes damage largely confined to the hippocampus resulted in a severe deficit in recollection but normal familiarity. A similar pattern of deficient recollection and preserved familiarity was reported in a patient with relatively selective hippocampal atrophy related to meningitis (Aggleton et al., 2005). Further consideration of the three properties of introduced above provides insights into both the fundamental elements of recollection and the role of the hippocampus in memory processing (see Figure 1).

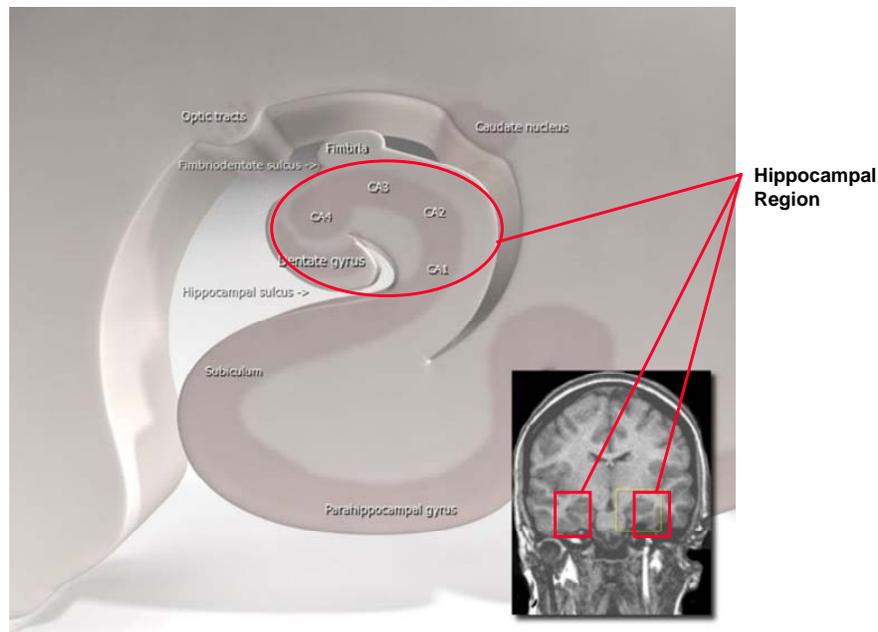


Figure 1. The hippocampal and parahippocampal regions within the brain

Events are represented as items in the context in which they were experienced. A fundamental feature of recollection is memory for the spatial, temporal, and associational context in which experiences occur. Functional imaging studies support the notion that the hippocampus is activated during the encoding or retrieval of associations among many elements of a memory, a characteristic of context-rich episodic memories (for review see Cohen et al., 1999; Eldridge et al., 2000; Maguire, 2001; Addis et al., 2004). For example, Henke et al., (1997) observed greater hippocampal activation when subjects associated a person with a house, as compared to making independent judgments about the person and house and others have found selective hippocampal activation during recollection of the context of learning in formal tests of memory (e.g. Davachi et al., 2003; Ranganath et al., 2003). The coding of associations extends beyond item and context associations such that the hippocampus is also selectively activated during the encoding or retrieval of verbal (Davachi & Wagner, 2002; Giovanello et al., 2003a) and face-name associations (Small et al., 2001; Zeineh et al., 2003; Sperling et al., 2003). Correspondingly, recent neuropsychological studies have found that recognition of associations is impaired even when recognition for single items is spared in amnesic patients (Giovanello et al., 2003, Turriziani et al., 2004). These studies reported impairment in recognition memory for associations between words or between faces or face-occupations pairs, as compared to normal performance in recognition of single items. At the same time, other functional imaging studies and characterizations of amnesia have suggested that the hippocampus is sometimes involved in both associative and single item recognition, highlighting the need to clarify the nature of associative information that composes an “event” (Squire et al., 2004). Nevertheless, these findings are generally consistent with the notion that the hippocampus plays a distinct role in

recollection associated with binding features of items and their context to represent salient events (Eichenbaum et al., 2007).

Studies that employ animal models can provide compelling evidence on the effects of selective hippocampal damage. Several studies have shown that damage limited to the hippocampus results in deficits in forming a memory for the context or location where items were once experienced (reviewed in Mumby, 2001). In one recent study, rats were initially exposed to two objects in particular places in one of two environmental chambers (Mumby et al., 2002). In subsequent recognition testing, the place of the object or the context was changed. Normal rats increased their exploration of objects that were moved to new places or put in novel contexts. By contrast, rats with hippocampal damage failed to recognize objects when either the place or context was changed (see also Eacott & Norman, 2004).

Several investigators have argued that animals are indeed capable of remembering the temporal as well as spatial context in which they experienced specific stimuli (Clayton et al., 2003; Day et al., 2003). To further explore these aspects of episodic memory, Eichenbaum (a team member of this LDRD) developed a task that assesses memory for events from a series of events that each involve the combination of an odor (“what”), the place in which it was experienced (“where”), and the order in which the presentations occurred (“when”; Ergorul & Eichenbaum, 2004). On each of a series of events, rats sampled an odor in a unique place along the periphery of a large open field. Then, memory for the when those events occurred was tested by presenting a choice between an arbitrarily selected pair of the odor cups in their original locations. Normal rats initially employed their memory of the places of presented cups and approached the location of the earlier experience. Then they confirmed the presence of the correct odor in that location. Animals with selective hippocampal damage fail on both aspects of this task even though their memory for independent features of location and odor items was intact. These findings indicate that the hippocampus is critical for effectively combining the “what”, “when”, and “where” qualities of each experience to compose the retrieved memory.

Studies on the firing properties of single neurons in animals provide insights into the nature of neural population representations in the hippocampus. There is a large body of evidence that hippocampal neurons encode associations among stimuli, actions, and places that compose discrete events. Many studies have shown that hippocampal neurons encode an animal’s location within its environment, and some view this as the principle function of hippocampal populations (Muller et al., 1999; Best et al., 2001). However, many other studies have shown that hippocampal neurons also fire associated with the ongoing behavior and the context of events as well as the animal’s location (Eichenbaum et al., 1999). In the most direct examination of this issue, Wood et al (1999) directly compared spatial and non-spatial coding by hippocampal neurons by training animals to perform the same memory judgments at many locations in the environment. A large subset of hippocampal neurons fired only associated with a particular combination of the odor, the place where it was sampled, and the match-non-match status of the odor. In a similar study on the coding properties of hippocampal neurons in humans, Ekstrom et al. (2003) recorded in subjects as they played a taxi driver game, searching for passengers picked up and dropped off at various locations in a virtual reality town. They observed

that many of these cells fired selectively associated with specific combinations of a place and the view of a particular scene or a particular goal. These and other studies indicate that, in rats, monkey, and humans, a prevalent property of hippocampal firing patterns involves the representation of unique associations of stimuli, their significance, specific behaviors, and the places where these events occur (see Eichenbaum et al., 2004).

Episodes are represented as sequences of events. We live our lives through personal experience, and our initial construction of reality within consciousness is a form of episodic buffer that contains a representation of the stream of events as they just occurred (Baddeley, 2000). In an early characterization of episodic recollection, Tulving (1983) distinguished episodic memory as organized in the temporal dimension, and contrasted this scheme with a conceptual organization of semantic memory. Tulving (1983) argued that the central organizing feature of episodic memory is that “one event precedes, co-occurs, or follows another.” This is reminiscent of Aristotle’s (350BC) characterization of vivid remembering: “Acts of recollection, as they occur in experience, are due to the fact that one thought has by nature another that succeeds it in regular order.” These characterizations emphasize the temporal organization of episodic memories.

The order of events within human memory depends on hippocampal function. In a study using a design similar to that described above, Hopkins et al. (1995) found that patients with hypoxic brain injury involving in shrinkage of the hippocampus are impaired in memory for the order of a series of 6 words, pictures, or spatial locations. These patients were, however, also impaired in recognition of the items, undermining an unambiguous interpretation of a deficit in the order of the events independent of memory for the events. More recently, Spiers et al. (2001) reported a selective deficit in order memory independent of item memory in a patient with selective hippocampal damage due to perinatal transient anoxia (Vargha-Khadem et al., 1997). In this study the patient explored a virtual reality town in which he received objects from virtual characters. His recognition of the familiar objects was intact, but he was severely impaired in memory for the order in which he received objects, as well as for where he received them. Also, Downes et al. (2002) reported that patients with medial temporal lobe damage that included bilateral hippocampal damage were impaired in memory for the order of presentation of words for which recognition of the items was equivalent. Also, evidence from the deferred imitation task, where subjects are required to remember an action sequence, indicate a critical role for the hippocampus (McDonough et al., 1995; Adlam et al., 2005). Thus, humans with hippocampal damage are impaired in memory for the order of events in unique episodes even in cases where recognition memory is intact.

Studies on animals also show that the representation of memories by the hippocampus incorporates not only items that must be remembered, but also the events that precede and follow. For example, Honey et al. (1998) provided a simple demonstration of the importance of temporal order in hippocampal processing, reporting that hippocampal lesions disrupted animals’ normal orienting response when a pair of stimuli are presented in the opposite order of previous exposures. The specific role of the hippocampus in remembering the order of a series of events in unique experiences has been explored using a behavioral protocol that assesses memory for episodes composed of a unique sequence of olfactory stimuli (Fortin et al., 2002; see also Kesner et al., 2002). Memory for the sequential order of odor events was directly compared with recognition of the

odors in the list independent of memory for their order. On each trial rats were presented with a series of five odors, selected randomly from a large pool of common household scents. Memory for each series was subsequently probed using a choice test where the animal was reinforced for selecting the earlier of two of the odors that had appeared in the series. In later sessions we also tested whether the rats could identify the odors in the list independent of their order, by rewarding the selection of a novel odor against one that had appeared in the series. Normal rats performed both tasks well. Rats with hippocampal lesions could recognize items that had appeared in the series but were severely impaired in judging their sequential order.

How do hippocampal neuronal populations represent the sequences of events that compose distinct episodes? A common observation across many different behavioral protocols is that different hippocampal neurons become activated during every event that composes each experience, including during simple behaviors such as foraging for food (e.g., Muller et al., 1987) as well as learning related behaviors directed at relevant stimuli that have to be remembered in studies that involve classical conditioning, discrimination learning, and non-matching or matching to sample tasks to tests and a variety of maze tasks (e.g. Hampson et al., 1993; for review, see Eichenbaum et al, 1999). In each of these paradigms, animals are repeatedly presented with specific stimuli and rewards, and execute appropriate cognitive judgments and conditioned behaviors. Corresponding to each of these regular events, many hippocampal cells show time-locked activations associated with each sequential event. Also, as described above, many of these cells show striking specificities corresponding to particular combinations of stimuli, behaviors, and the spatial location of the event. Thus, hippocampal population activity can be characterized as a sequence of firings representing the step-by-step events in each behavioral episode.

Furthermore, these sequential codings can be envisioned to represent the series of events and their places that compose a meaningful episode, and the information contained in these representations distinguishes related episodes that share common events and therefore could be confused. Recent studies on the spatial firing patterns of hippocampal neurons as animals traverse different routes that share overlapping locations provide compelling data consistent with this characterization. In one study, rats were trained on the classic spatial alternation task in a modified T-maze (Wood et al., 2000; see also Frank et al., 2000; Ferbinteanu and Shapiro (2003). Performance on this task requires that the animal distinguish left-turn and right-turn episodes that overlap for a common segment of the maze and requires the animal to remember the immediately preceding episode to guide the choice on the current trial, and in that way, the task is similar in demands to those of episodic memory. If hippocampal neurons encode each sequential behavioral event and its locus within one type of episode, then most cells should fire only when the rat is performing within either the left-turn or the right-turn type of episode. This should be particularly evident when the rat is on the “stem” of the maze, when the rat traverses the same set of locations on both types of trials. Indeed, a large proportion of cells that fired when the rat was on the maze stem fired differentially on left-turn versus right-turn trials. The majority of cells showed strong selectivity, some firing at over ten times the rate on one trial type, suggesting they were part of the representations of only one type of episode. Other cells fired substantially on both trial types, potentially

providing a link between left-turn and right-turn representations by the common places traversed on both trial types.

Functional imaging studies in humans have also revealed hippocampal involvement in both spatial and non-spatial sequence representation. Several studies have shown that the hippocampus is active when people recall routes between specific start points and goals, but not when subjects merely follow a set of cues through space (Hartley et al. 2003). In addition, the hippocampus is selectively activated when people learn sequences of pictures (Kumaran & Maguire, 2006). Even greater hippocampal activation is observed when subjects must disambiguate picture sequences that overlap, parallel to the findings on hippocampal cells that disambiguate spatial sequences (Wood et al., 2000).

Memories are networked to support inferential memory expression. Further consideration of the cognitive properties of episodic memory suggest that related episodic representations might be integrated with one another to support semantic memory and the ability to generalize and make inferences from memories. Referring to how related memories are integrated with one another, William James (1890) emphasized that "...in mental terms, the more other facts a fact is associated with in the mind, the better possession of it our memory retains. Each of its associates becomes a hook to which it hangs, a means by which to fish it up by when sunk beneath the surface. Together they form a network of attachments by which it is woven into the entire tissue of our thought." James envisioned memory as a systematic organization of information wherein the usefulness of memories was determined by how well they are linked together.

There are two main outcomes of the linking of representations of specific experiences. One is a common base of associations that are not dependent on the episodic context in which the information was acquired. Thus when several experiences share considerable common information, the overlapping elements and common links among them will be reinforced, such that those items and associations become general regularities. The representation of these general regularities constitutes semantic "knowledge" that is not bound to the particular episode or context in which the information was encoded. The networking of episodic memories by common elements provides a mechanism for the commonly (albeit not universally, see Tulving, 2002) held view that semantic knowledge is derived from information repeated within and abstracted from episodic memories.

There is considerable evidence that hippocampal neurons indeed extract the common features among related episodes. In all the studies described above, a subset of hippocampal neurons encode features that are common among different experiences – these representations could provide links between distinct memories. For example, in the Wood et al. (1999) study on odor recognition memory, whereas some cells showed striking associative coding of odors, their match/non-match status, and places, other cells fired associated with one of those features across different trials. Some cells fired during a particular phase of the approach towards any stimulus cup. Others fired differentially as the rat sampled a particular odor, regardless of its location or match-non-match status. Other cells fired only when the rat sampled the odor at a particular place, regardless of the odor or its status. Yet other cells fired differentially associated with the match and nonmatch status of the odor, regardless of the odor or where it was sampled. Similarly, in Ekstrom and colleagues' (2003) study on humans performing a virtual navigation task, whereas some hippocampal neurons fired associated with combinations of views, goals,

and places, other cells fired when subjects viewed particular scenes, occupied particular locations, or had particular goals in findings passengers or locations for drop off. In studies that have recorded hippocampal neuronal activity as rats perform alternation tasks in a T-maze (Wood et al., 2000; Frank et al., 2000; Ferbintineau & Shapiro, 2003), whereas many cells distinguish overlapping actions and locations on the maze, some cells capture the common places and events between the different types of episodes.

The notion that hippocampal cells might reflect the linking of important features across experiences and the abstraction of common information was also highlighted in recent studies on monkeys and humans. Hampson et al. (2004) trained monkeys on matching to sample problems, then probed the nature of the representation of stimuli by recording from hippocampal cells when the animals were shown novel stimuli that shared features with the trained cues. They found many hippocampal neurons that encoded meaningful categories of stimulus features and appeared to employ these representations to recognize the same features across many situations. Kreiman et al., (2000a) characterized hippocampal firing patterns in humans during presentations of a variety of visual stimuli. They reported a substantial number of hippocampal neurons that fired when the subject viewed specific categories of material, e.g., faces, famous people, animals, scenes, houses, across many exemplars of each. A subsequent study showed that these neurons are activated when a subject simply imagines its optimal stimulus, supporting a role for hippocampal networks in recollection of specific memories (Kreiman et al., 2000b). A subsequent study showed that some hippocampal neurons are activated a subject views any of a variety of different images of a particular person, suggesting these cells could link the recollection of many specific memories related to that person (Quiroga et al., 2005). This combination findings across species provides compelling evidence for the notion that some hippocampal cells represent common features among the various episodes that could serve to link memories obtained in separate experiences.

The second outcome from a network of linked memories is a capacity to use the common elements to retrieve multiple memories that include that element. Furthermore, hippocampal representations could support a capacity to “surf” the network of linked memories and identify relationships and associations among items that were experienced in distinct memories and therefore are only indirectly related. A single cue could generate the retrieval of multiple episodic and semantic memories, and cortical areas can access these multiple memories to analyze the consequential, logical, spatial, and other abstract relationships among items that appeared separately in distinct memories. These logical operations on indirectly related memories can support inferences from memory. The activity of searching and surfing networks of memories, and then comparing and contrasting memories could underlie our awareness of memories and the experience of conscious recollection. The organization of linked experience-specific and experience-general memories with the capacity for association and inference among memories is called a “relational memory network.”

In a series of studies, Eichenbaum has used a model system of rodent olfactory memory to explore the importance of the hippocampus in the linking memories and using the resulting relational networks to make associational and logical inferences from memory. One study examined the role of the hippocampus in making indirect associations between stimuli that were each directly associated with a common stimulus. Initially, normal rats

and rats with hippocampal lesions were trained on a series of overlapping “paired associates” (Bunsey & Eichenbaum, 1996). On each trial, the rat was initially presented with one of two initial items in a pairing, and then had to select the arbitrarily assigned associate. For example, for training on the pairs A-B and X-Y, if A was the initial item, then the rat had to select B and not Y; conversely, if X was the initial item the rat had to select Y and not B. Then the rats were trained on a second paired associated list where the initial items were the second items in the first list and new items were the associates (B-C and Y-Z). Thus, when B was presented initially, the rat was required to select C and not Z; when Y was presented initially, the rats was then required to select Z and not C. After training on all four paired associates, the rats were tested on their knowledge of the indirect relations among the pairings. These tests involved presentations of an initial item from the first learned paired associates (A or X) followed by a choice between the second items of the later learned associates (C versus Z). Normal rats demonstrated their ability to express these indirect relations by selecting C when A was presented and Z when X was presented, whereas rats with selective hippocampal damage showed no capacity for this inference from memory. These findings, combined with observations on another transitive inference task (Dusek & Eichenbaum, 1997), indicate that the hippocampus is critical to binding distinct memories into a relational network that supports flexible memory expression.

In another experiment, rats learned a hierarchical series of overlapping odor choice judgments (e.g., $A > B$, $B > C$, $C > D$, $D > E$), then were probed on the relationship between indirectly related items ($B > D$?). Normal rats learned the series and showed robust transitive inference on the probe tests. Rats with hippocampal damage also learned each of the initial premises but failed to show transitivity (Dusek & Eichenbaum, 1997). The combined findings from these studies show that rats with hippocampal damage can learn even complex associations, such as those embodied in the odor paired-associates and conditional discriminations. But, without a hippocampus, they do not interleave the distinct experiences according to their overlapping elements to form a relational network that supports inferential and flexible expression of their memories (see also Buckmaster et al., 2004).

Complementary evidence on the role of the hippocampus in networking of memories comes from two recent studies indicating that the hippocampus is selectively activated when humans make inferential memory judgments. In one study, subjects initially learned to associate each of two faces with a house and, separately, learned to associate pairs of faces (Preston & Gabrieli, 2004). Then, during brain scanning, the subjects were tested on their ability to judge whether two faces who were each associated with the same house were therefore indirectly associated with each other, and on whether they could remember trained face pairs. The hippocampus was selectively activated during performance of the inferential judgment about indirectly related faces as compared to during memory for trained face-house or face-face pairings. In the other study, subjects learned a series of choice judgments between pairs of visual patterns that contained overlapping elements, just as in the studies on rats and monkeys, and as a control they also learned a set of non-overlapping choice judgments (Heckers et al., 2004). The hippocampus was selectively activated during transitive judgments as compared to novel non-transitive judgments.

These findings indicate that the hippocampal relational network mediates the linking of distinct episodes that may contain items that have not been experienced in the same episode or in the same context. In doing so, the hippocampus plays a role in more than simply binding items within memories, but also mediates associations between distinct memories. During recollection, the hippocampus supports a capacity to generate multiple memories that share a common element, and the information contained within these memories can be used by many brain systems to make judgments about causal, logical, temporal, and spatial relations among the items in those memories (Cohen & Eichenbaum, 1993). Iterations of association, retrieval, and re-coding memories according to deduced relationships among the items would lead to the development of a systematic organization of items and episodes in memory wherein facts and events are linked to one another by a broad range of causal, logical, temporal, spatial, and other relevant relationships among the items. And this organization supports flexibility in the expression that is characteristic of recollective memory, specifically involving inferences between items that are only indirectly related.

2.2 The Anatomy of Memory

How do the above described memory functions emerge from the circuitry of the hippocampus? The brain system that mediates retrospective and prospective memory is composed of several cortical association areas interacting with structures in the medial temporal lobe (MTL), and in particular, the hippocampus. There is a general consensus that areas of the cerebral cortex are specialized for distinct aspects of cognitive and perceptual processing that are essential to memory, and that the cortex is the repository of detailed representations of perceptions and thoughts. The MTL is the recipient of inputs from widespread areas of the cortex and supports the ability to bind together cortical representations such that, when cued by part of a previous representation, the MTL reactivates the full set of cortical representations that compose a retrospective memory. Areas of the cortex both direct the storage of memories in the MTL and interpret the reconstructed memories generated by the MTL to support prospective memory. This simple, anatomically based scheme provides the framework on which our model is built. In the following sections, we will describe in greater detail the functional components of this system and the pathways by which information flows among them, and a qualitative model of how they interact to support retrospective and prospective memory.

The anatomy of the brain system that supports memory is remarkably conserved across mammalian species (Manns & Eichenbaum, 2007). Information processing in this system occurs in three main stages. The first stage involves virtually every neocortical association area (Burwell et al., 1995; Suzuki, 1996). Each of these neocortical areas projects to one or more subdivisions of the parahippocampal region, which includes the perirhinal cortex, the parahippocampal cortex, and the entorhinal cortex. The subdivisions of the parahippocampal region are interconnected and send major efferents to multiple subdivisions of the hippocampus itself. Thus, the parahippocampal region serves as a convergence site for cortical input and mediates the distribution of cortical afferents to the hippocampus. Within the hippocampus, there are broadly divergent and convergent connections that could mediate a large network of associations (Amaral & Witter, 1989), and these connections support plasticity mechanisms that could participate in the rapid coding of novel conjunctions of information (Bliss & Collingridge, 1993).

The outcomes of hippocampal processing are directed back to the parahippocampal region, and the outputs of that region are directed in turn back to the same areas of the cerebral cortex that were the source of inputs to the MTL.

Only highly pre-processed sensory information reaches the MTL, but these inputs come from virtually all higher-order cortical processing areas. Perhaps the most thoroughly studied cortical area afferent to the hippocampus is the inferotemporal (IT) cortex, the highest-order visual object processor in primates. Ablation (removal of material from the surface of an object by vaporization or other erosive processes) of the inferotemporal cortex results in a visual-guided learning and deficits without impairment in visual fields, acuity, or threshold. The behavioral physiology of inferotemporal cortex is consistent with the data from ablation studies, showing that IT neurons are maximally driven by complex visual patterns, and the response properties of these cells are dependent on attentional mechanisms and reward association. Many IT neurons are preferentially responsive to a particular pattern, often one that is of obvious significance to the animal, including cells that respond selectively to faces. IT neurons respond differently to the same stimuli when they appear as stimuli to-be-remembered, or when they were novel versus familiar, and some cells maintain firing during the memory delay periods during performance of short term memory tasks. In humans, distinct ventral temporal areas that include and surround IT are activated by presentation of different categories of visual cues, including faces, tools, and animate objects (Martin, 2007; Kanwisher, 2007).

Other major inputs to the MTL arise from the posterior parietal area. Damage to this cortical area results in impairment in neglect of contralateral sensory stimulation across sensory modalities (Mountcastle et al., 1975; Andersen, 1989). One area within parietal cortex that has received particular interest is area 7a where most cells are visually driven. These cells have very large receptive fields and neuronal responsiveness is highly dependent on attentional factors. These cells respond best when the stimulus is the target of an eye or hand movement and they prefer moving stimuli but show little preference for stimulus form or color. These and other data indicate that the posterior parietal area is specialized for attention and egocentric spatial analyses including localization and visual and manual acquisition of targets in space. Also, areas of the parietal and temporal cortex are involved in complex perceptual processing essential to configuration of the conceptual contents of information that is the subject of recollection (e.g., Uncapher et al., 2006).

Additional major inputs to the MTL arise from several areas within the prefrontal cortex, a sensory-motor-limbic integration area involved in the highest-order cognitive functions including motor programming, vicarious trial and error, and memory (Fuster, 1995). In humans components of the prefrontal cortex mediate working memory, effortful retrieval, source monitoring, and other processing currently being specified that contribute critically to cognitive functions essential to recollection (Dobbins et al., 2002). In addition, midline structures within the prefrontal and cingulate cortical areas have been identified as activated during processing of self-referential information that may be strongly related to autobiographical memory (Northoff & Bermpohl, 2004; Fink et al., 1996; Cabeza & St Jacques, 2007).

The nature of cortical inputs to the MTL differs considerably across mammalian species (Manns & Eichenbaum, 2006). The proportion of inputs derived from different sensory

modalities also varies substantially between species, such that olfaction (e.g., rats), vision (e.g., primates), audition (e.g., bats), or somatosensation (e.g., moles) have become disproportionately represented in the brain in different animals (Krubitzer and Kaas, 2005). Nevertheless, the sources of information derived from prefrontal and midline cortical areas, as well as posterior sensory areas, are remarkably consistent across species.

Despite major species differences in the neocortex, the organization of cortical inputs to the hippocampus is remarkably similar in rodents and primates. Across species, most of the neocortical input to the perirhinal cortex comes from association areas that process unimodal sensory information about qualities of objects (i.e., “what” information), whereas most of the neocortical input to the parahippocampal cortex comes from areas that process polymodal spatial (“where”) information (Suzuki & Amaral, 1994; Burwell et al., 1995). There are connections between the perirhinal cortex and parahippocampal cortex, but the “what” and “where” streams of processing remain largely segregated as the perirhinal cortex projects primarily to the lateral entorhinal area whereas the parahippocampal cortex projects mainly to the medial entorhinal area. Similarly, there are some connections between the entorhinal areas, but the “what” and “where” information streams mainly converge within the hippocampus. The cortical outputs of hippocampal processing involve feedback connections from the hippocampus successively back to the entorhinal cortex, then perirhinal and parahippocampal cortex, and finally, neocortical areas from which the inputs to the MTL originated.

2.3 Towards a functional organization of a cortical-hippocampal memory system

The anatomical evidence reviewed above suggests the following hypothesis about how information is encoded and retrieved during memory processing. During encoding, representations of distinct items (e.g., people, objects, events) are formed in the perirhinal cortex and lateral entorhinal area. These representations along with back projections to the “what” pathways of the neocortex can then support subsequent judgments of familiarity. In addition, during encoding, item information is combined with contextual (“where”) representations that are formed in the parahippocampal cortex and medial entorhinal area, and the hippocampus associates items and their context. When an item is subsequently presented as a memory cue, the hippocampus completes the full pattern and mediates a recovery of the contextual representation in the parahippocampal cortex and medial entorhinal area. Hippocampal processing may also recover specific item associates of the cue and reactivate those representations in the perirhinal cortex and lateral entorhinal area. The recovery of context and item associations constitutes the experience of retrospective recollection.

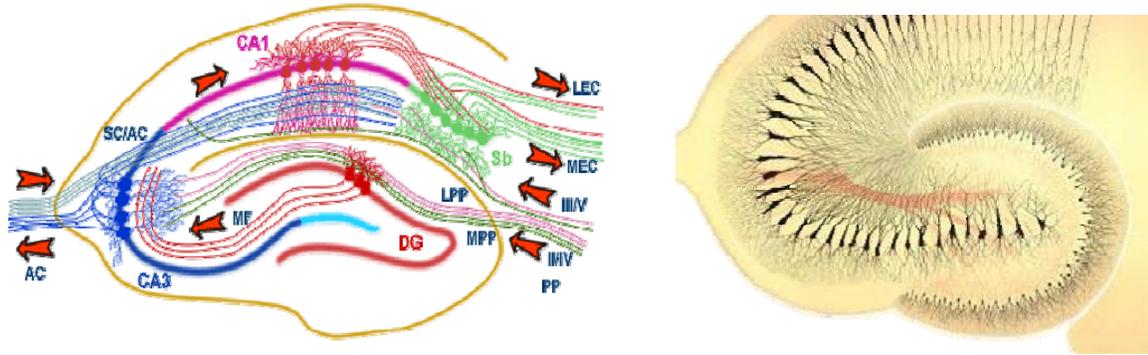


Figure 2. The Hippocampal Structures of the Brain

Perirhinal cortex and lateral entorhinal area. Substantial evidence indicates that neurons in the perirhinal cortex and lateral entorhinal cortex are involved in the representation of individual perceptual stimuli. Electrophysiological studies on monkeys and rats performing simple recognition tasks have identified three general types of responses (Brown & Xiang, 1998; Suzuki & Eichenbaum, 2000). First, many cells in these areas exhibit selective tuning to memory cues such as odors or visual stimuli. Second, some cells maintain firing in a stimulus-specific fashion during a memory delay, indicating the persistence of a stimulus representation. Third, many cells have enhanced or suppressed responses to stimuli when they re-appear in a recognition test, indicating involvement in the recognition judgment. Similarly, in humans, among all areas within the medial temporal lobe, the perirhinal area selectively shows suppressed responses to familiar stimuli (Henson et al., 2003). Complementary studies in animals with damage to the perirhinal cortex indicate that this area may be critical to memory for individual stimuli in the delayed non-matching to sample task in rats (Mumby & Pinel, 1994; Otto & Eichenbaum, 1992) and monkeys (Suzuki et al., 1993). These and other data have led several investigators to the view that the perirhinal cortex is specialized for identifying the memory strength of individual stimuli (e.g. Brown & Aggleton, 2001; Henson et al., 2003; Aggleton et al., 2004).

Parahippocampal cortex and medial entorhinal area. The parahippocampal cortex and medial entorhinal area may be specialized for processing spatial context. Whereas perirhinal and lateral entorhinal neurons have poor spatial coding properties, parahippocampal and medial entorhinal neurons show strong spatial coding (Burwell & Hafeman, 2003; Hargreaves et al., 2005). Further, the immediate early gene *fos* is activated in perirhinal cortex by novel visual cues, but *fos* is activated in the postrhinal cortex by a spatial re-arrangement of the cues (Wan et al., 1999). In addition, whereas object recognition is impaired following perirhinal damage, object-location recognition is deficient following parahippocampal cortex damage in rats (Gaffan et al., 2004) and monkeys (Alvarado & Bachevalier, 2005). Similarly, perirhinal cortex damage results in greater impairment in memory for object pairings whereas parahippocampal cortex lesions results in greater impairment in memory for the context in which an object was presented (Norman & Eacott, 2005). Parallel findings from functional imaging studies in humans have dissociated object processing in perirhinal cortex from spatial processing in

the parahippocampal cortex (Pihlajamaki et al., 2004). Furthermore, whereas perirhinal cortex is activated in association with the memory strength of specific stimuli (Henson et al., 2003), the parahippocampal cortex is activated during recall of spatial and non-spatial context (Ranganath et al., 2003; Bar and Aminoff, 2003).

Hippocampus. Compelling in support for differentiation of functions associated with recollection come from within-study dissociations that reveal activation of the perirhinal cortex selectively associated with familiarity and activity in the hippocampus as well as parahippocampal cortex was selectively associated with recollection (Deselaar et al., 2006; Davachi & Wagner, 2002; Davachi et al., 2003; Ranganath et al., 2003). These and many other results summarized in a recent review suggest a functional dissociation between the perirhinal cortex, where activation changes are consistently associated with familiarity, and the hippocampus and parahippocampal cortex, where activation changes are consistently associated with recollection (Eichenbaum et al., 2007). An outstanding question in these studies is whether the parahippocampal cortex and hippocampus play different roles in recollection. In particular, the above described findings on parahippocampal activation associated with viewing of spatial scenes suggests the possibility that this area is activated during recollection because recall involves retrieval of spatial contextual information. By contrast, the hippocampus may be activated associated with the combination of item and context information.

CA1 versus CA3. Several recent studies have suggested that subregions of the hippocampus may play distinct roles in memory. A particularly striking contrast comes from a comparison between two studies by Kesner and colleagues (Gilbert and Kesner, 2003; Kesner et al., 2005). In one experiment, normal rats learned associations between a particular object or odor and their locations in specific places in an open field. On each trial, one of two objects (differentiated by visual or olfactory cues) was placed at one of two locations on a large open field. If object A was in place 1, a reward could be found underneath. Similarly, if object B was in place two a reward could be obtained by displacing the object. However, no reward was available if either object was presented in the alternate location. Normal animals improved in performance across days, as reflected in differentiating their latencies to approach object in rewarded vs non-rewarded locations. Selective lesions of CA3 completely blocked acquisition of object-place associations, whereas CA1 lesions had no effect. In contrast, the opposite pattern of results was found in another study where rats were taught associations between an object and an odor that were separated by a short delay. The animals learned that if object A was presented before the delay, then a cup of sand would contain a food reward if it was scented with odor 1 (but not with odor 2). Conversely, if object B was presented first, then a cup of sand would contain a food reward if it was scented with odor 2 (but not odor 1). Memory was measured by a briefer latency to approach the scented cup on rewarded pairings (A-1 and B-2) than on non-rewarded pairings (A-2 and B-1). In normal rats, the latency to approach rewarded cups gradually decreased over daily training sessions, at about the same rate as observed in the previous object-place association study. In contrast, rats with selective CA1 lesions showed no sign of acquiring the associations between temporally separated objects, whereas rats with CA3 lesions acquired the task just as rapidly as normal animals. These results are consistent with the idea CA1 is specialized for representation of the order of events that are separated in time (Manns & Eichenbaum, 2006).

3. CURRENT COMPUTATIONAL REPRESENTATION

3.2. General Description of Architecture

For this LDRD project we focused on higher-level processing, occurring after eye foveation and movements such as saccades. The model starts at the point where the visual input images have been separated into two components. The first sub-image corresponds to the area seen by the focal area of the eye, and the second sub-image contains the entire field-of-view for the eye. This division models the higher resolution present in the fovea, as well as the way focus and context information are treated separately through some parts of biological cortex. We will begin the description of the computational system by describing the bottom-up behavior involved in encoding episodic memories into the representation of the system. As illustrated in Figure 3, the first sub-image (the blue sphere at bottom of image) is directed to the ventral stream of neural processing where object detection and categorization is handled (i.e., “what” information). The second sub-image (the upper-right portion of the visual input) is directed to the dorsal stream where the spatial context of what the eye is seeing is determined and categorized (i.e., “where” information). This stream consists of a lower resolution (zoomed out) view of the entire field of view including the focal area itself. Note that both sub-images are further segmented into overlapping sub-sections for even greater specificity in category formation of our episodic memory.

Our cortical model is comprised of stacked layers of fuzzy Adaptive Resonance Theory (ART) networks. ART is a well established self-organizing neural technique for classifying input activations. The interested reader can find a wealth of literature on ARTs details, performance, and stability (see Vila, 1994; Walczak, 2005). While ART is a good choice for the classification modules, other unsupervised learning techniques could be expected to render similar results, though probably with subtle and interesting differences. Between each pair of ART modules there is a layer of Temporal Integrators (Taylor et al., 2009), with an adjustable time constant of integration for each layer. This TIART network is meant to model a biological cortical column, which receives afferent connections from either a particular subsection of the input field or a particular subsection of the previous layer. These subsections overlap in a manner inspired by the biological cortex (Kingsley, 2000). The system as a whole is a simple but powerful cortical classifier. Progressively higher layers encode progressively more abstract objects or spatial locations. Low levels correspond to simple perceptual primitives (edges etc.), high levels might correspond to whole objects or other semantic concepts. However, the cortical ART networks have an interesting modification. Each ART network has a “top down” recall mode driven by input from higher-level, more abstract layers as well as the traditional “bottom up” mode driven by stimulus input. While in “top down” mode, an activated F2 node in the ART network reinstates the prototypical input pattern which it encodes in synaptic weights between layers F1 and F2 (see Figure 4). So, for example if a particular F2 node encoded the concept of “dog” it could read out the features (“fur,” “tail,” “barks,” etc.) in the F1 layer. These features would in turn activate the F2 layer of the next lowest ART module and so on. In this way the network learns new inputs from the bottom up, and can then recall and reconstruct these features from the top down.

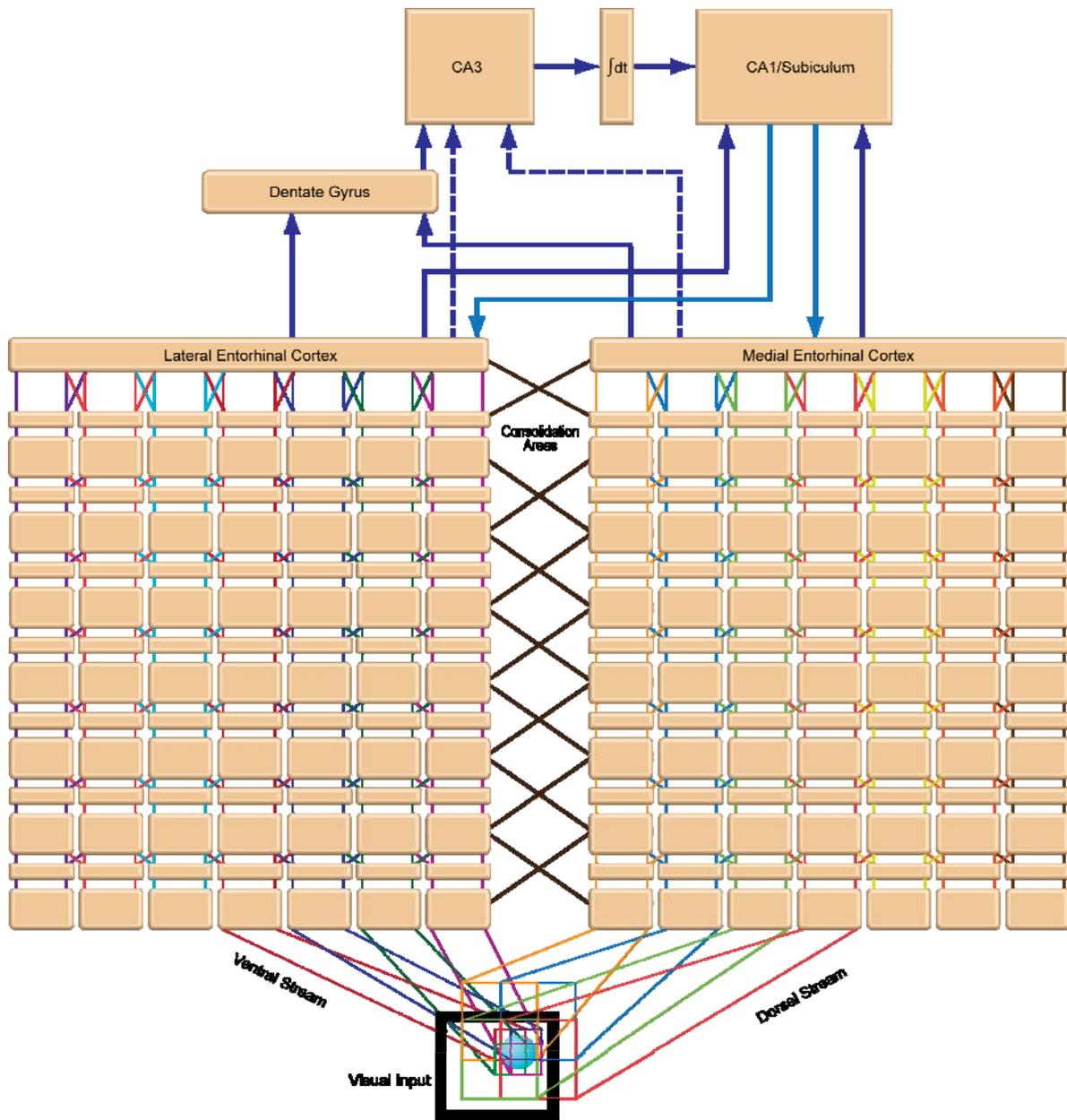


Figure 3. A conceptual view of the computational architecture

In our version of the cortex, one half of the cortical input columns receive high resolution input from the center of the visual field. This causes the network to develop templates which corresponds to "objects" and is meant to simulate the fovea near the sensory level and the ventral visual stream near the associative levels. The other half of the cortical columns receive low resolution input from the periphery which leads the development of spatial representations. This is meant to correspond to the off-center visual field at low levels and the dorsal visual stream at associative levels.

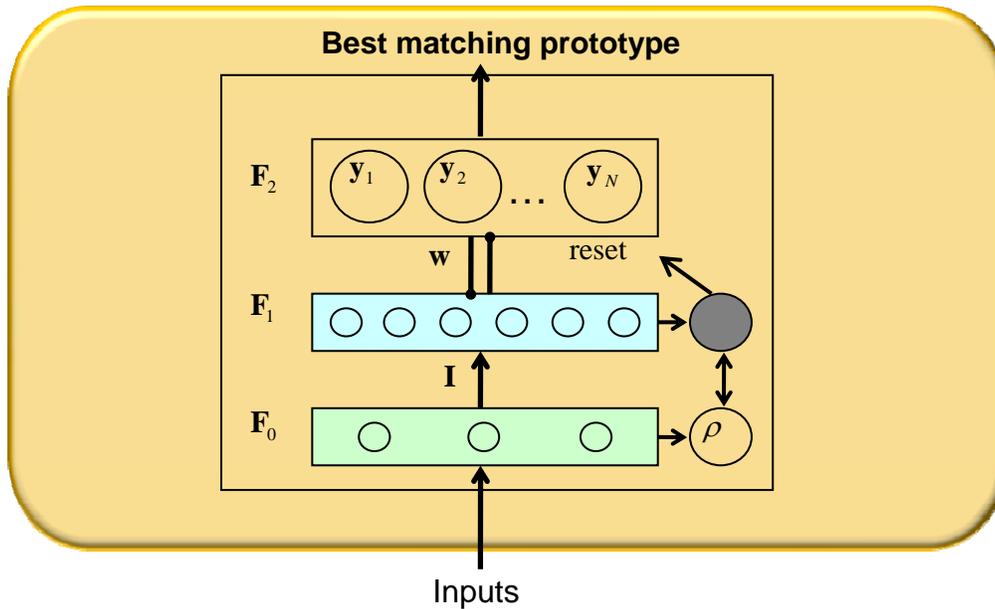


Figure 4. An example of an Adaptive Resonance Theory (ART) module

For each cortical modality (i.e., focus and context) there are three levels of cortex. For each level of cortex, we have implemented a grid of fuzzy ART modules (the current version has 6 layers at 7 X 7 modules; see Appendix 1). There exist temporal integrators between the fuzzy ART levels, but for these experiments the time constant of integration is dialed down to where they are inconsequential from the point of view of pasting together temporal events. The temporal integrators are dialed down because the experimental tasks are static in nature rather than temporal. A future extension of this architecture would be a single mechanism that dynamically adapts both static and temporal tasks.

3.1.1 The Hippocampus

The hippocampal system makes use of several ART variations, so it is related to the cortical system at a single unit level, but has a very different architecture and accomplishes fundamentally different information processing. While the cortex attempts to represent the conceptual structure of its inputs, the hippocampus attempts to quickly bind snapshots of high level cortical activity. Behaviorally, this gives us an episodic memory mechanism where concepts originating in multimodal sensory input are bound together. By way of this binding, the hippocampal representation can also be used to recover neocortical representations from partial activations.

Hippocampus is modeled as a loop of neural modules starting at entorhinal cortex, proceeding to dentate gyrus, continuing to CA3, then returning to entorhinal cortex through CA1, where some of the function of subiculum is implicitly captured in CA1. Entorhinal Cortex (EC) is the last level of cortex on the way from sensory input to the hippocampus. The EC is where all information that will be encoded in episodic memory

must converge. The Dentate Gyrus (DG) provides a pattern separation function for the information received from entorhinal cortex. The CA3 provides pattern completion and semanto-spatial association. CA1 closes the loop and provides temporal association.

The EC is an area of multi-modal convergence, where several data streams from different senses come together (Anderson, et al., 2007). The EC's cytoarchitecture resembles that of the cortex, so it here is modeled as is the rest of cortex. However, the ART networks which make up the EC have two sets of connections to the hippocampus, a feed forward connection to the DG and CA3 meant to simulate the perforant path, and bidirectional connections to the CA1/Subiculum component of the hippocampal model. The forward connections provide inputs to the hippocampal module. The back connections use LAPART rules to learn associative links between activity in CA1 and EC, thereby closing the autoassociative hippocampal loop. When a CA1 representation is activated, these back connections can drive top-down cortical recall (Figure 5).

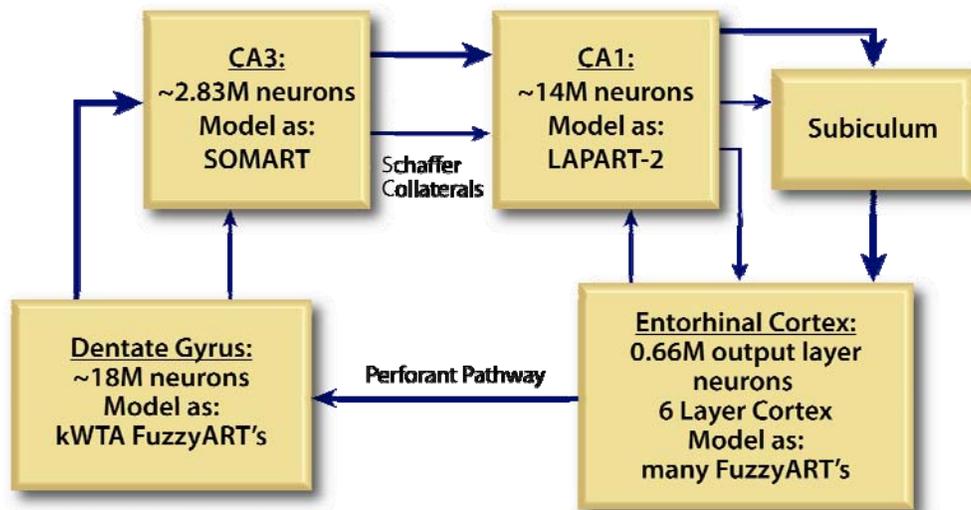


Figure 5. The modeled representation of the hippocampal system

In our implementation, the pre-Medial Temporal Lobe (MTL) sensory cortex and EC are represented by layers of fuzzy-ART modules which are modified to encode temporal semantic data. Individually, these temporally integrated adaptive resonance theory (TIART) modules are capable of encoding categorical representations of their given input vectors over time (Taylor, et al., 2009). By combining layers of TIART modules, our EC creates categories of categories to represent larger semantic concepts and combine the "dorsal stream" containing contextual information and the "ventral stream" of focal information before these streams enter the hippocampus. Within the hippocampal representation in our model, each of the primary regions is represented by a different ART variant selected to achieve the particular functionality of the individual region. The relative size of each module is scaled in accordance with approximate human neuroanatomy.

The DG has peculiar anatomical properties. It has a large number of neurons with relatively low activity and it is one of the few places in the brain in which new neurons are generated in the adult brain. These properties have led to the suggestion that the DG creates sparse, non-overlapping codes for unique events via pattern separation (Leutgeb, et al., 2007).

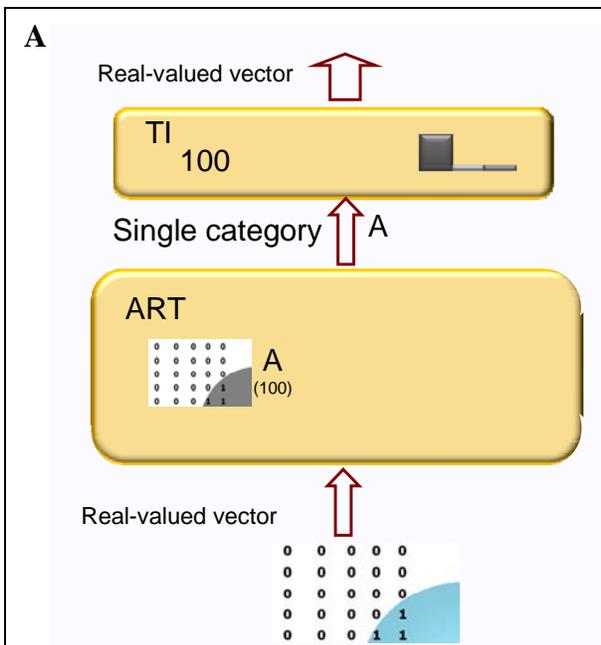
The DG in our model receives the conjoined multimodal sensory signals from EC. It performs pattern separation on this abundance of sensory information to produce sparse output activation, which ensures different semantic concepts are given unique encoding (Rolls & Kesner, 2006). Computationally, a series of k-winner-take-all (k-WTA) fuzzy-ART modules constitute the DG module of our model. A WTA module is a competitive network in which a single concept beats out competing concepts to represent the input vector. Effectively, a sparse encoding is created as each of the k WTA modules yields a single output. Similar input vectors will be represented by the same single winning output, and dissimilar inputs will be represented by a differing winning output, yielding pattern separated outputs. These outputs serve as the input for CA3.

Anatomical studies of the hippocampus proper reveal cytoarchitecture which differs radically from that of the cortex (Anderson, et al., 2007). While both CA1 and CA3 both contain pyramidal cells like the cortex, existence of extensive recurrent connections in CA3 and the presence of inhibitory and excitatory interneurons have led some investigators to suggest that CA3 may be involved in pattern completion (O'Reilly & Rudy, 2001).

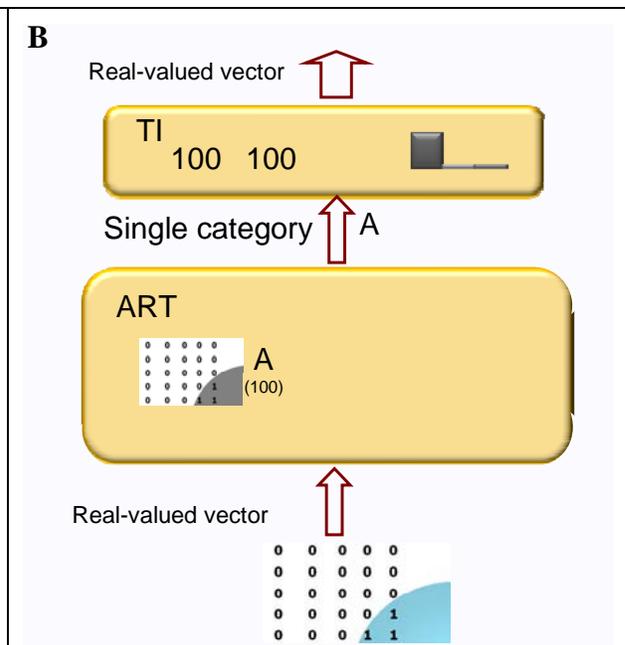
In this implementation the sparse output pattern from DG serves as input to CA3. Functionally, CA3 assists with episodic binding through auto-association. In our model, this functionality is represented by a Self-Organizing Map (SOM) structure. A standard SOM transforms a given input vector into a distinct topological region without supervision guiding the classification (Haykin, 1999). Incorporating the neighborhood updating capabilities of a SOM within a fuzzy-ART module, we have created a SOMART module to represent CA3. This module is capable of mapping semantically similar inputs to proximate topological regions. Thus the learning algorithm creates "islands" of activity which respond to similar input sets, but avoids a global topology. In effect, related concepts are clustered together to help associate episodic memories and these "islands" of relational bindings form the inputs to CA1.

Anatomically, the output of CA3 proceeds to CA1 and then to the subiculum as the major output region of the hippocampus. However, the exact functionality of the subiculum is largely unknown, so we have merged the capabilities of CA1 and subiculum in our model. CA1 has been implicated in learning relational information for temporal sequences and connecting these episodic encodings back to the original sensory inputs from EC. This ability to link sequences allows for temporal packaging of episodes. Since our CA3 can only encode momentary conjunctions, we need a mechanism which can capture sequences of changing relations. Thus, CA1 contains a unit which temporally integrates CA3 outputs using a set of leaky integrators. This provides a temporal gradient of input conjunctions coming from CA3, the oldest bindings will have the weakest signal in the temporal integrator, while the most recent bindings will be most strongly represented. This temporally coded sequence of CA3 activity is used by CA1 to create a topology of temporal sequence.

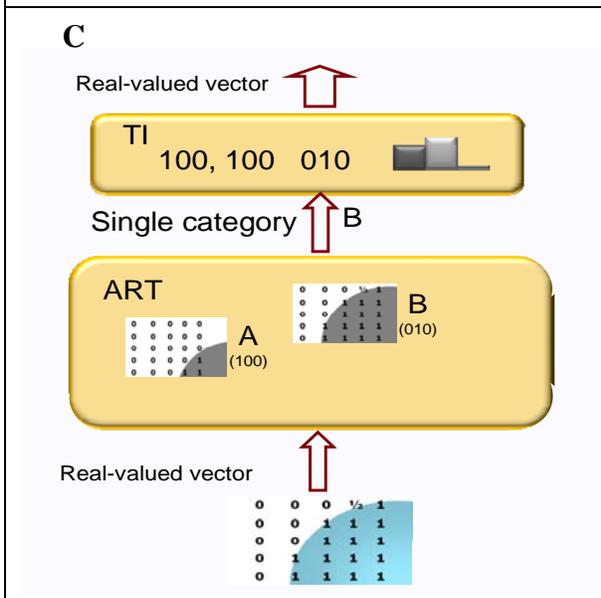
Once this temporal topology has been established, activity in CA1 is associated with activity in the EC via a Laterally Primed Adaptive Resonance Theory (LAPART) partially-supervised learning paradigm. Local CA1 learning is supervised in that a certain sequence of CA3 activations corresponds with certain EC activation. CA3 sequence *A*, where that sequence is translated to an instantaneous representation through the temporal integrator, is bound through learning to EC activation *B*. LAPART uses two ART modules connected by a lateral activation field, so the activations on each side are generalized via the ART classification mechanism. Through experience, a connection weight is learned to bind the node that corresponds to each classified CA3 sequence to a node that corresponds to some EC activation. This mapping of sequences onto the high-level cortical representations closes the hippocampal loop, and allows activations in CA1 to cue top-down recall in the cortex and unspool the temporal representations it has created. This entire process is graphically described in Figure 6 (A-K).



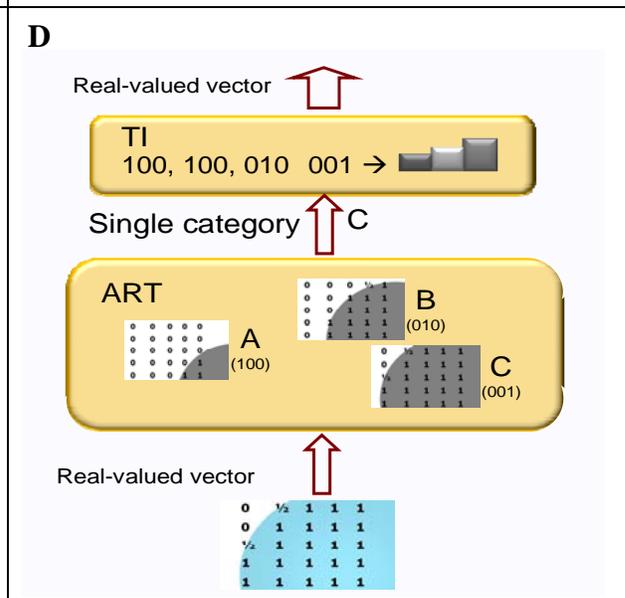
When presented with a novel input pattern, an ART module creates a representative category. The temporal integrator connected to the ART module raises the activation of the corresponding category to threshold.



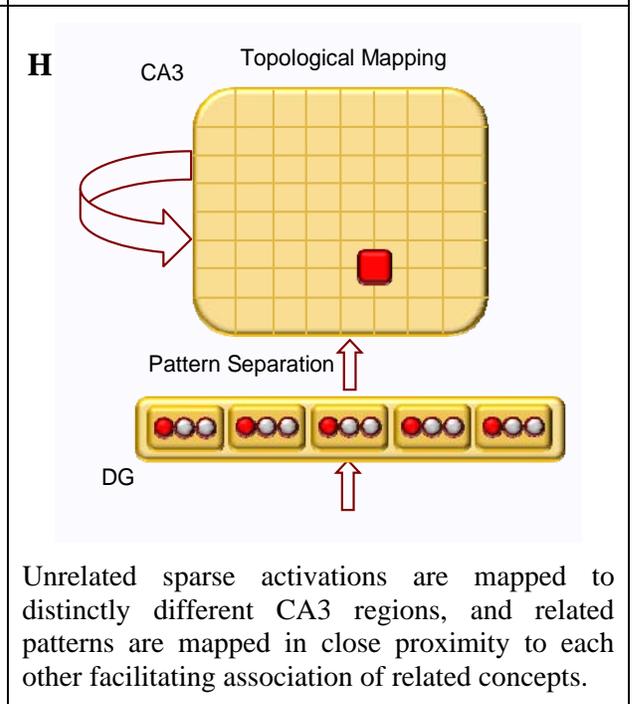
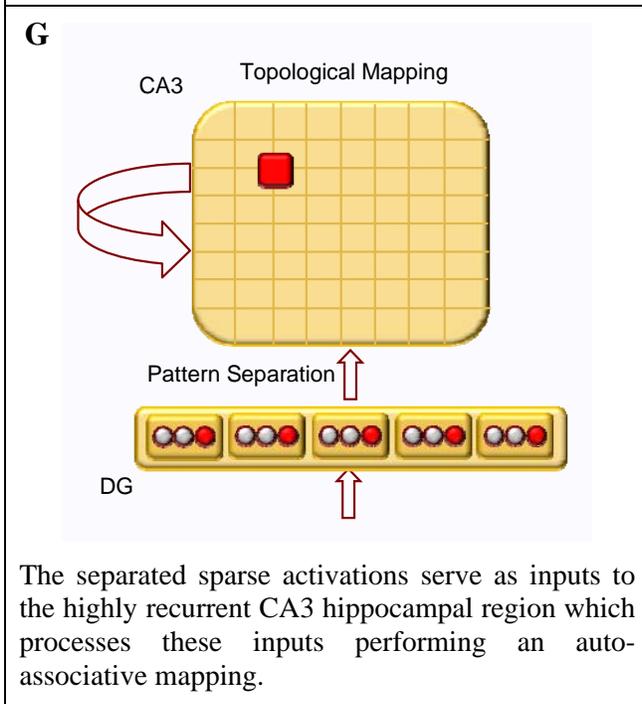
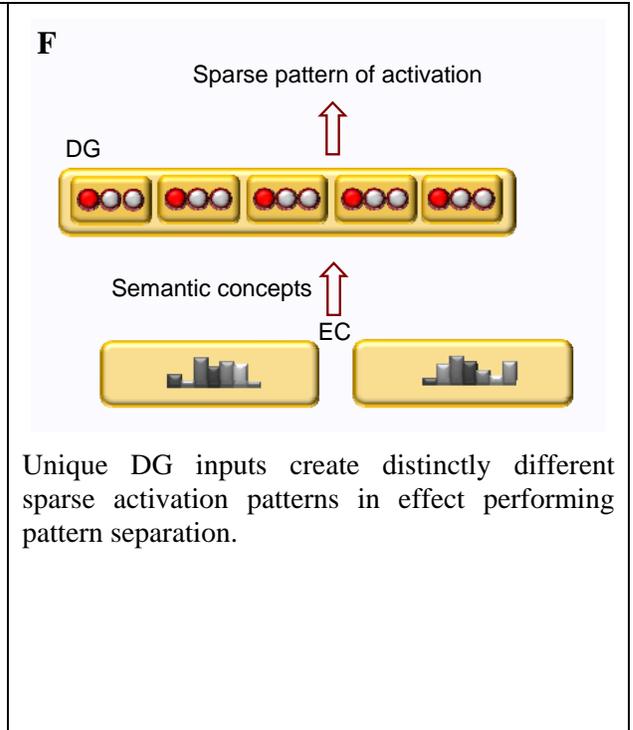
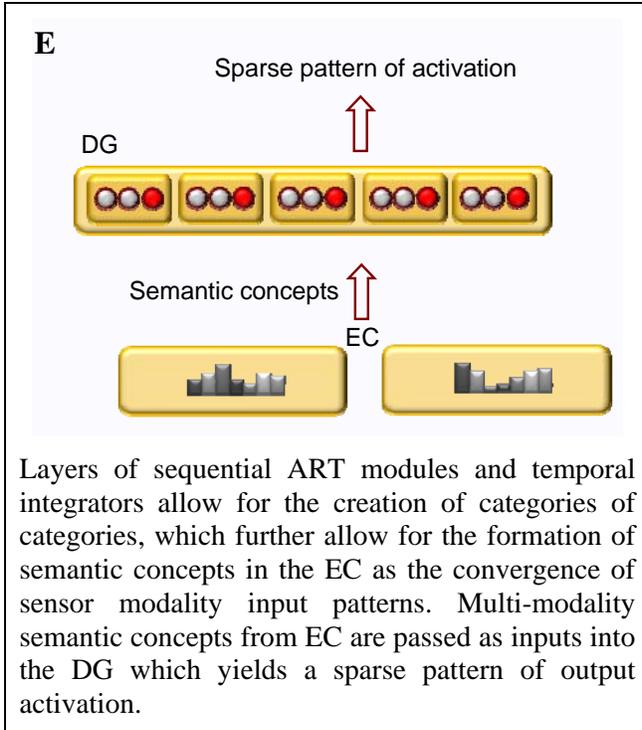
When the same input pattern is presented, once again, the ART module uses the same category encoding & the corresponding temporal integrator activation remains at threshold.



When a sufficiently different input pattern is presented, the ART module forms a new category & primes the corresponding temporal integrator value. The temporal integrator decays the activation value of the previous category.



With the presentation of a third distinct input pattern, a corresponding third category is formed which primes the respective temporal integrator value. The temporal integrator decays the activations of previously viewed categories creating a real-valued gradient vector of temporally integrated single categories.



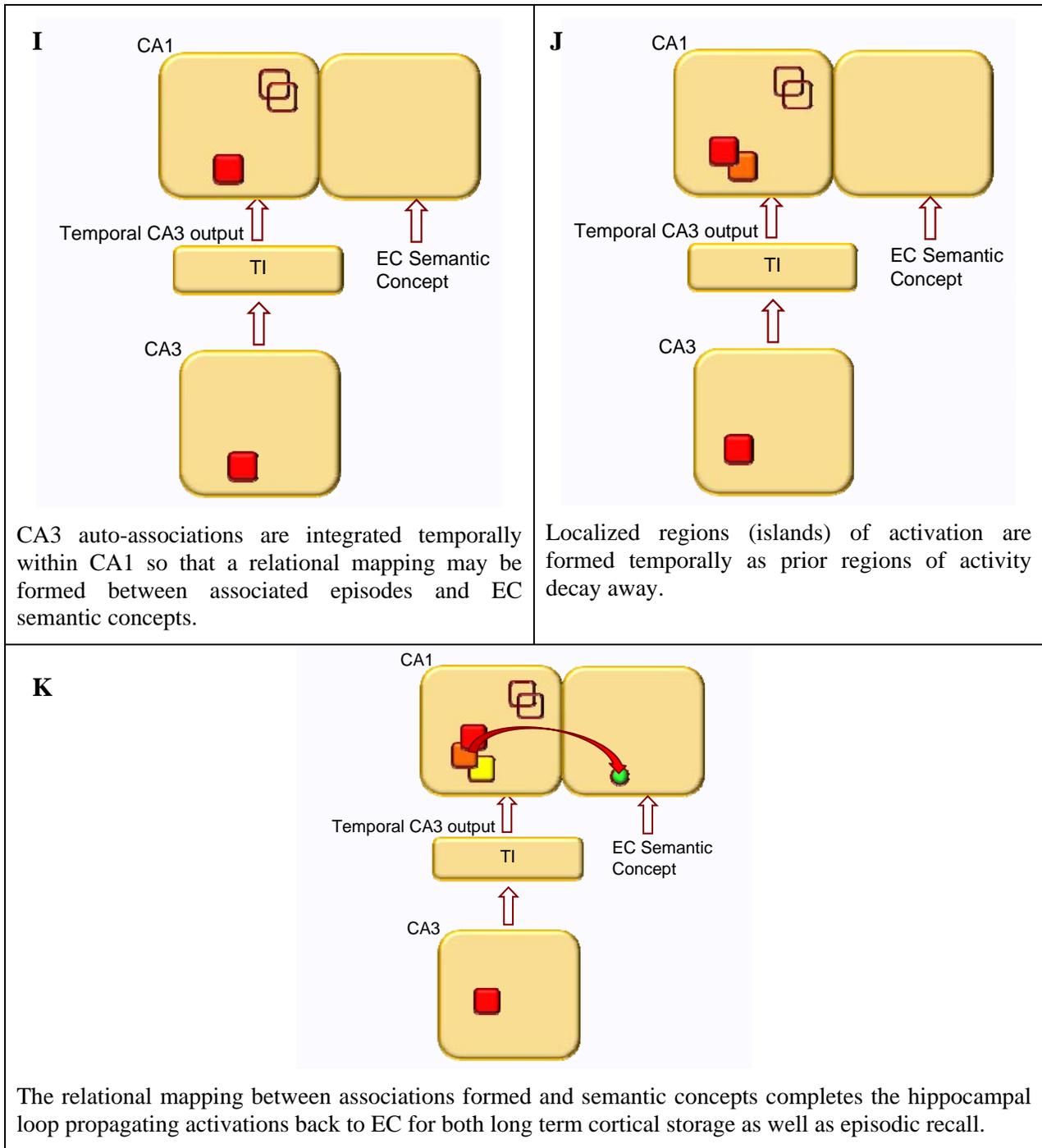


Figure 6. Step-by-step process of episodic/semantic activation

In our computational system, when an episodic memory is recalled, the hippocampus activates one of the categories in CA1, and it begins the top-down recall of the episodic trace which it encodes. In the recall, the active category in the CA1 ART unit encodes a temporal sequence of conjunctive associations which were formed in CA3 during the

episodic memory formation. During recall of a specific conjunctive association in CA3, each stream (ventral and dorsal) that makes up part of the conjunctive association is activated in top-down fashion. The top-down activation of each stream is similar, thus the top-down description of a single stream will be given in the interest of brevity. In the top-down activation of a cortical stream (either “what” or “where”), the top-level ART category is activated (in the EC or DG). This top-level category consists of a concatenation of ART categories from each column in the cortical stream. Top-down recall of this concatenation of categories consists of a simultaneous recall of each column starting at the ART unit which is at the top of each column. It is this ART unit which provides its category for the concatenation during the formation of the episodic memory. The top-down recall for each column is similar, thus the description of a single column will be given in the interest of brevity. During recall of a cortical column, the category in the top-most ART unit is activated in top-down fashion. This category contains a temporal sequence of categories from the next lower ART unit as integrated through the integration unit between them. The recall continues from ART unit to ART unit downward through the connecting temporal integration units until the bottom ART unit is reached. In top-down activation, each temporal integration unit contains a temporal sequence of ART category activations which were fed to it as input during episodic memory formation. During recall of the temporal sequence of ART categories in a temporal integration unit, each category in the sequence is re-activated in top-down fashion in the same temporal order as was originally experienced and encoded in the episodic memory trace. When the recall reaches the lowest level ART unit, it is ready for “replay.” During replay, the memory is re-activated in forward or bottom-up fashion in the same temporal sequence it was originally experienced. In this system, temporal information is stored in the activation potentials of temporal nodes. Local semantic information is stored in the synaptic weights of the ART modules. Long-term, memory can occur through Hebbian-like adaptation of synaptic connection weights between local cortical areas. In our model, a local cortical area at a given level is comprised of a collection of nodes that all influence the activation of the same ART output node.

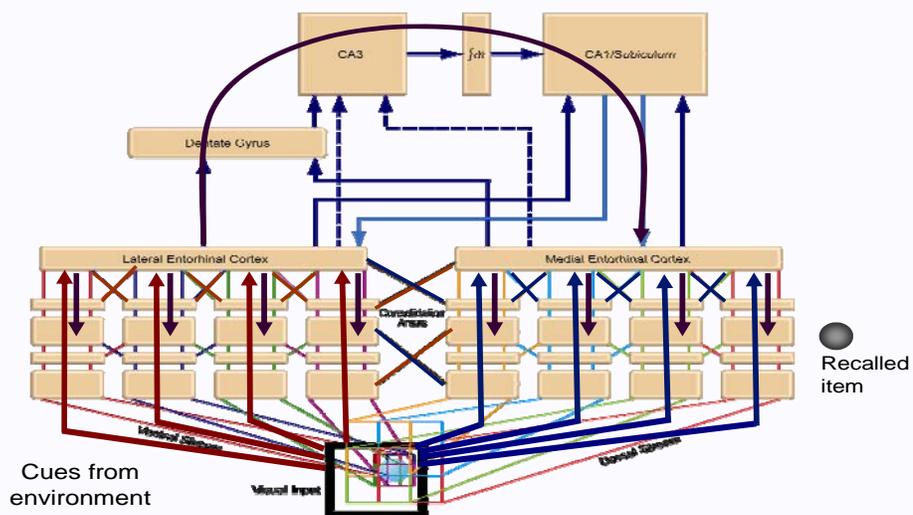


Figure 7. The bottom-up/top-down flow of information that both consolidates memory into LTM and produces recall when prompted

4. ASSESSMENT OF THE MODEL

Assessing a computational model for the degree of neuro-cognitive plausibility is a significant challenge. No one qualitative or quantitative method is sufficient to adequately evaluate the level of agreement between a computational model and the analogous brain system it seeks to represent. Thus, we employed several evaluative methods that were based on empirical, human studies as a means to quantitatively compare the model. These comparisons helped to qualitatively assess the accuracy of the model, whereby the more qualitative comparisons, the greater potential for an accurate assessment of the model. Each comparison described below was meant to address a key aspect or function of episodic memory.

4.1 Temporal and Sequential Memory and Recall of Objects

This section details our approach to assessing the model's learning of temporal semantics. First, the method of assessment is described and then a step-by-step example of its operation is given. Finally, the experimental method of assessment is demonstrated with a discussion of the results.

This explanation and the first experimental assessment have been presented in Taylor et al., 2009. We encode temporal semantic data as a recency gradient of generalized classifications. Fuzzy ART is used as the classifier that creates an active output on a certain $F2$ node for any given input vector. We implemented an ART network with a fixed number of $F2$ nodes available for recruitment, where each available $F2$ node (whether it had yet been recruited or not) is connected to a leaky temporal integrator. This implements a static architecture, which is convenient for explanation. The method might be extended to dynamic architecture creation (allowing ART to recruit new $F2$ nodes indefinitely) in modeling neuro-development. The integrators need be leaky, otherwise the output would continually increase over time (assuming continual input greater than zero), eventually saturating. A general leaky integrator is modeled as (Carpenter, Grossberg, & Rosen, 1991). Where y is the integrator output, x the input, and a an integration constant.

$$dy/dt = -ay + x \quad (0.2)$$

$$y(n+1) = (1-a)y(n) + x(n) \quad (0.3)$$

We implement a discretized (using Euler's method with an arbitrary sample period of one time unit) version of the leaky integrator (0.2) as formula (0.3). Where $1-a$ is a decay constant.

This method is simple enough in concept that we can provide a comprehensive example of its operation. The following example assumes three arbitrary sensory input vectors. We use sensory inputs to create grounded, stand-alone, examples. However, if the sequence of inputs over time were internal cortical activation patterns, the example could be describing an additional functional level of cortex, above that which created the inputs.

Let there be three distinct sensory input activation vectors A , B , C that form a temporal semantic sequence which we wish to encode. Inputs feed into an ART module (see Figure 9). Each input results in a different active node output on the ART. By placing a leaky temporal integrator on each ART output node, we encode a temporal sequence of inputs as a single real valued vector. The integrated vector is a recency gradient, where the order of element amplitudes (from low to high) represents the order of occurrence of the input vectors (from oldest to most recent). As mentioned, the value of a given integrator output node will decrease over time. As a result, the

farther in the past a given input was observed, the smaller a value the corresponding integrator output will have (until at some small activation level, the integrator output is lost in the noise of the system). A more biologically faithful sequence recall scheme (Sun & Giles, 2001), could involve an extra step where by leaky integrator activations provide an inhibitory signal to other nodes, such that highest activations could be ordered first in temporal sequence.

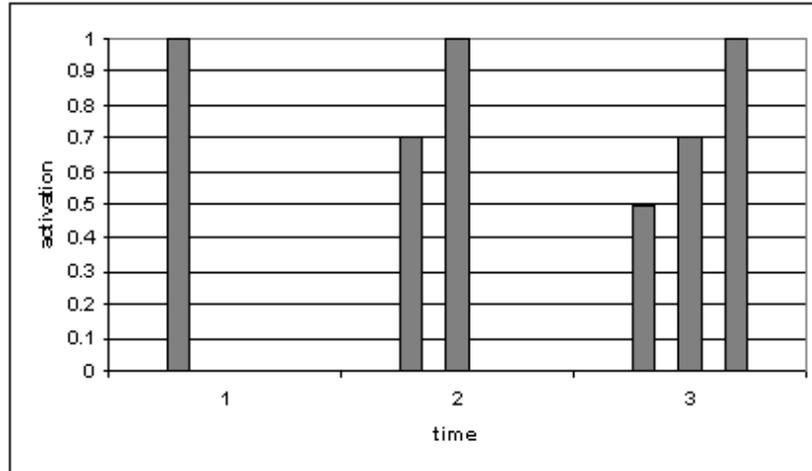


Figure 8. Time series of temporal integrator outputs given example input: "A", "B", "C"

If the first system input is a semantic concept *A* (the specific vector of input data is not important, the simplest assumption would be that the semantic concept *A* is represented by an array of pixel values that visually correspond to the letter *A*), then let the output of the ART be $[1\ 0\ 0]$, considering only the first three outputs for simplicity. Likewise, let the ART outputs corresponding to inputs *B* and *C* be $[0\ 1\ 0]$ and $[0\ 0\ 1]$ respectively. The temporal integration array initializes to $[0\ 0\ 0]$. Figure 8 illustrates the temporal sequence of integrator node outputs that result from presenting *A* at timestep 1, *B* at timestep 2, and *C* at timestep 3.

$$[0.5\ 0.707\ 1.0] \tag{0.4}$$

Finally then, the temporal input sequence *A*, *B*, *C* is encoded at a single point in time as vector (0.4). Ascending values in the vector indicated encoded temporal order, while connectivity to the rest of the architecture encodes semantic content. In the context in which it was formed, and to the level of detail that the ART categories have formed generalized templates encodes both the temporal and semantic information of the input sequence (Sun & Giles, 2001).

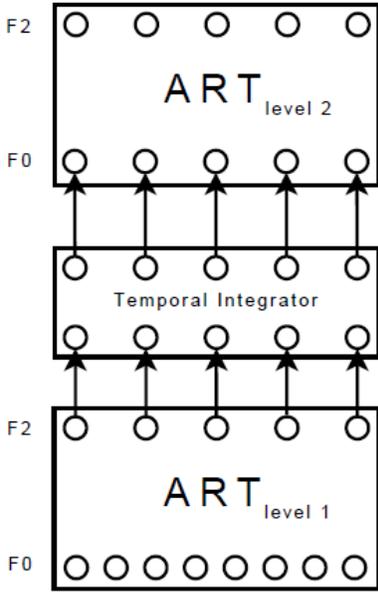


Figure 9. ART with temporal integrator

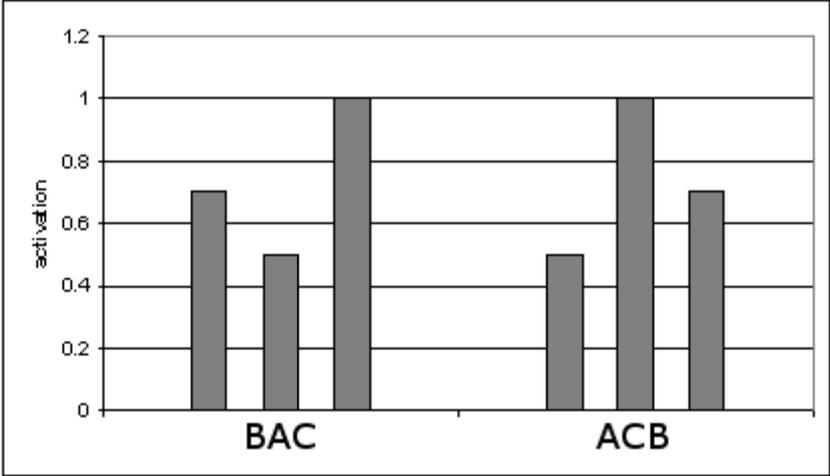


Figure 10. Example temporal integrator outputs

Vector (0.4) can now be encoded by a level 2 ART (see Figure 9) to uniquely represent the sequence *A, B, C*. Figure 10 illustrates some other possible temporal integrator outputs, given the indicated input sequence. Level 2 ART could encode these other vectors as representations of the corresponding temporal semantic sequence. Because the level 2 ART representations are unique (to the level of precision determined by the ART operational parameters), a top-down recall operation can recover the original sequence *A, B, C* from the level 2 ART encoding of (0.4). For example, let the temporal integrator output (0.4) result in activation of the first F2 output on level 2 ART. Then, given an augmented ART with top-down recall ability, as well as a playback

mechanism for the temporal integrator, top level stimulation of the first F2 node on level 2 ART will result in the sequence *A, B, C* being played back at the system input level.

Later work will delve into neural implementations of these top-down recall mechanisms. A brief overview of possible mechanisms is given now, so that their functionality can be used in the experiments that confirm temporal semantic encoding has occurred. We supplement ART by specifying a top-down behavior. When an *F2* node is stimulated from above, it plays down the associated template's activation levels to the ART input layer. Template activation level adjustment is where all memory storage in ART occurs, so playback of a given template represents recall of one memory component at the scale of that ART unit.

We also supplement the bottom-up temporal integration scheme with a top-down behavior. As the output of a temporal integrator array is a pattern encoding the order of input activation, recall of that gradient should play back the temporal integrator inputs in that order. When a pattern is placed on the temporal integrator array from above, the array will first activate the input corresponding to the lowest value in the pattern, then the next lowest, and so on. This behavior will play back the input activations in their original order. As a manner of implementation, we can imagine the top-down stimulation of the temporal integrator array as setting a threshold for each element of the array. The integrators then start integrating up from zero and fire the associated input node when their internal value reaches the threshold. The lowest threshold will be reached first, which is correct because it represents the input that occurred farthest in the past and therefore the input that should be played back first.

4.1.1 Experimental Results

This experiment demonstrated successful encoding of temporal semantic data. In this case, the semantic meaning will be visual sensory observation. The formation of the encodings themselves can be tracked by probing node activations in the architecture. However, this can at most show that some representations were formed (not that those representations are correct). We demonstrate that the representations are valid encodings of the input information by initiating a recall process that decodes the temporal semantic representations into whatever those representations encode in the context of the system. If the encoding, in the context in which it is stored and recalled, decodes to the original information (to the resolution of the system), then we have solid evidence that the desired functionality is captured by the described method.

Figure 9 shows the architecture used for the recall experiment. The visual sensory input is a two dimensional array of 10 by 10 pixels. The first level ART forms templates to represent the visual input symbols. The temporal integrator forms semanto-spatial patterns to represent sequences of symbols observed by the first layer ART as passed on by the first layer ART's *F2* node activation. The second layer ART forms templates that represent the patterns output by the first layer temporal integrator array that represent temporal sequences of input symbols. In this case, the timescales are such that the second layer ART captures temporal integrator patterns corresponding to three input timesteps. Figure 11 shows an input sequence. Figure 12 shows one recall sequence, in this case recall corresponds to the middle three symbols of the input.

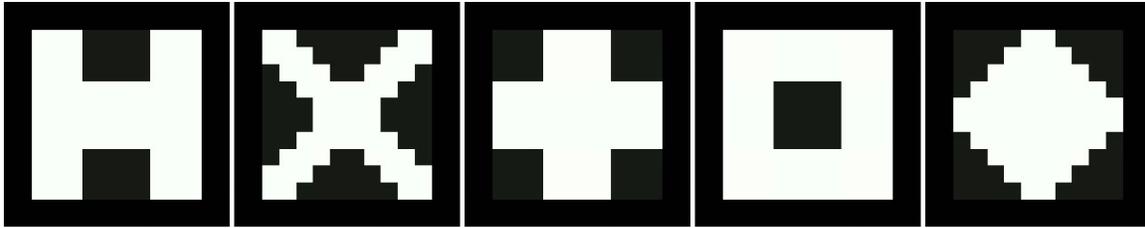


Figure 11. Memory and recall input sequence

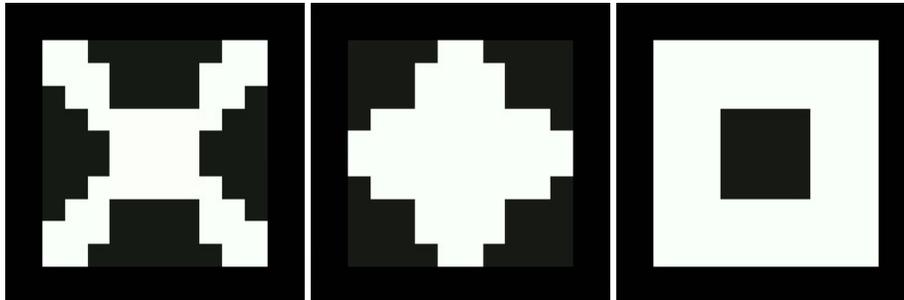


Figure 12. A sample memory and recall output sequence

We note that the recalled sequence is not an exact copy of the input sequence. The plus and diamond symbols have been aliased to a common symbol. This aliasing occurs because in the metric used by ART, the plus and diamond symbols are sufficiently close for classification to the same $F2$ template. An easy parallel can be drawn in human memory formation. To the non-expert, transient observation of either a grey Toyota Corolla or a grey Honda Accord is likely to form an aliased memory of "grey import sedan." In ART, the vigilance parameter determines how close two inputs must be to be classified into the same template.

The vigilance parameter was set arbitrarily at 0.8, in a range of 0 to 1, for the described experiment. In our architectures, the vigilance parameter of the ART units is an independent variable that can be used to tune the performance of the network. One of the trade offs inherent in tuning with the vigilance parameter is memory space vs. precision, greater vigilance will form a greater number of more precise templates. We can tune a part, or all, of this sort of architecture for more precise (higher resolution) temporal semantic encoding, but only at the expense of using up more memory space.

As stated earlier, the temporal semantic encoding method is capable of encoding information to the precision of the generalization used in the ART module(s). This experiment demonstrates that valid temporal semantic encodings were formed because when the encodings are decoded, the original temporal and semantic information is recovered.

In addition to the limitations of the ART parameters used, the user would want to consider supplemental mechanisms (possibly such as those found in (James, 2001)) if exact encoding of sequences is important. The anticipated use of this method is the simulation of biological cognitive processes, so some deficit in perfect memory is acceptable. Further research can characterize the consequences of potential imperfections in the system, as compared to

imperfections in human cognition. Research shows better relative memory performance, as opposed to absolute memory performance in humans (Sejnowski & Rosenberg, 1987).

Another consideration for future applications is the packaging of input episodes. The above example uses static episode size for simplicity of analysis. A more interesting system would dynamically package episodes, possibly based on rate of change and/or novelty.

$$V_m = \frac{RT}{F} \ln \frac{P_K[K^+]_o + P_{Na}[Na^+]_o + P_{Cl}[Cl^-]_i}{P_K[K^+]_i + P_{Na}[Na^+]_i + P_{Cl}[Cl^-]_o} \quad (0.5)$$

There exists a biological correlate, in the behavior of neurons, to the sort of temporal integration described here. Membrane potentials integrate down (assuming sub-threshold stimulation to some level above resting equilibrium) over time due to leakage current through resting ion channels. The Goldman equation (Vila, 1994), describes the influence of ionic concentrations on the neuron membrane potential (Waibel, et al., 1989). An elevated membrane potential would reflect deviation from an equilibrium (where the Goldman equation expressed that equilibrium for the pertinent ions in neurons) of ionic concentrations between the inside and outside of the neuron. Ions would then flow through resting (non-gated) ion channels until equilibrium was restored. The time course of this equilibrium restoration can be described by temporal integration.

The rate of change of the membrane potential is a function of the number of resting channels and the number, connection strength, and activity of afferent neural connections. We abstract beyond gated ion channels and outgoing action potentials as we only seek to explain neural plausibility rather than a full model of neuron. With proper balance and biasing, the rate of change in membrane potential (and hence the rate of decay of the integration) could be tuned over a wide range. This tuning allows the arbitrary time scale representation that we mentioned earlier.

We illustrate the biological correlate both to show neurological plausibility of our technique, but also as evidence for pervasive temporal integration (and thereby pervasive co-encoding of both temporal and semantic information at multiple levels through cortex). No neurons function without some form of temporal integration, though a counter argument is that the time scale of ion-channel temporal integration is not relevant to information encoding.

Structure is a critical characteristic of any neural system (Wan, 1994; Heathcote, 1995), no less so in the method described here. Temporal semantic information encoded as a recency gradient only has meaning within the structural context in which it was encoded. The activation of a temporal integration node, which represents the information that a square was the most recent symbol in a certain sequence, only has that specific meaning because there exists a connection between that node and a classifier node that represents square. Furthermore, that classifier node only represents square because of the particular connects between the classifier functional subsection inputs and visual sensor outputs. The guidance to be appreciated by the consideration of structure is that when building up larger architectures with these methods, one must keep in mind invariant structural mappings of the processed information, lest one lose or corrupt the information being encoded.

4.2. Associating Object/Scenes Pairs

For the second assessment we compared the model results to study results of Hannula et al. (Haykin, 1999). The Hannula study presented human subjects with a series of face-scene pairs in a study block, and then tracked eye movements for sets of three faces, with a background scene, presented in a test block. The set of three faces can be from one of three categories: match, re-pair, or novel. The match face sets contain three known (previously seen) faces, one of which is correctly paired with the background scene. The re-pair face sets contain three known faces, but none of them are correctly paired with the background scene. The novel face sets contain three unknown faces. This task is an exercise in episodic memory for associating people and places.

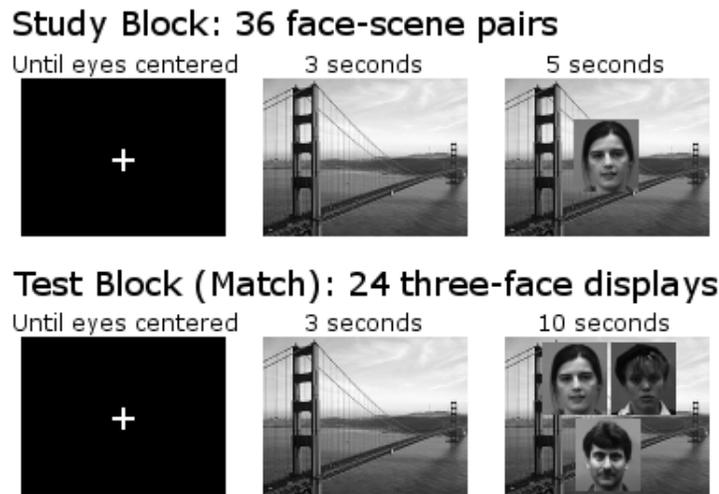


Figure 13. Hannula image scheme

Published results show that normal subjects will spend a larger proportion of viewing time directed to a face that correctly matched the background. Subjects with hippocampal damage did not exhibit this proportional increase in dwell time on the matching face. This result indicates that hippocampus is required for the recognition of previously observed episodes.

The goal of this assessment is to show evidence that our model exhibits some of the same function as biological brain with regard to scene/object pair association. In the interest of correlating behavior from our model to human behavior, we create a mapping of the human experimental setup to an experiment that we can run, in simulation, on our model. To mitigate visual processing effects, we map the face-scene focus-context images to simple geometric images (initially using squares and triangles, then going to orthogonal lines and dashed lines). Our input images were ten pixels by ten pixels.

We present arrangements of our focus and context images that correspond with the study and test image presentations of the original experiment. The study presentation sequence of the original experiment is fixation, scene, face. As our model lacks a mechanism for separating focus and context information in the visual field, we must simulate that separation by presenting separate images to the focus and surround modality inputs. The presentation sequence for the focus modality is fixation, scene, face. The presentation sequence for the context modality is fixation,

scene, scene. These sequences reflect the fact that the scene image is a focus image during the second element of the original study presentation because the scene is the only image on the screen. The context modality only ever sees the scene because even when the face is present in the original study presentation, the scene still forms the background of the image. Part of the original experimental setup is that visual dwell time on an image is a measure of recognition of that image. As the model is lacking eyes, an alternate measure of recognition must be developed. The simulated measure of recognition is equivalent to directly probing neural activation in a human brain. Modeled neural activation can be evaluated by observing the ART classifier module output in the cortex model and the grid node outputs in the hippocampus model. A representative output report is shown in Figure 14. Model recognition scores are computed by summing contributions from each cortical classification module, and the hippocampus. The cortical classification modules can each contribute one point, and the hippocampus can contribute a point. This scoring convention was arbitrary and was sufficient for our purposes. The first ART module in the cortical focus modality contributed a point if it identified an existing template (i.e. it had previously learned a generalization) for the current input. As inputs are presented in sequences of three (fixation, scene, face), the first ART module will make its contribution based on the last element of the sequence. The second ART module in each cortical modality is located after a temporal integrator, and so it will score familiarity based on the whole sequence.

Another biological brain mechanism that our model lacks is the ability to concentrate on different portions of an image. As such, we must simulate that ability for the purpose of the test images. Instead of a single sequence with the last image containing three faces, we present three sequences with the last image each containing one face. This way, the model does not need to consider three sub-images as the human subjects do when looking at the single test image of three faces. The experiment outputs from Figure 5 reflect a test sequence in the original experiment where faces 1, 2, and 3 are shown against scene 1, then faces 2, 5, and 6 are shown against scene 2. In the study portion of the experiment, face 1 was viewed with scene 1 and face 2 was viewed with scene 2.

4.2.1 Results

Figure 14 shows an example of the experimental results. This example reflects two test sequences from the Hannula experiment. As this example is from the match category of faces, all three faces are known. One face should have an episodic memory associating it with the tested scene.

Presentation timestep	Focus-Input Index	Context-Input Index	Total Familiarity Score
1	1	1	5
2	2	1	4
3	3	1	4
4	2	2	5
5	5	2	4
6	6	2	4

Figure 14. Experiment outputs

The intact model exhibits higher familiarity scores when previously studied matching focus-context (face-scene) image pairs are presented, as opposed to pairs that were not studied

together. This behavior correlates to the eye dwell time of human subjects in the Hannula study. If the hippocampus section of the model is lesioned, familiarity scores are the same between matching and non-matching image pairs. This behavior also correlates with human subjects, where subjects with hippocampal damage do not preferentially dwell on particular faces during the matching and non-matching face-scene pairs. Further results show that there is no difference in familiarity scores between different focus images in the re-pair and novel tasks, with either intact or lesioned models. These results correlate with the Hannula data where hippocampally damaged and normal subjects both view faces in the re-pair task with no preference.

4.2.2 Repeated Experiment

The aforementioned results were obtained from initial model architectures. With significant advances made to the architecture we re-ran the experiment and were subsequently able to improve upon the mapping between the procedure performed by Hannula et al. and our work. Model fidelity advances allowed us to use actual images of faces and scenes rather than representative geometric shapes.

Furthermore, advancements made to the hippocampus representation enabled us to restrict the neural activation analysis to CA3 where associations are formed. By doing so, we were subsequently able to analyze the model CA3 activations and attain the results similar to Hannula et al. These results correlate well with human performance data in which greater hippocampal activation is observed when an existent encoding may be retrieved. Likewise, partial hippocampal activity is required to explore the representation of a novel input, and little hippocampus involvement is required for re-paired episodes which do not fit prior encodings.

4.3 Co-occurrence of Shared Scenes with Novel Objects

For the third assessment we compared the model to a study performed by Preston et al. (Preston, 2004). In the Preston study, human subjects were trained on black and white photographs of face-house pairs and face-face pairs in three sets. The first set consists of pairs of faces and houses. The second training set introduces new faces paired with the same set of houses shown in the first set. And finally, the third training set consisted of face-face pairs which were previously unseen. During the testing phase of the Preston study, subjects performed forced-choice judgment tasks. Two of the tasks presented either a face or a house and required the subject select the corresponding house or face to complete the pair. The other two tasks focused on face-face pairs. One task was simply a test of the learned face-face pairs, whereas the other task tested subject's ability to recall related face-face pairs which shared a common house but which were never explicitly seen together.

Similarly, we trained our model using face-house pairs such that a face is processed by the ventral stream and a house is processed by the dorsal stream. An example of the input presented to the model may be seen in the left half of Figure 15. See Appendix B for the full training sequence. Our model lacks the ability to perform the forced-choice judgment task. So rather we first trained the model on face-house pairs including faces with a common house. Then, we turned off learning in the model so that no new concepts could be formed, but rather only existent concepts could be used. We then presented the model an ambiguous partial input cue by inputting a blank image to the ventral stream and one of the houses previously seen during training to the dorsal stream. This partial cue presentation may be seen on the right half of Figure 15. Rather than selecting between possible choices as in the forced-choice judgment task, we get our model to reconstruct the image that it has stored in memory associated with the house. The

resultant image is not an exact copy of the original input, but rather is an amalgamation of categorical representations distributed throughout the hierarchy of TIART modules comprising sensory cortex. A sample recalled face may be seen at the bottom right of Figure 15.

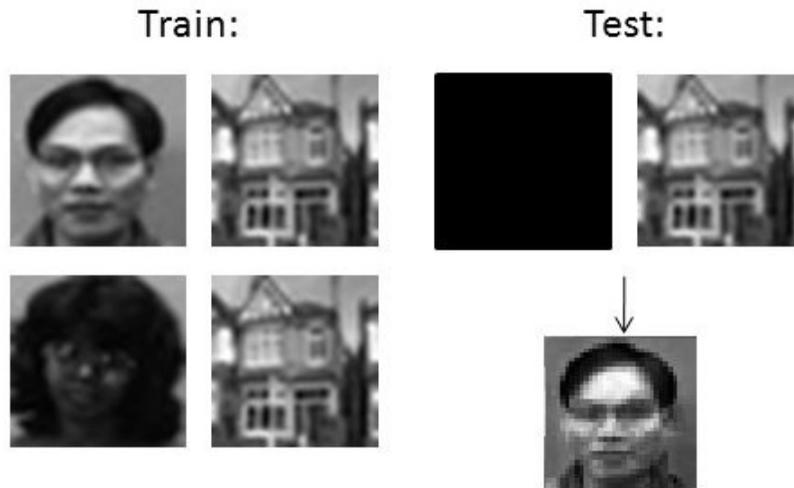


Figure 15. Face and scene pairs

4.3.1 Qualitative comparison results

In addition to the example shown in Figure 15, the model was successfully able to recall correct corresponding faces for each of the houses shown during testing. All of the recalled faces clearly resembled a particular face shown during the training phase; however, each was subject to slight distortions yielding an imperfect recall.

The ability to associate related face-face pairs can be observed qualitatively within the model by noting the activation regions within CA3. Portrayed in the upper left region of the graphical user interface (GUI), shown in Figure 16, informally one can observe whether or not the same CA3 activations are employed to encode the association of cohabitation. The visual input presented to the model can be seen in the lower left of the GUI. As displayed in Figure 16, when presented with face A and house A, a distinct region of CA3 is activated. Likewise, as shown in Figure 17, when presented with a different face B also paired with house A, an overlapping region of CA3 is indeed active indicative of the shared encoding between the related face-face pairing. On the other hand, as shown by Figure 17, when presented a distinctly different face C and a different house B an entirely different CA3 region is utilized for the encoding.

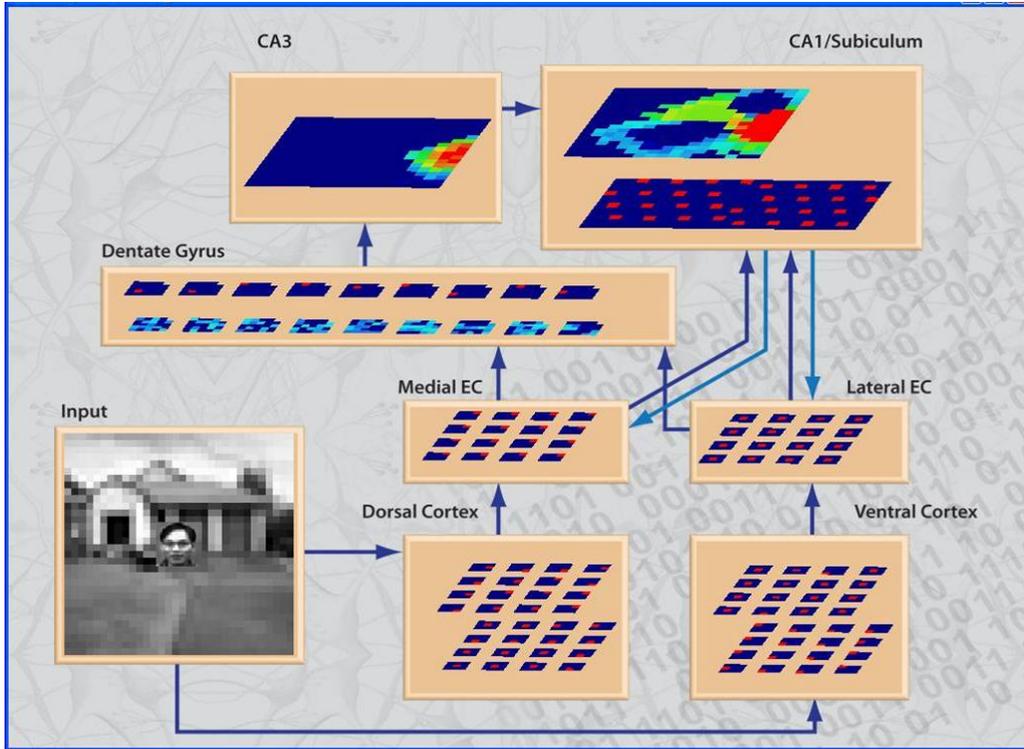


Figure 16. Person 'A' with house 'A'

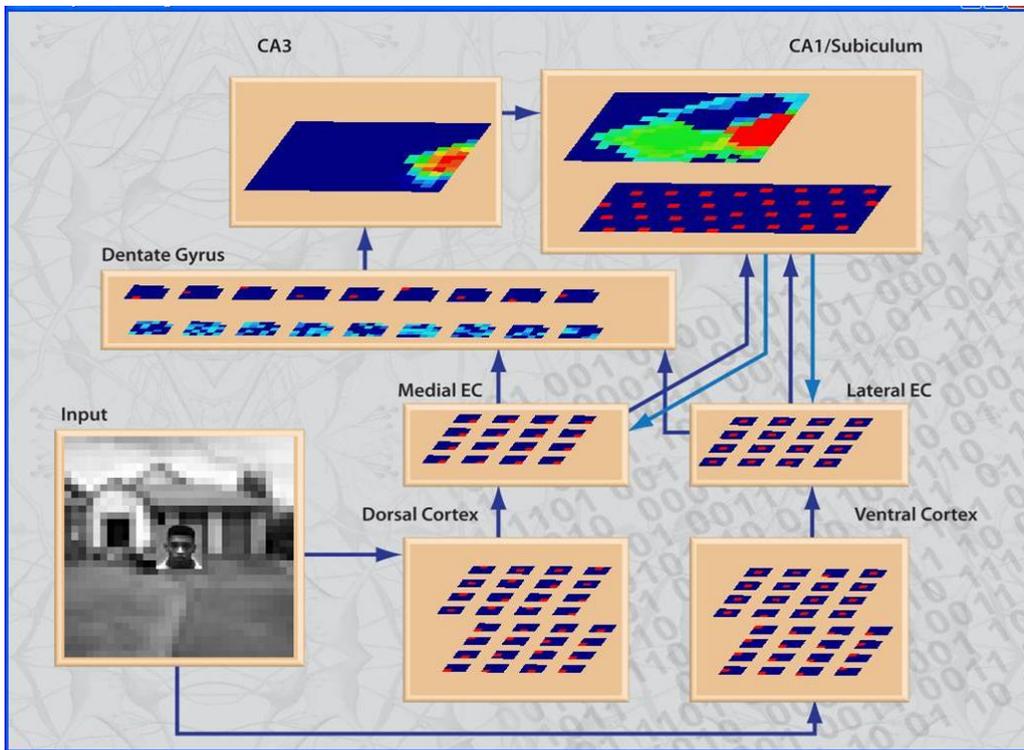


Figure 17. Person 'B' with house 'A'

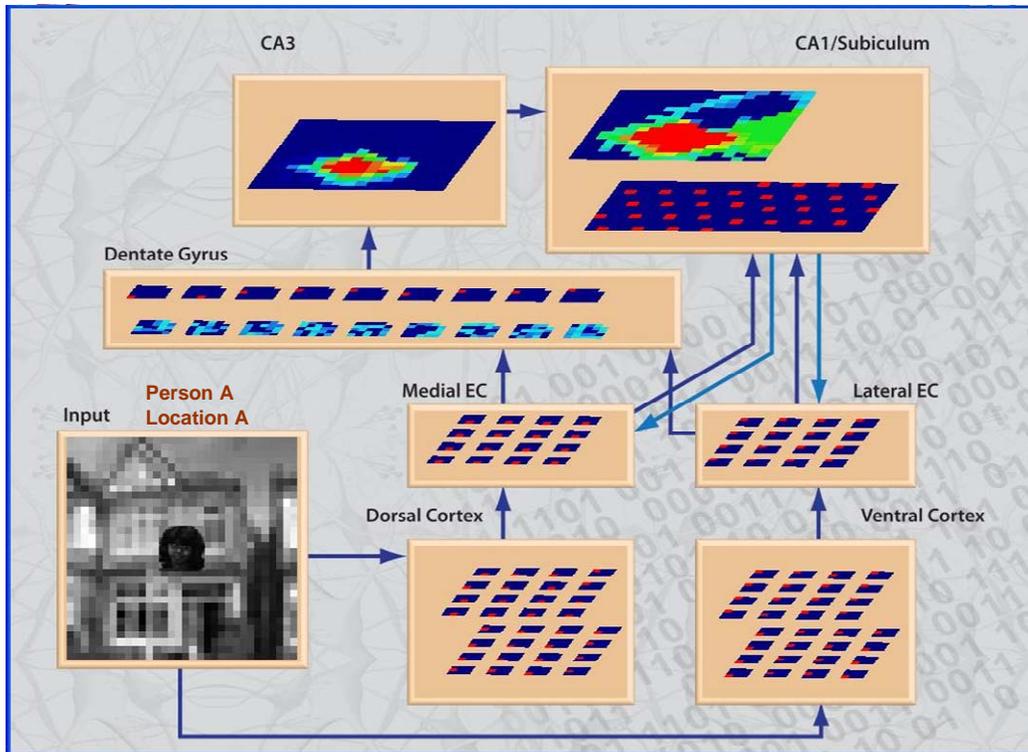


Figure 18. Person ‘C’ with house ‘C’ produces distinct mapping of activation

4.3.2 Quantitative comparison results

In the absence of a fully embodied model, with output modality to articulate the envisioned associations beyond cued recall, we have applied the mathematics of information theory to quantify the relationship between semantic concepts within the architectural implementation of CA3. Information theory allows for a quantitative evaluation of the information content independent of the particular computational implementation or the underlying neuroanatomical processes modeled.

More specifically, within the context of information theory, mutual information is a measure of the dependence between two random variables (Cover, 2005), and is computed by the double summation given in equation (6).

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (6)$$

Treating a conjoined face-house pair semantic concept as a random variable, the various CA3 encoding regions represent the alphabet of possible states the random variable may take on. In other words, a single random variable, such as X, represents the combined concept of a particular face and a specific house. Any specific pattern of activation within the CA3 may be used to represent the random variable, and thus the entire CA3 grid is the set of all possible values the random variable may express. From this perspective, mutual information may then be used to

quantitatively evaluate whether or not our architecture recognizes and auto associates inferred relationships.

In comparison to the Preston study, this technique allows us to evaluate whether or not our model is capable of forming an association between unseen related face-face pairs. A single face is only part of a random variable, and so for two different faces to share a relationship they must have a common context. The left column of Figure 19 lists the mutual information for the related face-face pairs our model was trained on. For instance, the first entry is the mutual information for two random variables A and E. A different face is represented by A than that of E; however both random variables share the same house. The right column on the other hand presents an averaged mutual information value of all the unrelated faces in reference to a particular face. As an example, the face represented by A is only related to the face represented by E. All other random variables (in this case B, C, D, F, G, and H) represent unrelated faces. Therefore, in column 2 of Figure 19, we represent the average mutual information values for non-matching (i.e. non-auto-associated) faces.

Mutual Information of Related Face-Face Pairs	Average Mutual Information of Unrelated Pairs
$I(A;E) = 0.3657$	$I(A;\sim E) = 0.0254$
$I(B;F) = 0.3628$	$I(B;\sim F) = 0.0303$
$I(C;G) = 0.3303$	$I(C;\sim G) = 0.0294$
$I(D;H) = 0.3570$	$I(D;\sim H) = 0.0322$
$I(E;A) = 0.3657$	$I(E;\sim A) = 0.0247$
$I(F;B) = 0.3628$	$I(F;\sim B) = 0.0293$
$I(G;C) = 0.3303$	$I(G;\sim C) = 0.0296$
$I(H;D) = 0.3507$	$I(H;\sim D) = 0.0323$

Note: \sim denotes negation

Figure 19. Model comparison to Preston study

Furthermore, we have tested the capabilities of our model on even more complex associations than those in the Preston experiment, to demonstrate the flexibility available in forming novel arbitrary associations. As shown in Figure 20, we have tested our model using a vehicle context in addition to houses. In addition to contextual relationship, a more advanced partially overlapping association occurs in this more advanced example.

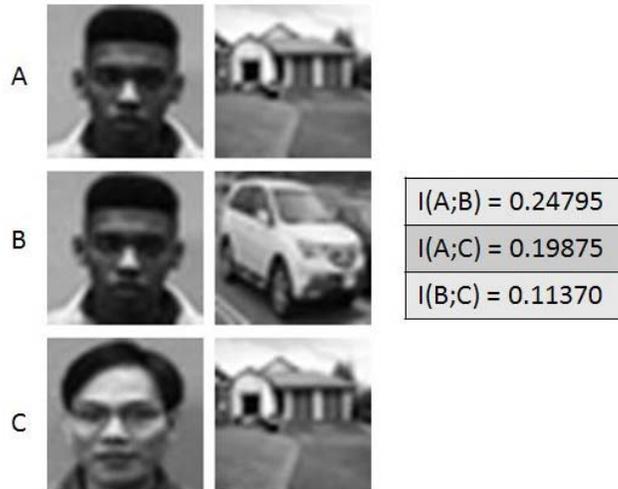


Figure 20. Faces paired with different contexts

The motivation behind the Preston et al. study was to investigate the role of the human hippocampus in the novel expression of declarative memories (Preston, 2004). Comparable performance by our model on an equivalent test demonstrates an apropos functional appropriation to the modules comprising our architecture.

In the Preston study, human performance was near perfect for the learned face-house pairs (Preston, 2004). Likewise, our model was successfully able to recall a correct face for each house presented such that the recalled image incurred only slight distortion. Due to algorithmic limitations, the present version of the model can only recall one of the faces associated with a given house. However, this could be corrected by allowing the model to retrieve all association pairs rather than only the single best match.

The Preston study observed increased hippocampal activation during fMRI scans when subjects were tested on related face-face pairs compared with learned face-face pairs (Preston, 2004). This observation demonstrates the important role of the hippocampus in relational tasks.

Beyond simply leveraging the hippocampus to form associations within our model, in particular mutual information has quantifiably shown our model is capable of forming associations between novel concepts. In our mutual information measure, we approximate the joint probability distributions for two semantic concepts. This approximation is calculated by computing the normalized fuzzy conjunction of the respective CA3 activations when the concepts are processed by the model individually. As can be seen by the mutual information approximation values given in Figure 20, the related face-face pairs have a significantly larger CA3 mutual information measure than that of unrelated pairs.

Furthermore, by incorporating vehicles as an additional context, we are able to demonstrate that our model is capable of processing a variety of concepts, as is true for humans, and is not only capable of processing houses. This more complex association additionally demonstrates the ability to associate multiple contexts with a single focus in addition to associating multiple foci with a single context. For example, as illustrated in Figure 8, the same person represented in concepts A and B is associated with a house in one concept and a vehicle in the next. A second

person is additionally associated with the same house as shown in concept C. While both people cohabitate the same house, only the first person is associated with the vehicle. The ability to differentiate between these overlapping associations is evident by the mutual information measures. Both the mutual information value associating the first person with his house and vehicle, as well as the mutual information value associating the two people cohabitating the same house are considerably larger than the relationship between the second person and the vehicle.

Comparing human CA3 activation during a co-occurrence task (Leutgeb, et al., 2005) to the CA3 activation of the computational model during a similar co-occurrence task yielded similar results. This is shown in Figure 21.

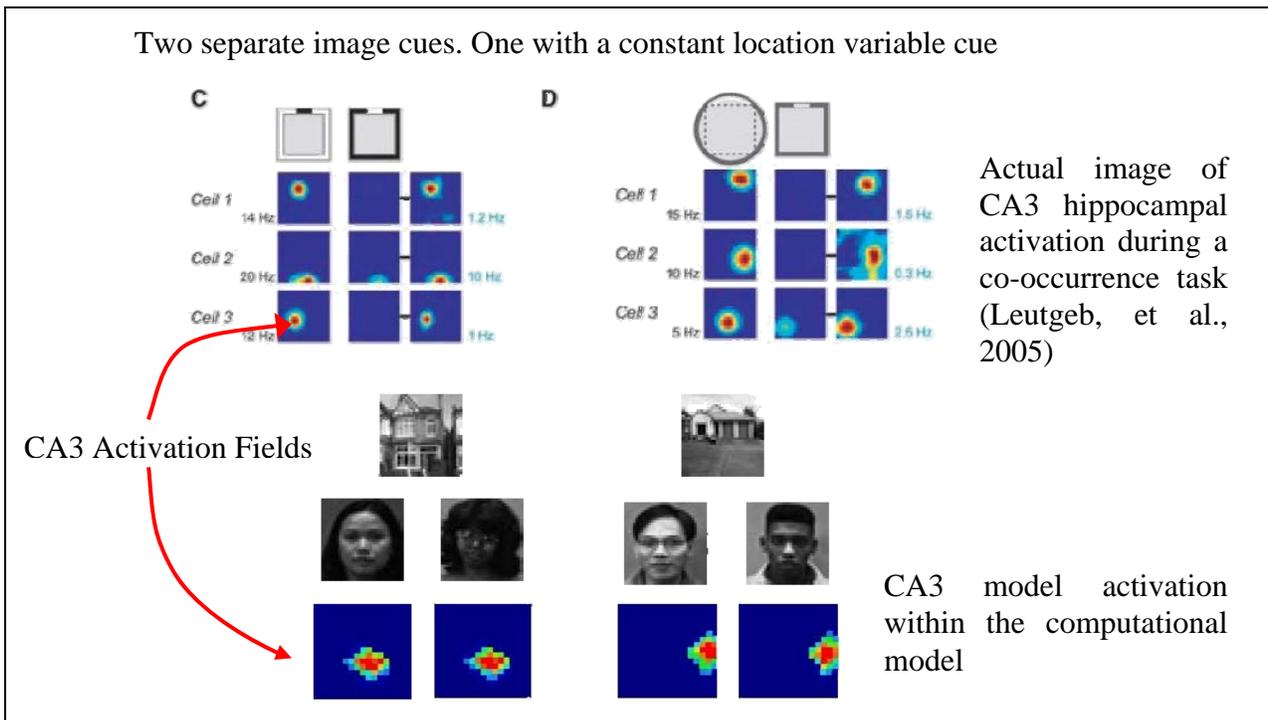


Figure 21. Experimental and model image of CA3 hippocampal activation during a co-occurrence task

5. GENERAL DISCUSSION AND CONCLUSIONS

The progression of our computational model is driven by attempts to improve model fidelity in relation to neurobiology. Rather than striving to implement the most efficient machine learning algorithms to achieve a desired goal, our approach has been to model the neuroanatomy and processes underlying declarative memory and recall. In doing so, we have demonstrated the ability to model elements of cognitive behavior such as familiarity and recognition.

As a result of continuous improvement to the model we are also able to create automatic associations of various semantic concepts. Additionally we have presented mutual information as a means of quantitatively analyzing the dependence between semantic concepts within the CA3 region of the model. Overall, information theoretic analysis provides a mathematically rigorous means of analyzing the information storage and propagation capabilities of a model in an implementation dependent manner. In general, the artificial neural network computation model we have presented processes sensory inputs and in effect is capable of exhibiting qualitative memory phenomena such as auto-association of episodic memory concepts.

We have made both a neurophysiological and a psychological behavioral case for our model. We assert that this approach is of great potential benefit to the field because it puts computational modelers and neuropsychological investigators into interdisciplinary communication. By engineering a structural neuro-cognitive model, we have highlighted areas where neuroscience could most profitably shine the light of discovery to push our understanding further forward. For instance, all information that traverses through our model goes through the entorhinal cortex to dentate gyrus connections. That connectivity scheme was modeled based on our best anatomical understanding, but what is the merit of bringing together all of the modalities before they are hippocampally bound? Why is this evolutionarily more valuable than retaining the higher information content possible with separate modalities? This is an area in need of a neuroscientific theoretical approach and an answer could in turn, help us to construct a more veridical, powerful and explanatory model. We believe that our model provides the experimentalist with a useful tool to explore cognitive processes. The behavioral effects we suggest should be confirmed in human subjects, but the model can be used to run exhaustive trials that would not be plausible for human studies. As this model continues to be developed, the computational-to-human study paradigm will only become more attractive and the potential for interdisciplinary collaboration more alluring. This is exemplified in the statement by Neal Cohen, professor and Director of the Head, Brain & Cognition division at the Department of Psychology, & Beckman Institute for Advanced Science and Technology, & Neuroscience Program:

This model supports the ability to do classification/categorization of a range of visual inputs, to remember the prior occurrence of each of those inputs individually, to do pattern completion permitting recovery of those items based on partial or incomplete cues, to represent different locations in the visual environment, to remember which individual items occurred in which locations, and to bind together in memory representations of any arbitrary collection of items with one another and with their spatial or other contexts. And all of these capabilities are implemented in a model with biological realism greater than in any previously implemented model. Finally, it is done in a way that permits us to test the contributions of each of the individual components of the model and to compare that with what is seen in humans and animals.

We believe this work will also benefit the Science and Defense national security mission of the DOE and other federal agencies by increasing the understanding of key aspects of cognition as well as creating a higher fidelity human modeling architecture. This will enable the DOE to better understand the thought processes underlying human behavior, as well as enhance human modeling in areas such as action/counter-action predictive simulations, training, and assistive decision making.

5.1 Model Limitations

Although our model is built upon understood neuroanatomical hippocampal function using biologically plausible computational mechanisms, it is not an identical reproduction of neural anatomy and function. Our model is not an exact neuron for neuron replica of the HC. Indeed, even if we had the computational resources to implement it, a reference map of every neuron and synapse in biological sensory cortex and hippocampus does not exist. Our work provides evidence for some specific connection schemes that we consolidated from the best existing literature. Future models can iterate and improve upon our assumptions. While not implementing the absolute volume of neural nodes in modeled biological structures, the model does take into consideration the neuron density and type within distinct regions, and attempts to preserve the same ratios in allocating computational resources.

In terms of scope, the model is constrained to sensory cortex, parahippocampus and HC. This partial neural representation does not include an output modality, consequently constraining the means by which we may query and test the model. As addressed in the relevant sections describing the means by which we have tested the model, we have accounted for this limitation by constraining the means in which we extract information from the model. Rather than simply making inferences regarding model performance or knowledge based upon the underlying computational implementation, we have restricted our analysis to mechanisms such as neural activation which is somewhat analogous to brain imaging approaches.

As the fidelity of the model is not at the neuron level, likewise it does not operate via action potentials. Rather, our model requires a clocking system regulating the flow of information through the model. This seemed a reasonable abstraction as we are running the model on digital computers anyway. The temporal integrators through the system do provide a means of buffering up a sequence of inputs, but include a design tradeoff impairing the ability to encode a sequence containing a repeated input separated by a different input. The temporal integrator functionality, as described formerly, decays the activation value of a category representation over time. However, the activation is replenished upon subsequent presentations of the same input. For example, while an input sequence of ABC could correspond to an integrated vector output $\langle 0.5, 0.75, 1 \rangle$, a sequence which repeats such as ABA is indistinguishable from BA (and various other possibilities). Computationally there are several simple means of compensating in a non-biological manner, but that would contradict the design intentions of this project. We see two possibilities for reconciling this approach with the biology. First, the biology may implement a more complex temporal sequence encoding scheme. We implemented a fairly simple scheme in part through a desire not to make any unreasonable demands upon what biological neural networks might be capable of. Second, the inability to distinguish ABA from AB may not be an issue at sufficiently abstract conceptual levels. It seems a reasonable claim that you can never have exactly the same experience twice, therefore the brain will never see exactly the same pattern of activations twice, hence "A" will never repeat as in ABA.

6. REFERENCES

1. Addis, D.R., Moscovitch, M., Crawley, A.P., and McAndrews, M.P. (2004) Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus* 14: 752-762.
2. Adlam, A.R., Vargha-Khadem, F., Mishkin, M., and de Haan, M. (2005) Deferred imitation of action dequences in developmental amnesia. *Journal of Cognitive Neuroscience* 17: 240-248.
3. Aggleton, J.P., Kyd, R.J., & Bilkey, D.K. (2004). When is the perirhinal cortex necessary for the performance of spatial memory tasks? *Neuroscience & Biobehavioral Reviews*, 28, 611-24.
4. Aggleton J.P., Vann S.D., Denby C., Dix S., Mayes A.R., et al. (2005) Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia* 43(12):1810-1823.
5. Alvarado, M.C., & Bachevalier, J. (2005). Comparison of the effects of damage to the perirhinal and parahippocampal cortex on transverse patterning and location memory in rhesus macaques. *Journal of Neuroscience*, 25, 1599-1609.
6. Amaral, DG, Witter, MP. 1989. The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience* 31: 571-591
7. Amaral, D. G., & Witter, M. P. (1995). Hippocampal formation. In G. Pacinos (Ed.), *The Rat Nervous System*, (2nd ed., pp. 443-493). San Diego, CA: Academic Press.
8. Andersen, R.A. (1989) Visual and eye movement functions of the posterior parietal cortex. *Annual Review of Neuroscience* 12:377-403.
9. Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum
10. Annon, A. (2009). Episodic memory modeled by an integrated cortical-hippocampal neural architecture, *Human Behavior-Computational Modeling and Interoperability*, Oak Ridge, TN.
11. Annon, A. (2009). Temporal semantics: An adaptive resonance theory approach, *Proceedings of the 2009 International Joint Conference on Neural Networks*. Atlanta, GA.
12. Aristotle (350 BC). *On Memory and Reminiscence*. (J. I. Beare, Trans.).
13. Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4, 417-423.
14. Bar, M. & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, 38, 347-58.

15. Barnes, J. M., & Underwood, B. J. (1959). Fate of first-list associations in transfer theory. *Journal of Experimental Psychology*, 58, 97–105.
16. Best, P.J, White, A.M., and Minai, A. (2001) Spatial processing in the brain: the activity of hippocampal place cells. *Annual Review of Neuroscience* 24, 459-86.
17. Bliss, T. V. P., Collinridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature*, 361, 31-39.
18. Brown, M.W. & Xiang, J.Z. (1998). Recognition memory: Neuronal substrates of the judgment of prior occurrence. *Progress in Neurobiology*, 55, 149-89.
19. Brown, M.W. & Aggleton, J.P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, 2, 51-61.
20. Buckmaster, C.A., Eichenbaum, H., Amaral, D.G., Suzuki, W.A. and Rapp, P. (2004) Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. *Journal of Neuroscience* 24: 9811-9825.
21. Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, 379, 255-257.
22. Burwell, R.D., Witter, M.P. & Amaral, D.G. (1995). Perirhinal and postrhinal cortices of the rat: a review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus*, 5(5), 390-408.
23. Burwell, R.D., & Hafemanm, D.M. (2003). Positional firing properties of postrhinal cortex neurons. *Neuroscience*, 119, 577-88.
24. Cabeza, R. and St. Jaques, P. (2007) Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences* (in press).
25. Carpenter G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine, *Computer Vision, Graphics, and Image Processing*, 37, 54–115.
26. Carpenter, G. A., Grossberg, S. & Rosen. D. B. (1991). FuzzyART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, 4:759–771.
27. Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nature Reviews Neuroscience*, 4, 685-691.

28. Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., Nash, C. (1999). Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus*, 9: 83-98.
29. Cohen, N. J. & Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System*. Cambridge, MA: M.I.T. Press.
30. Cohen, N.J. & Squire, L.R. (1980). Preserved learning of pattern-analyzing skill in amnesia: Dissociation of "knowing how" and "knowing that." *Science*, 210, 207-210
31. Cover, T. M. *Elements of information theory*. Hoboken, N.J: J. Wiley, 2005.
32. Daselaar S.M., Fleck M.S, & Cabeza R. (2006) Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *Journal of Neurophysiology* 96:1902-1911.
33. Davachi, L. and Wagner, A. G. (2002). Hippocampal contributions to episodic encoding, Insights from relational and item-based learning. *Journal of Neurophysiology*, 88, 982-990.
34. Davachi, L., Mitchell, J. P. and Wagner, A. D. (2003). Multiple routes to memory, Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*. 100, 2157-2162.
35. Day, M., Langston, R., & Morris, R. G. M. (2003). Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. *Nature*, 424, 205-209.
36. Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron*, 35, 989-996.
37. Downes, J. J., Mayes, A. R., MacDonald, C., & Humkin, N. M. (2002). Temporal order memory in patients with Korsakoff's syndrome and medial temporal amnesia. *Neuropsychologia*, 40, 853-861.
38. Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Science, U.S.A.*, 94, 7109-7114.
39. Eichenbaum, H. & Cohen, N.J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. Oxford: Oxford University Press.
40. Eichenbaum, H. (2007). Declarative memory: Insights from cognitive neurobiology. *Annual. Review of Psychology*, 48, 547-572.

41. Eichenbaum, H., P. Dudchencko, E. Wood, M. Shapiro and H. Tanila (1999) The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron* 23: 209-226
42. Eichenbaum H. 2004 Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron* 44:109-20.
43. Eichenbaum, H., Yonelinas A.R., and Ranganath, C. (2007) The medial temporal lobe and recognition memory. *Annual Review of Neuroscience* (In press).
44. Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, 425, 184-87.
45. Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Brookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, 3, 1149-1152.
46. Elman, J. L. (1990). Finding structure in time, *Cognitive Science*, 14, 179–211, 1990.
47. Ergorul, C., & Eichenbaum, H. (2004). The hippocampus and memory for “What”, “Where”, and “When”. *Learning and Memory*, 11, 397-405.
48. Ferbinteanu, J. & Shapiro, M.L. (2003) Prospective and retrospective memory coding in the hippocampus. *Neuron* 40: 1227-1239.
49. Fink, G.R., Markowitsh, H.J., Reinkemeier, M., Bruckbauer, T., Kessler, J., and Heiss, W.-D. (1996). Cerebral representation of one’s own past: neural networks involved in autobiographical memory. *Journal of Neuroscience* 16: 4275-4282.
50. Fortin, N. J., Agster, K. L., & Eichenbaum, H. (2002). Critical Role of the Hippocampus in Memory for Sequences of Events. *Nature Neuroscience*, 5, 458-462.
51. Frank, L. M., Brown, E. N. & Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* 27: 169-178.
52. Fuster, J. M. (1995). *Memory in the cerebral cortex*. Cambridge, MA: M.I.T. Press.
53. Gaffan, E. A., Healey, A. N., & Eacott, M. J. (2004). Objects and positions in visual scenes: effects of perirhinal and postrhinal cortex lesions in the rat. *Behavioral Neuroscience*, 118, 992-1010.
54. Gilbert, P. E., & Kesner, R. P. (2003). Localization of function within the dorsal hippocampus: the role of the CA3 subregion in paired-associate learning. *Behavioral Neuroscience*, 117, 1385-1394.

55. Giovanello, K. S., Verfaellie, M., and Keane, M. M. (2003). Disproportionate deficit in associative recognition relative to item recognition in global amnesia. *Cognitive, Affective, and Behavioral Neuroscience* 3: 186-194.
56. Giovanello, K. S., Schnyer, D. M., and Verfaellie, M. (2003) A critical role for the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus* 14: 5-8.
57. Hampson, R. E., Heyser, C. J., and Deadwyler, S. A. (1993) Hippocampal cell firing correlates of delayed-match-to-sample performance in the rat. *Behavioral Neuroscience*, 107, 715-739.
58. Hampson, R. E., Pons, T. P., Stanford, T. R., Deadwyler, S. A. (2004) Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory.
59. Hannula, D. E., Ryan, J. D., Tranel, D. & Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *Journal of Cognitive Neuroscience*, 19(10):1690–1705.
60. Hannula, D. E.; Tranel, D.; Cohen, N. J. (2006). The long and the short of it: Relational memory impairments in amnesia, even at short lags, *Journal of Neuroscience* 26, (32), 8352-8359.
61. Hargreaves, E. L., Rao, G., Lee, I., & Knierim, J. J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science*, 308, 1792-4.
62. Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron*, 37, 877-888.
63. Haykin, Simon S. (1999). *Neural networks a comprehensive foundation*. Upper Saddle River, N.J: Prentice Hall.
64. Heckers, S., Zalezak, M., Weiss, A. P., Ditman, T., and Titone, D. (2004) Hippocampal activation during transitive inference in humans. *Hippocampus* 14, 153-162.
65. Henson, R. N., Cansino, S., Herron, J. E., Robb, W. G., & Rugg, M. D. (2003). A familiarity signal in human anterior medial temporal cortex? *Hippocampus*, 13, 301-304.
66. Honey, R. C., Eatt, A. & Good, M. (1998). Hippocampal lesions disrupt an associative mismatch process. *Journal of Neuroscience* 18 :2226-2230.
67. Hopkins, R. O. & Kesner, R. P. (1995). Item and order recognition memory in subjects with hypoxic brain injury. *Brain and Cognition*, 27, 180-201.

68. James, W. (1890). *The Principles of Psychology*. (1918 ed.). New York: Holt.
69. Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of Neural Science*. New York: McGraw-Hill Co.
70. Kanwisher, N., Downing, P., Epstein, R., and Kourtzi, Z. (2007) Functional neuroimaging of visual recognition. In R. Cabeza & A. Kingstone, eds. *Handbook of Functional Neuroimaging of Cognition*. MIT Press: Cambridge. pp. 109-152.
71. Kesner, R. P., Gilbert, P. E., and Barua, L. A. (2002). The role of the hippocampus in memory for the temporal order of a sequence of odors. *Behavioral Neuroscience*, *116*, 286-290.
72. Kesner, R.P., Hunsaker, M.R., & Gilbert, P.E. (2005). The role of CA1 in the acquisition of an object-trace-odor paired associate task. *Behavioral Neuroscience*, *119*, 781-786.
73. Kingsley, R (2000). *Concise Text of Neuroscience*. Lippincott, Williams & Wilkins.
74. Kohonen, T. (2001). *Self-Organizing Maps*. Third, extended edition. Springer.
75. Kreiman, K., Kock, C., and Fried, I. (2000a). Category specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, *3*, 946-953.
76. Kreiman, K., Kock, C., & Fried, I. (2000b). Imagery neurons in the human brain. *Nature* *408*: 357-361.
77. Krubitzer, L., & Kaas, J. (2005). The evolution of the neocortex in mammals: how is phenotypic diversity generated? *Current Opinions in Neurobiology*, *15*, 444-453.
78. Kuhn, G., Watrous, R. L. & Ladendorf, B. (1990). Connected recognition with a recurrent network, *Speech Communication*, *9*, 41-48.
79. Kumaran, D. & Maguire, E. A. (2006). The dynamics of hippocampal activation during encoding of overlapping sequences. *Neuron*, *49*, 617-629.
80. Lawrence, S., Giles, C. L., & Fong, I. (1996). Can recurrent neural networks learn natural language grammars, *Proceedings of the IEEE International Conference on Neural Networks*, 1853-1858.
81. Leutgeb, et al. Independent Codes for Spatial and Episodic Memory in Hippocampal Neuronal Ensembles. *Science* *309*, 619 (2005)
82. Maguire E. A, (2001). Neuroimaging studies of autobiographical events memory. *Philosophical Transactions of the Royal Society of London, Series B*: *356*, 1441-1452.
83. Manns J. R and Eichenbaum H. (2006). Evolution of the hippocampus. In J.H. Kaas, ed. *Evolution of Nervous Systems*. Vol 3. Academic Press: Oxford pp. 465-490.

84. Martin, A. (2007). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone, eds. *Handbook of Functional Neuroimaging of Cognition*. MIT Press: Cambridge. pp. 153-186.
85. McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419-457
86. McDonough, L., Mandler, J. M., McKee, R. D., and Squire, L. R. (1995) The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Sciences* *92*: 7580-7584.
87. Mountcastle, V. B., J. C. Lynch, A. Georgopoulos (1975) Posterior parietal association cortex of the monkey. Command functions for operations within personal space. *Journal of Neurophysiology* *38*:871-908.
88. Muller, R. U., Kubie, J. L., & Ranck, J. B., Jr. (1987). Spatial firing patterns of hippocampal complex spike cells in a fixed environment. *Journal of Neuroscience* *7*, 1935-1950.
89. Muller, R. U, Poucet, B., Fenton A. A., & Cressant, A. (1999). Is the hippocampus of the rat part of a specialized navigational system? *Hippocampus*. *9*, 413-22.
90. Mumby, D. G., & Pinel, P. J. (1994). Rhinal cortex lesions and object recognition in rats. *Behavioral Neuroscience*, *108*, 11-18.
91. Mumby, D. G. (2001) Perspectives on object recognition memory following hippocampal damage: lessons from studies on rats. *Behavioural Brain Research* *127*, 159-181.
92. Mumby, D. G., Gaskin, S., Glenn, M. J., Scharamek, T. E. and Lehmann, H. (2002) Hippocampal damage and exploratory preferences in rats: memory for objects, place, and contexts. *Learning and Memory* *9*: 49-57.
93. Norman, G., & Eacott, M. J. (2005). Dissociable effects of lesions to the perirhinal cortex and the postrhinal cortex on memory for context and objects in rats. *Behavioral Neuroscience*, *119*, 557-66.
94. Northoff, G. & Bermpohl, F. (2004) Cortical midline structures and the self. *Trends in Cognitive Sciences* *8*:102-107.
95. O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition." *Science*, 91-94.

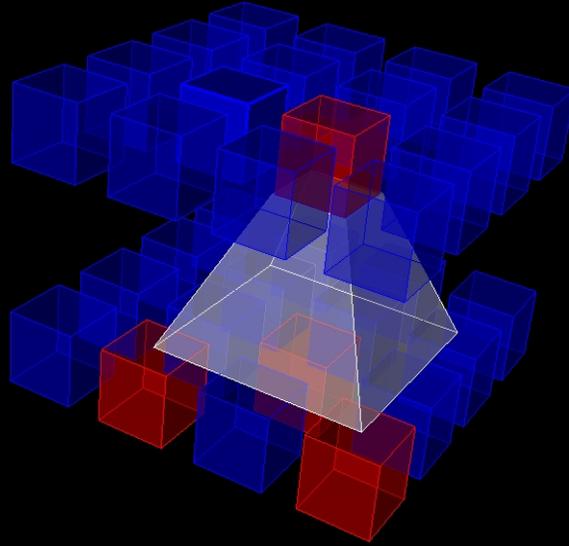
96. O'Reilly, R. C., Rudy, J. W. (2001). Conjunctive representations, the hippocampus, and contextual fear conditioning, *Cognitive Affective Behavior Neuroscience*, 1, 1, 66-82.
97. O'Reilly, Randall C., & Yuko Munakata. (2000). *Computational Explorations in Cognitive Neuroscience Understanding*. MIT Pr.
98. Otto, T., & Eichenbaum, H. (1992). Complementary roles of orbital prefrontal cortex and the perirhinal-entorhinal cortices in an odor-guided delayed non-matching to sample task. *Behavioral Neuroscience*, 106, 763-776.
99. Pihlajamaki, M., Tanila, H., Kononen, M., Hanninen, A., Soininen, H., & Aronen, H.J. (2004). Visual presentation of novel objects and new spatial arrangements of objects differentially activates the medial temporal lobe areas in humans. *European Journal of Neuroscience*, 19, 1939-49.
100. Preston, A. R. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, 148-52.
101. Preston, A., Shrager, Y., Dudukovic, N.M. & Gabrieli, J.D.E. (2004) Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, 14, 148-152.
102. Quiroga R.Q., Reddy L., Kreiman G., Koch C., Fried I. (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102-1107.
103. Ranganath, C., Yonelinas, A.P., Cohen, M.X., Dy, C.J., Tom, S.M., and D'Esposito, M.D. (2003) Dissociable correlates of recollection and familiarity with the medial temporal lobes. *Neuropsychologia* 42: 2-13.
104. Rolls, E. T., & Kesner, R. P. (2006). A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology*, 79, 1-48.
105. Rosenbloom, P. S., Laird, J. E., & Newell, A. (editors). *The Soar Papers: Research on Integrated Intelligence*. MIT Press, Cambridge, MA, 1993.
106. Scoville, W. B., & Milner, B. (2000). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neuropsychiatry Clinical Neuroscience* 12, 103-13.
107. Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce english text," *Complex Systems*, 1, 145-168.
108. Shastri L. & Fontaine, T. (1995). Recognizing Hand-printed digit strings using modular spatio-temporal connectionist networks, *Connection Science*, 7, 211-235.
109. Sperling, R., Chua, E., Cocchiarella, A., Rand-Giovannetti, E., Poldrack, R., Schacter, D.L, and Albert, M. (2003) Putting names to faces: Successful encoding of associative memories activates the anterior hippocampal formation. *NeuroImage* 20: 1400-1410.
- 110.

111. Spiers, H. J., Burgess, N., Hartley, T., Vargha-Khadem, F., & O'Keefe, J. (2001). Bilateral hippocampal pathology impairs topographical and episodic memory but not visual pattern matching. *Hippocampus*, *11*, 715-725.
112. Sun, R. & Giles, C. (2001). Sequence learning: from recognition and prediction to sequential decision making, *IEEE Intelligent Systems*, *16*, 67-70.
113. Suzuki, W. A., Zola-Morgan, S., Squire, L.R., & Amaral, D.G. (1993). Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactual modalities. *Journal of Neuroscience*, *13*, 2430-51.
114. Suzuki W. A. & Amaral D. G. (1994) Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology* *350*, 497-533.
115. Suzuki, W., & Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the NY Academy of Sciences*, *911*, 175-91.
116. Svarer, C., Hansen, L. K., & Larsen, J. (1993). On design and evaluation of tapped-delay neural network architectures, *Neural Networks*, 46-51.
117. Suzuki W.A. & Amaral D.G. (1994) Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology* *350*, 497-533.
118. Suzuki, W., & Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the NY Academy of Sciences*, *911*, 175-91.
119. Taylor, S. E., Bernard, M. L. Caudell, T. P., Cohen, N. J. Healy, M. J., Morrow, J. D., Verzi, S. J., Vineyard, C. M., & Watson, P. (2009). Memory in silico: Building a neuromimetic episodic cognitive model. *Proceedings of the 2009 International Joint Conference on Neural Networks*. Atlanta, GA.
120. Taylor, S. E., Bernard, M. L. Caudell, T. P., Cohen, N. J. Healy, M. J., Morrow, J. D., Verzi, S. J., Vineyard, C. M., & Watson, P. (2009) Memory in silico: Building a neuromimetic episodic cognitive model. *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering*. Los Angeles, California USA. 733-37.
121. Taylor, S. E., Healy, M. J., and Caudell, T. P. (2007). Categorical mapping from ontology to neural network: Initial studies of simple neural networks' concept capacity, *Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, 2020-2025*.
122. Taylor, S. E., Bernard, M. L., Verzi, S. J., Morrow, J. S., Vineyard, C. M., Healy, M. J. & Caudell, T. P. (2009). Temporal semantics: An adaptive resonance theory approach." *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering*. Los Angeles, California USA. 733-37.
123. Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

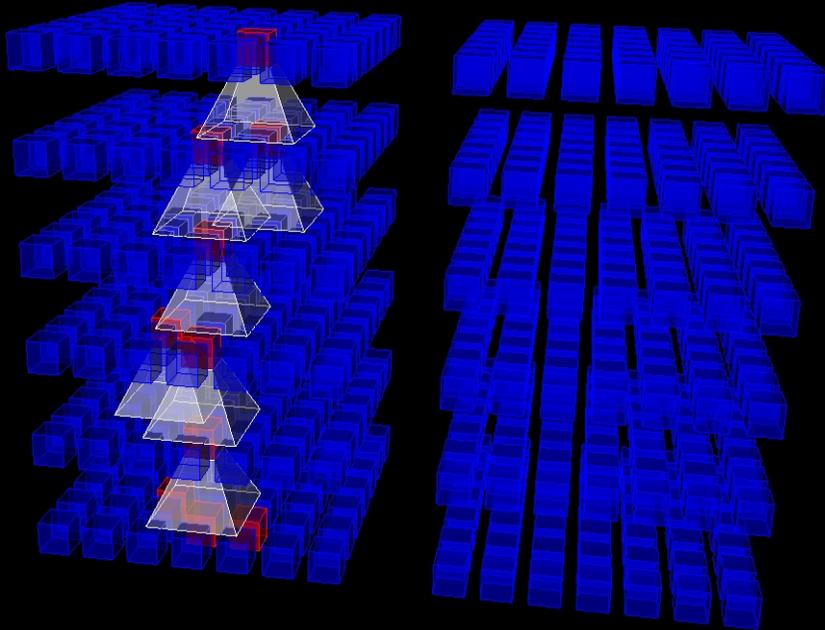
124. Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1-25.
125. Turriziani, P., Fadda, L., Caltagirone, C., and Carlesimo, G. A. (2004) Recognition memory for single items and associations in amnesia patients. *Neuropsychologia* 42: 426-433.
126. Uncapher M. R., Otten L. J., Rugg M. D. (2006) Episodic encoding is more than the sum of its parts: an fMRI investigation of multifunctional contextual encoding. *Neuron* 52:547-556.
127. Vargha-Khadem, F., Gadin, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W. & Mishkin, M. (1997). Differential Effects of Early Hippocampal Pathology on Episodic and Semantic Memory. *Science*, 277, 376-80.
128. Vila, L. (1994). A survey on temporal reasoning in artificial intelligence, *AI Communications*, 7, 4-28.
129. Waibel, A., Hanazawa, T., Hinton, K., Shikano, & Lang, J. (1989). Phoneme recognition using time-delay neural networks, *Acoustics, Speech, and Signal Processing* [see also IEEE Transactions on Signal Processing], IEEE Transactions on, vol. 37, no. 3, pp. 328-339.
130. Walczak, S. (2005). Artificial neural network medical decision support tool: Predicting transfusion requirements of ER patients, *Information Technology in Biomedicine, IEEE Transactions on*, 9, 468-474
131. Wan, E. A., (1994). *Time series prediction by using a connectionist network with internal delay lines*, In *Time Series Prediction*. Addison-Wesley, pp. 195-217.
132. Wan, H., Aggleton, J.P., Brown, M.W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *Journal of Neuroscience*, 19, 1142-48.
133. Wood E, Dudchenko PA, Eichenbaum H. (1999). The global record of memory in hippocampal neuronal activity. *Nature* 397: 613-16.
134. Wood E, Dudchenko P, Robitsek JR, Eichenbaum H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27, 623-33.
135. Yonelias, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauve, M. J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, 5, 1236-41.
136. Zeineh, M. M., Engel, S. A., Thompson, P. M., and Brookheimer, S. Y. (2003) Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science* 299: 577-580.

APPENDIX A: LAYOUT AND CONNECTION STRUCTURE OF THE MODEL

Modules from higher columns receive inputs from multiple modules of lower columns



Currently: 6 layers at 7X7 modules



**APPENDIX B: THE INPUT SET PRESENTED TO THE MODEL IN
COMPARISON WITH THE PRESTON STUDY**



DISTRIBUTION

- 4 Lawrence Livermore National Laboratory
Attn: N. Dunipace (1)
P.O. Box 808, MS L-795
Livermore, CA 94551-0808
- 1 Thomas Caudell, PhD
P.O. Box 1629
Edgewood, NM 87015
- 1 Neal J Cohen, PhD
3214 Lakeshore Drive
Champaign, IL 61822
- 1 Howard Eichenbaum, PhD
29 Wareland Road
Wellesley, MA, 02481
- 1 Mark McDaniel, PhD
375 S. Gore Ave
St. Louis, MO 63119
- 1 Patrick Watson
210 W. John St. Apt C.
Champaign, IL, 61820

1	MS0316	R. Schiek	1437
1	MS0321	J. Peery	01400
1	MS0321	J. Mitchiner	01430
1	MS0370	G. Backus	01433
1	MS0370	L. McNamara	01433
1	MS0370	T. Trucano	1411
1	MS0484	R. Skocypec	08004
1	MS0638	D. Peercy	0421
1	MS0721	C. Duus	06343
1	MS0830	C. Dornburg	12335
1	MS0830	K. Diegert	12335
1	MS0933	S. Taylor	09511
1	MS0839	C. Johnson	07000
1	MS1004	M. Skroch	06344
1	MS1004	P. Xavier	06344
1	MS1005	A. Nanco	06345
1	MS1009	D. Small	06474

5	MS1188	M. Bernard	06341
1	MS1188	M. Glickman	06343
1	MS1188	D. Morrow	06341
1	MS1188	E. Raybourn	06345
1	MS1188	F. Rothganger	06341
1	MS1188	W. Shanyfelt	06341
1	MS1188	G. Sleafte	06340
1	MS1188	A. Speed	06341
1	MS1188	S. Verzi	06343
1	MS1188	J. Wagner	06341
1	MS1188	J. Whetzel	06343
1	MS1217	I. Dubicka	05925
1	MS1316	M. D. Rintoul	01412
1	MS1318	C. Phillips	01415
1	MS1318	S. Rountree	01415
1	MS9406	D.Djordjevich	08116
1	MS0899	Technical Library	9536 (electronic copy)
1	MS0123	D. Chavez, LDRD Office	1011
1	MS0161	Patent and Licensing Office	11500



Sandia National Laboratories