A Model of Rapid Memory Formation in the Hippocampal System

Lokendra Shastri (shastri@icsi.berkeley.edu)
International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, CA 94704 USA

Abstract

Our ability to remember events and situations in our daily life demonstrates our ability to rapidly acquire new memories. There is a broad consensus that the hippocampal system (HS) plays a critical role in the formation and retrieval of such memories. A computational model is described that demonstrates how the HS may rapidly transform a transient pattern of activity representing an event or a situation into a persistent structural encoding via long-term potentiation and long-term depression.

Introduction

Our ability to remember events and situations in our daily life and acquire facts after reading a newspaper demonstrates our ability to rapidly acquire new memories. This form of memory has been the focus of considerable research in psychology and cognitive neuroscience and has been characterized variably as declarative, *locale*, and explicit. There is a broad consensus that this form of memory is distinct, both in its functional properties and its neural basis, from other forms of memories such as memories of perceptual-motor skills, priming, and classical conditioning (for a review see Cohen & Eichenbaum, 1993; Squire, 1992).

Memories of events and situations are acquired rapidly. It is reasonable to assume that the construal of an experience in terms of an event or a situation is initially expressed as a pattern of activity over neural structures. This expression, however, is per force transient, and hence, the neural encoding of a memorable event or situation must be transformed rapidly from a transient pattern of activity into a persistent structural encoding, or else it would be lost.

A battery of neuropsychological, neuroanatomical, neurophysiological, and imaging data suggests that the hippocampal system (see below) plays a critical role in the encoding and recall of events and situations. Several studies have shown that human patients with bilateral damage to the HS suffer from severe amnesia and are unable to remember events that occurred just a few minutes ago (e.g., Scoville & Milner, 1957). Such patients, however, can still acquire procedural skills and demonstrate priming effects. Studies of animal models (e.g., O'Keefe & Nadel, 1978; Squire & Zola-Morgan, 1991) also provide support for the putative role of the HS.

A number of researchers have proposed models to explain and understand the functionality of the HS based memory system. These include system-level models that attempt to describe the functional role of the HS (e.g., Cohen & Eichenbaum, 1993; Squire & Zola-Morgan, 1991) as well as computational models that attempt to explicate how the HS might realize its putative function (e.g., Marr, 1971; Treves & Rolls, 1994; Lynch & Granger, 1992; Gluck & Myers, 1993; Hasselmo, 1997; Schmajuk & DiCarlo, 1992; O'Riley & McClelland, 1994). This work has greatly enhanced our understanding of the HS and its role in memory formation and retrieval, but it has not dealt with some critical representational problems associated with the encoding and retrieval of specific events and situations. We discuss some of these problems below.

Representational Requirements of Encoding Events and Situations

Typically, memories of events and situations record who did what to whom where and when. Alternately, they may describe a state of affairs wherein multiple entities occur in a particular configuration or relationship, or record the state of an entity. In each of these cases an event or a situation may be viewed as a *relational instance* consisting of a collection of *bindings* between the *roles* of a generic relation and the *entities* that *fill* these roles in the given event or situation. For example, the event "John gave Mary a book on Tuesday" may be viewed as an instance of the generic relation GIVE with the *role*-`entity` bindings:

(GIVE: ⟨*giver*=`John`⟩,
⟨*recipient*=`Mary`⟩,
⟨*give-object*=`a-Book`⟩,
⟨*temporal-location*=`Tuesday`⟩)

There exists a vast body of work in traditional as well as cognitive linguistics that demonstrates how various aspects of conceptual knowledge can be expressed using appropriate relational structures composed of role-entity bindings. Such structures have been variably referred to as frames, schemas, scripts, and predicates.

The fact that an event or a situation is essentially a collection of role-entity bindings gives rise to a number of representational requirements.

First, it entails that a memory of events and situations must be capable of encoding and subsequently detecting role-entity bindings. A memory that only binds together the entities that occur in an event, but does not encode which entity fills which role, cannot function properly since it cannot distinguish between events such as "John gave Mary a book" and "Mary gave John a book". Observe that these two events are distinct even though they involve the same roles and entities because some of the entities fill different roles in the two events.

Second, the encoding of an event or a situation should respond positively to partial cues, but at the same time, it must respond negatively to a cue that specifies incompatible bindings — even if the cue is otherwise highly similar to the memorized event. For example, while the encoding of the event "John gave Mary a book in front of the library on Tuesday" must respond positively to the partial cue "Did John give Mary a book?", it must respond negatively to the highly similar but erroneous cue "Did Susan give Mary a book in front of the library on Tuesday?" These two requirements — the recognition of partial cues and the rejection of similar but erroneous cues — together entail that the encoding of an event or a situation should be capable of actively detecting *errors* (mismatches) between the bindings specified in a cue and those specified in the memorized event or situation. An encoding of the event "John gave Mary a book on Tuesday" that only detects binding matches and cannot detect binding errors will be unable to distinguish between the erroneous cue "Did John give Mary a book on Friday?" and the partial but matching cue "Did John give Mary a book?" To this encoding, the partial and the erroneous cues would appear similar since both contain *three* matching bindings. Hence the encoding of an event or a situation must also incorporate *binding error detectors*.

Third, the encoding of an event or a situation should support recall and respond to wh-queries by *selectively* retrieving entities that fill a specified role in the memorized event or situation. For example, the encoding of the event "John gave Mary a book on Tuesday" must selectively activate "Mary" in response to the wh-query "To whom did John give a book on Tuesday?" Hence the encoding must also include *binding extractor* circuits that can activate entities that fill specific roles within the memorized event or situation.

To summarize, the memorization of an event or a situation requires the rapid formation of: binding detectors, binding error detectors, circuits for integrating the responses of these detectors, and circuits for extracting role-fillers from bindings. Existing HS based memory models as well as purely computational models of rapid memorization proposed by Feldman (1982) and Valiant (1994) do not satisfy these representational requirements.

A Model of Memory Formation in the HS

The proposed model SMRITI[1] (Shastri, 1997) addresses the representational requirements discussed above and demonstrates how a system like the HS might rapidly transform a transient pattern of activity representing an event or a situation into a persistent encoding capable of supporting recognition and recall. This transformation leads to the rapid formation of distributed structures for detecting bindings and binding errors, integrating the outputs of these detectors, and performing binding extraction. The resulting encoding can recognize highly partial patterns, exhibit a high-degree of pattern separation, and respond to wh-queries.

While it is relatively straightforward to imagine how the HS might learn isolated binding detectors and binding extractors using associative learning, the concurrent learning of such detectors for all of the role-entity bindings pertaining to an event is problematic since cross-talk among active roles and entities can lead to the formation of spurious binding detectors and extractors.

The formation of binding error detectors is even more problematic given their paradoxical behavior. The crux of the problem is this: The formation of a binding error detector for the binding of a role r and an entity f must occur in response to the coincident activity of r and f. But subsequent to its formation, the binding error detector must *not* fire anymore in response to the coincident activity of r and f — *the very activity that led to its formation*. Instead, it must fire in response to the firing of r in the *absence* of the coincident firing of f. It is not obvious how such a detector might be learned rapidly within a neural circuit. One of the contributions of the present work is that it demonstrates how circuits that behave like binding error detectors can be learned rapidly within the HS via long-term potentiation (LTP) and long-term depression (LTD).

The model's architecture parallels the circuitry of the HS and provides a rationale for various components of the HS and their idiosyncratic interactions. It also predicts the memory deficits that would result from selective damage to components of the HS.

The encoding is sparse, but at the same time, it is physically distributed and redundant. While the sparseness of the encoding enables the model to memorize a large number of events, the physically distributed and redundant nature of the encoding makes the model robust against significant amounts of cell loss.

After a brief review of the HS architecture and the LTP and LTD phenomena, the paper describes the proposed model emphasizing its functional architecture and the mapping between its components and those of the HS. Finally, the paper lists some behavioral deficits predicted by the model. Limited space precludes a circuit level description of the model; such a description may be found in Shastri (1997).

The Hippocampal System

The hippocampal system (HS) refers to a collection of medial temporal lobe structures consisting of the entorhinal cortex (EC) and the hippocampal formation (HF). The HF in turn consists of the Ammon's horn, the dentate gyrus (DG) and the subicular complex (SC). Ammon's horn and DG together form a distinctive sea-horse shaped structure that arches around the mesencephalon and is referred to as the hippocampus. Ammon's horn in turn consists of distinct regions labeled CA1, CA2, and CA3.

Figure 1 depicts a schematic of the major pathways interconnecting the components of the HS. The EC acts as the principal gateway between the HS and other cortical areas; it funnels cortical outputs into the HF and in turn, relays the output of the HF back to cortical areas. EC receives direct and massive projections from higher-order polymodal associational areas (e.g., Van Hoesen, 1982) as well as major projections from the perirhinal and parahippocampal cortices. The latter in turn receive inputs from higher-order visual areas and several polymodal associational areas. Thus EC appears to be the locus of converging polymodal and high-level activity and it is plausible that this activity corresponds to a transient high-level representation of an agent's construal of its experience in terms of events and situations.

[1]The name is derived from the Sanskrit word for "memory". It is also an acronym for a "System for the Memorization of Relational Instances from Transient Impulses".

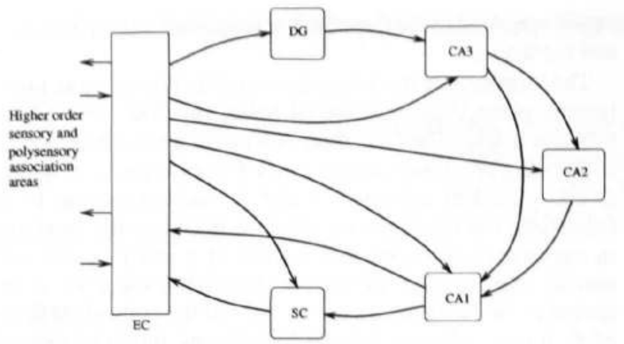

Figure 1: Architecture of the hippocampal formation (after Amaral, 1993). See text for abbreviations.

The major pathways connecting the components of the HS form an idiosyncratic network. These connections give rise to a large number of distinct pathways that start and terminate in EC. Thus the major pathways of the HS form a *complex loop* around the HS. [2]

The HS also interacts with several other brain regions. For example, the HS receives afferents from the amygdala which is implicated in the autonomic and emotive aspects of behavior and cognition and the septal nuclei which in turn receive inputs from the reticular formation, a brain-stem network mediating arousal. These inputs are believed to play an important regulatory role and may provide a global control signal that enables or disables learning (Hasselmo, 1997).

LTP and the Emergence of Committed Cells

Long-term potentiation (LTP) involves long-term increase in synaptic strength resulting from the pairing of presynaptic activity with postsynaptic depolarization, and has emerged as the most promising cellular mechanism underlying activity dependent memory formation (Lynch & Granger, 1992).

LTP involves the unusual receptor NMDA which is activated by the neurotransmitter glutamate, but only if the postsynaptic membrane is already depolarized. Once the NMDA receptor is activated, calcium ions flood into the postsynaptic cell and lead to a complex series of biochemical changes that result in the induction of LTP. The two conditions required for the activation of the NMDA receptor can be brought about by converging inputs arriving at a cell in close temporal proximity; one input can lead to postsynaptic depolarization and the other can cause the release of glutamate. Consequently, NMDA mediated LTP can form the basis of associative learning in neural circuits. LTP possesses several properties that make it suitable for rapid memory formation. It is induced very rapidly — within a few seconds, it is synapse specific, and once stable, it can persist for a long time.

In addition to potentiation, synapses can also undergo activity dependent long-term depression (LTD). The following describes how different forms of LTP and LTD have been modeled in SMRITI using four parameters, namely, Δw, ω, k, and *isp*.

[2] CA2 is often merged with CA3 when describing the rat hippocampal circuitry. In humans and other primates, however, CA2 forms a distinct region.

Associative LTP Coincident pre-synaptic activity at a pair of synapses x and y that share the same post-synaptic cell can lead to their LTP. Synaptic efficacy is modeled as a weight and Δw_{ltp} specifies the increase in weight upon potentiation. The parameter ω specifies the maximum duration by which impulses arriving at x and y can lead/lag one another and still be considered synchronous. Repeated synchronous activity at x and y is required for the induction of LTP. The parameter k specifies the number of times x and y must receive synchronous impulses for the induction of LTP. Finally, isp specifies the maximum permissible gap between the arrival of successive impulses at x (or y) in the above repetition.

Homosynaptic LTP Repeated activation arriving at a synapse can also cause its weight to increase by Δw_{ltp}. This is referred to as homosynaptic LTP. As before, k specifies the number of impulses x must receive before homosynaptic LTP is induced, and isp specifies the maximum permissible gap between successive impulses arriving at x in the above repetition.

Heterosynaptic LTD When a synapse undergoes LTP, neighboring synapses on the same post-synaptic cell may undergo heterosynaptic LTD if they do not receive sufficient pre-synaptic activity. Upon undergoing LTD, the weight of a synapse decreases by Δw_{ltd}.

Homosynaptic LTD A synapse receiving low-level presynaptic activity may undergo homosynaptic LTD if the low level of presynaptic activity is accompanied by post-synaptic hyperpolarization. Upon undergoing LTD, the weight of a synapse decreases by Δw_{ltd}.

Emergence of Committed Cells and Circuits

LTP and LTD can lead to an activity dependent transformation of a quasi-random network into a structure consisting of cells and circuits that are committed to specific functionalities. Typically, a cell receives a large number of afferents and hence, can participate in a potentially large number of functional circuits. If however, the weights of selected synapses impinging on the cell increase (say, via LTP) and optionally, the weights of other synapses decrease (say, via LTD), then the cell becomes highly selective and participates in only a small number of functional circuits. When this happens, we say that the cell has become *committed*. The process of commitment can also be viewed as a neurally plausible realization of the notion of long-term recruitment (Feldman, 1982).

Figure 2 describes how the cell C may become committed to the functional circuit $A\&B_1$. We assume that initially C is uncommitted and its synapses have low efficacy. The coincident activity of A and B_1 results in the associative LTP of synapses formed by the afferents from A and B_1 and the heterosynaptic LTD of synapses formed by afferents from other B_is. If we assume that the firing of C requires inputs at two or more potentiated synapses, then C fires if and only if both A *and* B_1 fire concurrently. In other words, C now behaves as the circuit $A\&B_1$. Observe that if A corresponds to a role and B_is correspond to some entities then C can be viewed as a binding detector for the binding $\langle A = B_1\rangle$.

The encoding of relational instances involves the commitment of more complex circuits for detecting and integrating binding errors. Shastri (1997) describes how local feedback

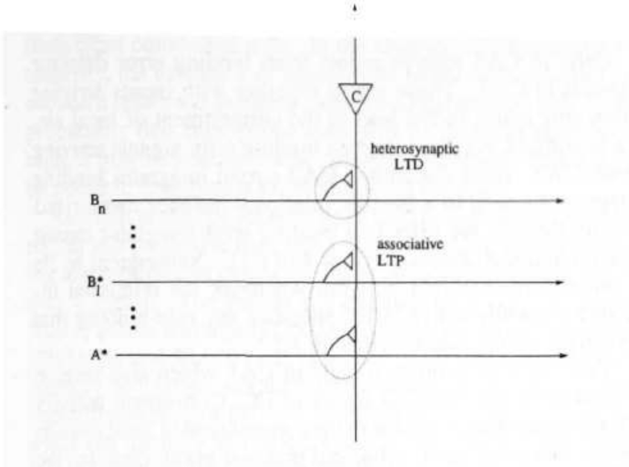

Figure 2: Cell C becomes committed to the circuit $A\&B_1$. A '*' indicates that the source is firing.

and feedforward circuits of the sort known to exist in the HS can get committed to form such functional units.

An Overview of SMRITI

At a macroscopic level, the overall functioning of SMRITI may be described as follows. It is assumed that our cognitive apparatus construes our experiences in terms of events and situations as a result of complex interactions between sensory, perceptual, categorical, linguistic, and inferential processes. These construals are expressed as transient and distributed patterns of activity over high-level cortical circuits (HLCC). The HLCCs in turn project to EC and give rise to transient patterns of activity in EC. The resulting activity in EC can be viewed as the presentation of an event or a situation to the HS by a HLCC for possible memorization. Alternately, a HLCC may present an event or a situation to the HS as a "query" and expect a certain type of response if the query matches one of the items previously memorized by the HS, and a qualitatively different type of response if it is novel.

The activity injected into EC by a HLCC propagates around the complex loop consisting of EC, DG, CA3, CA2, CA1, SC, and EC, and triggers a sequence of synaptic changes involving LTP and LTD. As a result of these changes, the event or situation presented to the HS is transformed from a transient pattern of activity into a persistent structural encoding. The structures committed during this transformation behave as distributed circuits for detecting and integrating bindings and binding errors, and extracting role-fillers from bindings.

The pattern of activation in EC resulting from the activity arriving from CA1 and SC constitutes the response of the HS. The reentrant activity in EC in turn propagates back to the HLCCs. Note that the full blown neural expressions of roles, entities, and generic relations involved in an event or a situation lie outside the HS.

Functional Architecture of SMRITI

Figure 3 shows the functional architecture of SMRITI and identifies how its components might map onto the HS. The memorization of an event or situation involves the rapid formation of:

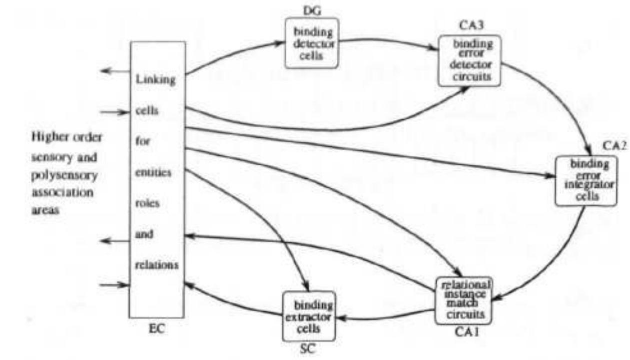

Figure 3: The functional architecture of SMRITI.

- *linking* cells in EC that connect entities, roles, and generic relations in HLCCs to the HS,
- *binding detector* cells in DG,
- *binding error detector* circuits in CA3,
- *binding error integrator* cells in CA2,
- *relational instance match* circuits in CA1, and
- *binding extractor* cells in SC.

As described in Shastri (1997) the above cells and circuits emerge from loosely organized quasi-random network structures as a result of LTP and LTD.

The Transient Encoding of an Event or a Situation

The model posits that the dynamic (active) representation of an event or a situation is a transient pattern of rhythmic activity wherein a role-entity binding is expressed by the synchronous firing of cells associated with the bound role and entity as described in Shastri & Ajjanagadde (1993). It is assumed that each generic relation is encoded as a focal cluster in some HLCC. The cluster for an *n-place* generic relation P contains n role nodes, an *enabler* node (*?P*), and two *collector* nodes ($+P$ and $-P$). The significance of the $?P$, $+P$ and $-P$ nodes is as follows: Assume that the roles of P are dynamically bound to some fillers. The activation of $?P$ means that the HLCC is querying whether or not the currently active instance of P is already encoded in the HS. In contrast, the HLCC activates $+P$ to assert the currently active dynamic instance of P, or it activates $-P$ to assert the negation of the currently active instance.[3] In response to a query about an instance of P, the HS activates the positive (negative) collector if the encoding of the instance (or its negation) exists in the HS.

The activity pattern shown in Figure 4 depicts the transient activity within an HLCC corresponding to an event RI given by: $(R_1 : \langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$ where R_1 is a generic relation, r_1 and r_2 are roles, and f_1 and f_2 are entities bound to r_1 and r_2 respectively. It is assumed that the cells associated with r_1 and f_1 fire in synchrony and so do cells associated with r_2 and f_2. The firing of these two groups of cells however is desynchronized with reference to each other. The transient representation of the query $(R_1 : \langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$? is similar except that the enabler cells $?R_1$, $?f_1$, and $?f_2$ are active instead of the collector cells $+R_1$, $+f_1$, and $+f_2$.

[3]An example of a negation being asserted is "John did not give a book to Mary."

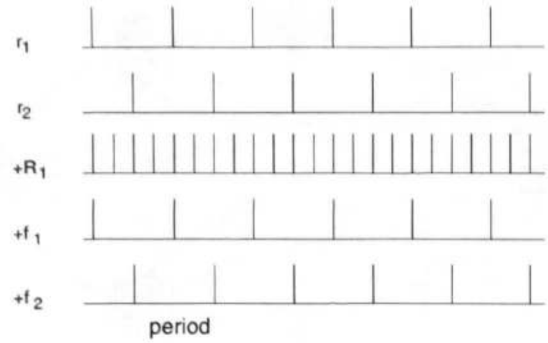

Figure 4: Transient encoding of the relational instance $(R_1 : \langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$

Stepping Through the Model

Interactions between EC and HLCC As a result of the activity arriving in EC from HLCCs, cells in certain regions of EC become committed to the *collector* nodes of entities and generic relations in the HLCCs. Similarly, cells in other regions of EC become committed to the *enabler* nodes of these entities and generic relations. At the same time, cells in a third region of EC become committed to the role nodes. Finally, links between the cells committed to a collector node and cells committed to the corresponding enabler node get potentiated. The commitment of cells in EC to particular generic relations, entities, and roles occurs the very first time the generic relation, entity, or role appears in a relational instance presented to the EC by a HLCC. Subsequently, a committed cell in EC fires whenever the HLCC node it is committed to, fires. Thus the resulting activity of committed enabler, collector, and role cells in EC in response to activity in HLCCs is similar to that shown in Figure 4.

Interactions within the HS Role and ?entity cells project to a large number of cells in DG. As a result of the synchronous activity of those role and ?entity cells in EC that correspond to bound roles and entities in RI, certain uncommitted cells in DG receive convergent activity from r_1 and $?f_1$ cells and become committed to serve as their binder cells. Similarly, for r_2 and f_2. We refer to such committed DG cells as $binder(\langle r_1 = f_1 \rangle)$ and $binder(\langle r_2 = f_2 \rangle)$ cells, respectively. Subsequent to their commitment, $binder(\langle r_1 = f_1 \rangle)$ cells will fire whenever r_1 and $?f_1$ in EC fire synchronously. The cell $binder(\langle r_2 = f_2 \rangle)$ will behave in an analogous manner.

Cells in DG project to cells in CA3 which also receive afferents from role cells in EC. Impulses along afferents from DG and EC lead to the commitment of nodes within CA3 to form circuits for detecting *binding errors*. Thus convergent impulses arriving from $binder(\langle r_1 = f_1 \rangle)$ cells in DG and r_1 cells in EC lead to the commitment of cells in CA3 to form a local feedback circuit for detecting an error in the binding $\langle r_1 = f_1 \rangle$. We refer to such a circuit as $bed(\langle r_1 = f_1 \rangle)$. Subsequent to its commitment, this circuit will fire whenever r_1 is bound to any entity other than f_1 in the relational instance currently active in EC. Similarly, convergent impulses arriving from $binder(\langle r2 = f2 \rangle)$ cells in DG and r_2 cells in EC lead to the commitment of a binding error detection circuit $bed(\langle r_2 = f_2 \rangle)$.

Cells in CA2 receive inputs from binding error detector circuits in CA3. These inputs together with inputs arriving from role nodes in EC lead to the commitment of local circuits within CA2 that integrate binding error signals arriving from CA3. Each committed CA2 circuit integrates binding errors pertaining to a specific relational instance memorized within the HS. We refer to a binding error integrator circuit for a relational instance RI as $bei(RI)$. Subsequent to its commitment, $bei(RI)$ will fire whenever the relational instance currently active in EC specifies any role-binding that is different from that specified in RI.

Cells in CA2 project to cells in CA1 which also receive inputs from the ?relation nodes in EC. Convergent activity along these pathways leads to the commitment of local circuits within CA1 that act as relational instance match circuits. We refer to such a match circuit for a relational instance RI as $rim(RI)$. Subsequent to its commitment, $rim(RI)$ will fire whenever the relational instance currently active in EC matches RI.

Cells in CA1 project to +relation cells in EC. This projection allows the afferents from rim circuits in CA1 to form potentiated links with the appropriate +relation cells in EC. Thus $rim(RI)$ cells in CA1 form potentiated links with $+R_1$ cells in EC. Subsequent to the potentiation of these CA1 to EC links, the firing of $rim(RI)$ cells in CA1 will lead to the firing of $+R_1$ cells in EC.

CA1 cells project to SC which also receives direct projections from role cells in EC. The incident activity along these pathways leads to the commitment of cells in SC that act as *binding extractor* (or bex) cells. We refer to binding extractor cells for the roles r_1 and r_2 of RI as $bex(\langle r_1 = ? \rangle | RI)$ $bex(\langle r_2 = ? \rangle | RI)$ respectively. Subsequent to their commitment, $bex(\langle r_1 = ? \rangle | RI)$ cells will fire whenever $rim(RI)$ fires in temporal proximity of the firing of r_1 and $bex(\langle r_2 = ? \rangle | RI)$ cells will fire whenever $rim(RI)$ fires in temporal proximity of the firing of r_2.

Cells in SC project back to +entity cells in EC. This projection allows afferents from bex cells to form potentiated links with the appropriate +entity cells in EC. Thus $bex(\langle r_1 = ? \rangle | RI)$ cells and $bex(\langle r_2 = ? \rangle | RI)$ cells in SC form potentiated links with $+f_1$ and $+f_2$ cells in EC respectively. Subsequent to their potentiation, the firing of $bex(\langle r_1 = ? \rangle | RI)$ and $bex(\langle r_2 = ? \rangle | RI)$ will lead to the firing of EC cells $+f_1$ and $+f_2$ respectively.

Encoding and Recognition Times

As shown in Shastri (1997), the cells and circuits mentioned above start off as indistinguishable cells and links embedded within loosely organized quasi-random networks but emerge rapidly as a result of LTP and LTD. SMRITI memorizes a relational instance within 20 periods (see Figure 4). Since synchronous activity encoding dynamic bindings is expected to lie in the γ-band, a plausible range of period values is 25—35 msec. Thus SMRITI demonstrates that an event can be memorized in less than a second. SMRITI takes between 5 and 8 periods to recognize and recall memorized instances.

Capacity Considerations

The memorization of relational instances occurs as a result of interactions between quasi-random networks and depends on the existence of target cells that receive suitable afferents from other committed cells. In the absence of complete connectivity, the existence of appropriate target cells required to encode a relational instance cannot be guaranteed. But if the probability that appropriate cells will be found is extremely high, it may be assumed with "practical certainty" that it will be possible to encode a given relational instance.

Relevant probabilities have been calculated using plausible region and projective field sizes, and by making the simplifying assumption that projective fields are distributed uniformly over a region. The results suggest that a capacity of about 50,000 events containing 200,000 distinct bindings involving 2000 roles and 50,000 entities is tenable. Even at this level of memory utilization, the odds of not finding suitable cells for commitment remain below 1 in 300,000. The odds of failure when the memory is loaded with 25,000 events containing 100,000 distinct bindings are less than 2 in a billion. Detailed quantitative results appear in Shastri (1997).

Since multiple cells redundantly encode each functional unit, and given that these cells are quasi-randomly distributed in a region, the probability that limited cell loss will destroy *all* the "copies" of a functional unit is extremely small. Thus the encoding is robust against cell loss. For example, by assuming that about 10 cells become committed to be binding error detectors for each memorized relational instance, it can be shown that the odds of more than 5 of these 10 cells being lost due to a 1% loss of cells are less than 1 in a billion.

Some Predictions

A few key predictions about the effect of focal damage to components of the HS are summarized here: Major damage to EC will lead to erroneous "don't know" responses. Behaviorally this amounts to forgetting. In contrast, major insult to CA3 or CA2 will lead to excessive false positive responses. Major damage to CA1, however, will lead to a catastrophic memory failure. Finally, major damage to SC will leave recognition memory intact but disrupt recall memory. Major cell loss in HF, in particular CA1, will prevent the formation of new memories. Major damage to SC will leave the formation of structures required to support recognition memory intact but prevent the formation of structures required to support recall.

Conclusion

The computational model described above demonstrates how the HS may rapidly transform a transient pattern of activity representing an event or a situation into a persistent structural encoding. It is hoped that detailed experimentation with the model will provide some useful insights into human memory.

The work outlined here has significance for other learning tasks besides the memorization of events and situations. In particular, both the proposed circuit for detecting binding errors and the manner in which such circuits can be formed rapidly within quasi-random network structures, have broad relevance for cognitive neuroscience. For example, this kind of circuit can perform the generic function of *coincidence error* detection; such a circuit is formed when two patterns A and B occur concurrently, and subsequently, it fires whenever A occurs without being accompanied by B. Moreover, the firing of this type of circuit can signify a failure of expectation, and hence, such circuits can form the basis of a system for novelty detection.

Acknowledgments

This work was partially funded by ONR grant N00014-93-1-1149. Thanks to the L0 group at ICSI for discussions.

References

Amaral, D.G. (1993). Emerging principles of intrinsic hippocampal organization. *Current Opinion in Neurobiology 3*:225–229.

Cohen, N.J. & Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System.* Cambridge: M.I.T. Press.

Feldman, J. A. (1982). Dynamic connections in neural networks, *Bio-Cybernetics, 46*:27-39.

Gluck, M.A. & Myers, C.E. (1993). Hippocampal Mediation of Stimulus Representation: A Computational Theory. *Hippocampus 3 (4)*: 491–516.

Hasselmo, M.E. (1997). A model of human memory based on the cellular physiology of the hippocampal formation. In *Neural Networks for Neuropsychologists* R. Parks & D. Levine (eds.) MIT Press. (in the press).

Lynch, G. & Granger, R. (1992). Variations in synaptic plasticity and types of memory in corticohippocampal networks. *Journal of Cognitive Neuroscience 4(3)*:189-199.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society, B 262*: 23-81.

O'Keefe, R.C. & Nadel, L. (1978). *The hippocampus as a cognitive map.* Oxford University Press. London.

O'Reilly, R.C. & McClelland, J.L. (1994). *Hippocampal Conjunctive Encoding Storage, and Recall: Avoiding a Tradeoff.* Technical Report PDP.CNS.94.4. June 1994. Carnegie Mellon University, Pittsburgh, PA.

Scoville, W.B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry 20*:11-21.

Schmajuk, N.A. & DiCarlo, J.J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. *Psychological Review, 99* (2), 268-305.

Shastri, L. (1997). From transient patterns to persistent structures: a computational model of rapid memory formation in the hippocampal system. In preparation.

Shastri, L. & Ajjanagadde V. (1993). From simple associations to systematic reasoning. *Behavioral and Brain Sciences 16:3*, 417-494.

Squire, L.R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195–231.

Squire, L.R. & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science 253*:1380-1386.

Treves, A & Rolls, E.T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus 4(3)*:374-391.

Valiant, L. (1994). *Circuits of the mind.* New York : Oxford University Press.

Van Hoesen, G.W. (1982). The primate hippocampus gyrus: New insights regarding its cortical connections. *Trends in Neuroscience, 5*, 345–350.