Appeared in Neurocomputing, 38-40: 889-897

A computational model of episodic memory formation in the hippocampal system

Lokendra Shastri
International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, CA 94704, USA
TEL: (510) 642-4274; FAX: (510) 643-7684
shastri@icsi.berkeley.edu

Key words: Episodic Memory; Hippocampus; Binding; Recruitment

Abstract

The memorization of events and situations (episodic memory) requires the rapid formation of a memory trace consisting of several functional components. A computational model is described that demonstrates how a transient pattern of activity representing an episode can lead to the rapid recruitment of appropriate circuits as a result of long-term potentiation within structures whose architecture and circuitry match those of the hippocampal formation, a neural structure known to play a critical role in the formation of such memories.

1 Introduction

We remember our experiences in terms of events and situations that record who did what to whom where and when, or describe states of affairs wherein multiple entities occur in particular configurations. This form of memory is referred to as episodic memory [23] and it is known that the hippocampal system (HS) serves a critical role in the formation of such memories [10, 21, 4, 22].

A number of researchers have proposed models to explain how the HS subserves the episodic memory function. These include macroscopic system-level models that attempt to describe the functional role of the HS as well as more detailed computational models that attempt to explicate how the HS might realize its putative function (e.g., [11, 14, 22, 8, 13, 12]. While our understanding of the HS and its potential role in memory formation and retrieval has been enhanced by this extensive body of work, several key representational problems associated with the encoding of episodic memories have remained unresolved. In particular, most existing computational models view an item in episodic memory as a feature vector or as a conjunction of features. But for reasons summarized below, such a view of episodic memory is inadequate for encoding events and situations (also see [16, 18]).

First, events are *relational objects*, and hence, cannot be encoded as a conjunction of features. Consider an event $E1$ *described* by "John gave Mary a book in the library on Tuesday". This event cannot be encoded by simply associating "John", "Mary", "a Book", "Library", "Tuesday" and "give" since such an encoding would be indistinguishable from that of the event "Mary gave John a book in the Library on Tuesday". In order to make the necessary distinctions, the memory trace of an event should specify the *bindings* between the *entities* participating in the event and the *roles* they play in the event. For example, the encoding of $E1$ should specify the following *role*-`entity` bindings: (⟨*giver*=`John`⟩, ⟨*recipient*=`Mary`⟩, ⟨*give-object*=`a-Book`⟩, ⟨*temporal-location*=`Tuesday`⟩, ⟨*location*=`Library`⟩).

Second, the memory trace of an event should be responsive to partial cues, but at the same time, it should be capable of distinguishing between the memorized event and other highly similar events. For example, the memory trace of $E1$ should respond positively to a partial cue "John gave Mary a book," but not to a cue "John gave Susan a book in the library on Tuesday" even though the latter event shares all, but one, bindings with $E1$.

Third, during retrieval, the memory trace of an event should be capable of reactivating the bindings associated with the event so as to recreate an active representation of the event. In particular, the memory trace of an event should support the retrieval of specific components of the event. For example, the memory trace of $E1$ should be capable of differentially activating John in response to the cue "Who gave Mary a book?"

It can be shown that in order to satisfy *all* of the above representational requirements, the memory trace of an event must incorporate functional units that serve as binding-detectors, binding-error-detectors, binding-error-integrators, relational-instance-match-indicators, and binding-reinstators [18]. Moreover, as explained in [18], it is possible to evoke an articulated representation of an event by reinstating the bindings describing the event and activating the web of semantic and procedural knowledge with these bindings. Cortical circuits encoding sensorimotor schemas and generic "knowledge" about actions such as *give* and entities such as *persons*, *books*, *libraries*, and *Tuesday* can recreate the necessary gestalt and details about the event "John gave Mary a book on Tuesday in the library" upon being activated with the small number of bindings associated with this event.

This article partially describes a computational model, SMRITI, that demonstrates how a transient pattern of activity representing an event can lead to the rapid formation of appropriate functional units as a result of long-term potentiation (LTP) [2] within structures whose architecture and circuitry match those of the HS. The model shows that the seemingly idiosyncratic architecture of the HS and the existence of different types of inhibitory interneurons, and different types of local inhibitory circuits in the HS, are ideally suited for the recruitment of the necessary functional units. A detailed description of the model appears in [18].

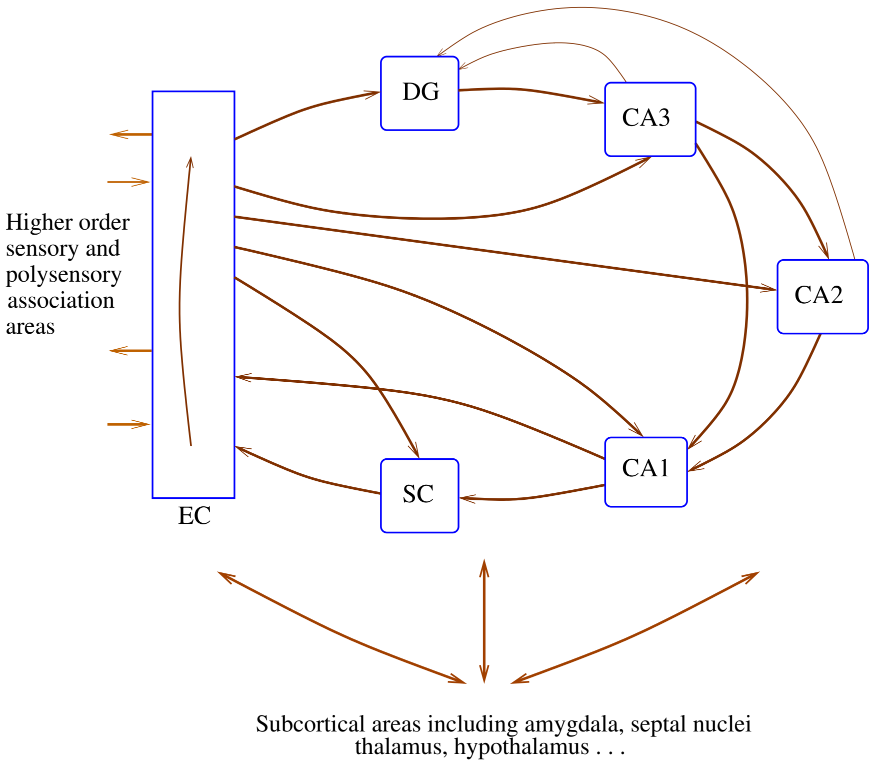

Figure 1: Summary of major pathways interconnecting the components of the hippocampal system. Note the multiple pathways from EC to CA1, CA2, CA3, and SC. Also note the backprojections from CA1 and SC to EC. The HS also receives a wide array of subcortical inputs (not shown). The subicular complex (SC) has been depicted as a single monolithic component for simplicity. See text for detail. Abbreviations: HS, hippocampal system; DG, dentate gyrus; EC, entorhinal cortex; SC, subicular complex.

1.1 The hippocampal system

The hippocampal system (HS) refers to a heterogeneous collection of neural structures including the entorhinal cortex (EC), the dentate gyrus (DG), Ammon's horn (or the hippocampus proper), and the subicular complex (SC). Ammon's horn consists of fields CA1, CA2, and CA3, and the SC consists of the subiculum, presubiculum, and parasubiculum regions. Note that CA2 is often merged with CA3 in the animal literature, but is a distinct field of Ammon's horn, especially in humans and other primates [5].

Fig. 1 depicts a schematic of the major pathways interconnecting the components of the HS. EC serves as the principal portal between the HS and the rest of the cortex. Higher-order unimodal sensory areas as well as polymodal association areas project to upper layers of EC (e.g., [24]). In turn, the upper layers of EC project to DG, CA3, CA2, CA1, and SC. Moreover, DG projects to CA3, CA3 projects to CA2 and CA1, CA2 projects to CA1, and CA1 projects to SC. CA1 and SC project back to the deeper layers of EC which in turn project back to the high-level cortical areas that project to EC. Thus activity originating in high-level cortical regions converges on the HS via EC, courses through the complex loop formed by the pathways of the HS, and returns back to high-level cortical regions from where it had originated.

In addition to the forward pathways mentioned above, there also exist pathways from CA2 and CA3 to DG, and recurrent connections within DG, CA3, CA2, and to a lesser extent, within CA1. Moreover, each region of the HS contains a variety of inhibitory interneurons that in conjunction with principal cells give rise to well-defined feedback and feedforward inhibitory local circuits. The HS also receives a rich set of afferents from subcortical areas that mediate arousal and other autonomic, emotional, and motivational aspects of behavior. These inputs can communicate the affective significance of an experience/stimulus to the HS.

1.2 Long-Term Potentiation

LTP refers to long-term activity dependent increase in synaptic strength and is believed to underlie memory formation [2]. In particular, convergent activity at multiple synapses that share the same postsynaptic cell can lead to their associative LTP. The proposed computational model uses a highly idealized, but computationally effective, form of associative LTP. In brief, the occurrence of LTP in the model is governed by the following parameters: the *potentiation threshold* θ_p, the *weight increment* Δw_{ltp}, the *window of synchrony* ω, the *repetition factor* κ, and the *maximum inter-activity interval* τ_{iai}. Consider a set of synapses $s_1, \ldots, s_n$ sharing the same postsynaptic cell. Convergent presynaptic activity at s_i's can lead to associative LTP of naive s_i's and increase their weights by Δw_{ltp} if the following conditions hold: (i) the total (convergent) activity arriving at s_i's exceeds θ_p, (ii) this activity is synchronous, i.e., arrives with a maximum lead/lag of ω, (iii) such synchronous activity repeats at least κ times, and (iv) the interval between two *successive* arrivals of convergent activity is at most τ_{iai}. It has been shown [17] that recruitment learning algorithms [6, 15] proposed for one-shot learning in connectionist networks can be firmly grounded in LTP.

2 A System-level description of the model

At a macroscopic level, the functioning of the model may be described as follows. Our cognitive apparatus construes our experiences as a stream of events and situations and these are expressed as transient and distributed patterns of activity over high-level cortical circuits (HLCCs). HLCCs in turn project to EC and give rise to transient patterns of activity. The resulting activity in EC can be viewed as the presentation of an event to the HS by HLCCs for possible memorization. Alternately, HLCCs may present an event to the HS as a "query" and expect a certain type of response from the HS if the query matches a previously memorized item, and a qualitatively different type of response if it does not. At a macroscopic level of description, the cortico-hippocampal interaction envisioned here is similar to that assumed by other models of the HS-based memory system (e.g., [11, 7, 4]).

The transient activity injected into EC propagates around the complex loop consisting of EC, DG, CA3, CA2, and CA1, SC, and back to EC, and triggers a sequence of synaptic changes in these structures. The model demonstrates that such synaptic changes can transform the transient pattern of activity into a persistent structural encoding (i.e., a memory trace) composed of the requisite functional circuits mentioned in Section 1. The level of significance of a presented event determines the *mass*, i.e., the number of cells, recruited for encoding an event. The activity 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 HLCCs and completes a cycle of cortico-hippocampal interaction.

Note that all entities and generic relations and their roles are expressed in various HLCCs. Only the functional units listed in Section 1 that are required for the encoding of specific events involving these entities, generic relations, and roles are encoded within the HS.

2.1 Functional architecture of SMRITI

The mapping between the functional units comprising each memory trace and the component regions of the HS is as follows: (refer to Fig. 1).

- *Linking* cells in EC. These cells connect the HLCC based representations of entities, generic relational structures and their roles to the HS.
- *Binding-detector* cells in DG,
- *Binding-error-detector* circuits in CA3,
- *Binding-error-integrator* cells in CA2,
- *Relational-match-indicator* circuits in CA1 for detecting a match between a cue and the memorized event based on the activity of the above cells and circuits, and
- *Binding-reinstator* cells in SC.

Several "copies" of each functional cell and circuit are recruited during the memorization of an event leading to a highly redundant encoding of the event. The recruitment of *binding-detector* cells and *binding-error-detector* circuits is described in detail in [16]. The recruitment of other functional units is discussed in [18].

The transient encoding of an event or a situation

The model posits that the dynamic encoding of an event situation is a transient pattern of rhythmic activity. A role-entity binding is expressed within this activity by the synchronous firing of cells associated with the bound role and entity [19, 25, 20]). It is speculated that the dynamic expression of role-entity bindings may involve γ band activity (minor cycle) and blocks of repeated activity (major cycle) may correspond to θ band activity (cf. [3]).

3 Results

Simulations of the model containing 1000-1500 cells and 8,000-15,000 links have been performed to verify that the functional units are recruited as prescribed by the model. However, these simulations do not reveal properties of the model arising from its large scale. To explicate these properties, a detailed quantitative analysis of the model has been carried out using plausible values of system parameters. The system specification involves over 75 parameters which specify among other things, region and projective field sizes, the ratio of pyramidal cells to various types of inhibitory interneurons, firing thresholds, time course of *epsps*, refractory periods, and the conditions governing the induction of LTP. In particular, the calculations assume that DG, CA3, CA2, CA1, and SC contain 15 million, 2.7 million, 800,000, 15 million, and 1 million principal cells, respectively, the ratio of principal cells to inhibitory cells is about 10:1, and the size of EC $\rightarrow$ DG and DG $\rightarrow$ CA3 projective fields are 17,000 and 14, respectively. These choices are based in part on data provided in [1][26].

The results of the quantitative analysis show that extremely robust memory traces can be formed even if one assumes an episodic memory capacity of 75,000 distinct events consisting of 300,000 bindings. Some of the results of this quantitative analysis are displayed in Table 1. The memory traces formed in the model exhibit a strong form of pattern separation and differentiate between two events even though they match along all but one dimension. This can be seen from the data presented in Fig. 2 which shows the response of *relation-match-indicator* (*remind*) cells in CA1 to cues containing different numbers of binding-errors. The firing of *remind* cells in response to a cue signals a match between the cue and the encoded event. Ideally these cells should fire if and only if a cue has zero binding-errors with respect to the event encoded by the memory trace. In keeping with the desired behavior, the response of *remind* cells is sharply lower in all the binding-error conditions compared to that in the match condition. Since *remind* cells lie at the apex of the memory trace of an event, their response is affected by cross-talk and errors in all preceding functional units. Consequently, the robust response of *remind* cells is significant.

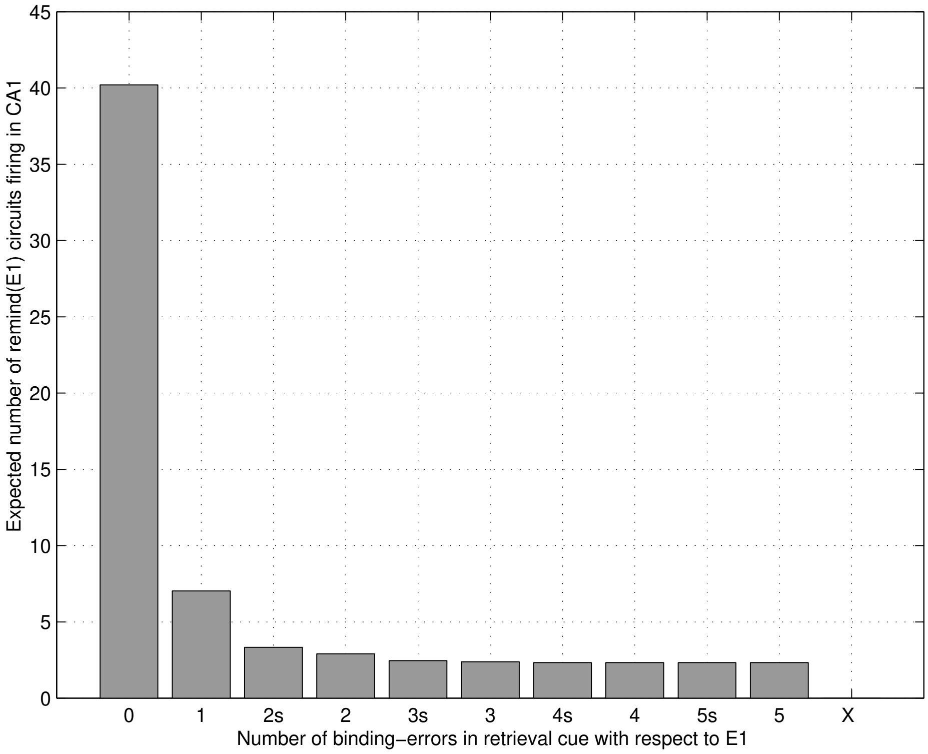

Figure 2: The response of *remind(E1)* circuits in CA1 as a function of the match between the retrieval cue and the memorized event $E1$. Condition X refers to a retrieval cue involving a different type of event (e.g., if $E1$ is an event involving the act of "buying", then a cue involving acts other than "buying" (say walking) would correspond to a cue involving a different type of event).

Statistic	DG *binder*	CA3 *bed*	CA2 *bei*	CA1 *remind*	SC *reinstate*
Pfail	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$
E⟨candidiates⟩	195.02	412.70	88.49	57.99	845.87
E⟨ recruits⟩	195.02	17.64	13.48	52.45	137.29

Table 1: *Pfail* denotes the probability that suitable cells or circuits will not be found in a target region for recruitment as a functional unit during the memorization of the event "John gave Mary a book in the library on Tuesday." E⟨candidate⟩ denotes the expected number of cells or circuits that receive adequate connections, and hence, are candidates for recruitment as a functional unit. E⟨recruits⟩ specifies the expected number of candidate cells or circuits that will be recruited for each functional unit. This number indicates the expected number of "copies" of each functional unit in the memory trace of E1 formed in the HS. The ratio of the number of candidate cells to the number of recruited cells is governed by inhibition from Type-1 interneurons. Abbreviations: DG: dentate gyrus; SC: Subicular complex; *binder*: binding-detector cells; *bed*: binding-error-detector circuits; *bei*: binding-error-integrator cells; *remind*: relation-match-indicator circuits; *reinstate*: binding-reinstator cells.

4 Predictions

The model predicts that significant damage to EC will lead to a catastrophic failure in the formation and retrieval of episodic memories. Damage to DG granule cells will lead to erroneous "don't know" responses (forgetting). Large scale loss of CA3 pyramids or loss of perforant path inputs to CA3 will lead to false positive responses (spurious memories). Large scale loss of mossy fiber inputs or loss of CA3 interneurons will lead to erroneous "don't know" responses (forgetting). Loss of CA1 pyramidal cells will lead to catastrophic forgetting, but loss of CA1 inhibitory interneurons will lead to false positive responses. Subjects with damage to SC will have good recognition memory, but impaired recall, and will have difficulty responding to wh-questions. Finally, subjects with damage to CA1, but with intact EC, DG and CA3 regions will continue to generate binding-error signals, and hence, continue to detect novelty [9] even though they may be amnesic.

5 Conclusion

The proposed computational model explains how a transient pattern of activity representing an event can lead to the formation of a *functionally adequate* memory trace in the HS. The memory traces capture the relational aspects of events and situations. They can differentiate between highly similar events and respond to partial cues. The model accounts for the idiosyncratic architecture and local circuitry of the HS and provides a rationale for its components and projections. It suggests specific and significantly different functional roles for regions CA3, CA2, CA1, and SC than those suggested by existing models. While most models of the HS view CA3 as an associative memory, this work suggests that (i) a key representational role of CA3 may be the detection of binding errors and (ii) CA3 interneurons may play a critical role in the formation of binding-error detector circuits critical to the proper functioning of episodic memory and novelty detection.

References

[1] D.G. Amaral and M.P. Witter, Hippocampal Formation, in: G. Paxinos, ed., *The Rat Nervous System*, second, edition, (Academic Press, London, 1995) 443-493.

[2] T.V.P. Bliss and G.L. Collingridge, A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361 (1993) 31-39.

[3] J.J. Chrobak and G. Buzsaki, Gamma Oscillations in the Entorhinal Cortex of the Freely Behaving Rat. *J. Neurosci.*, 18 (1998) 388-398.

[4] N.J. Cohen and H. Eichenbaum, *Memory, Amnesia, and the Hippocampal System* (M.I.T. Press, Cambridge, Massachusetts, 1993).

[5] H.M. Duvernoy, *The Human Hippocampus*. (J.F. Bergmann, Munich, 1988).

[6] J.A. Feldman, Dynamic connections in neural networks. *Bio-Cybernet.* 46 (1982) 27-39.

[7] E. Halgren, Human hippocampal and amygdala recording and stimulation: evidence for a neural model of recent memory, in: L.R. Squire and N. Butters, eds., *Neuropsychology of Memory* (Guilford Press, New York, 1984) 165-182.

[8] M.E. Hasselmo, B.P. Wyble and G.V. Wallenstein, Encoding and retrieval of episodic memories: Role of cholinergic and GABAergic modulation in the hippocampus, *Hippocampus* 6 (1996) 693-708.

[9] R.T. Knight, Contribution of human hippocampal region to novelty detection. *Nature* 382 (1996) 256-259.

[10] J. O'Keefe and L. Nadel, *The hippocampus as a cognitive map* (Clarendon Press, Oxford, 1978).

[11] D. Marr, Simple memory: a theory for archicortex, *Phil. Trans. R. Soc. B* 262 (1971) 23-81.

[12] J.M.J. Murre, TraceLink: a model of amnesia and consolidation of memory. *Hippocampus* 6 (1996) 674–684.

[13] R.C. O'Reilly and J.L. McClelland, Hippocampal Conjunctive Encoding Storage, and Recall: Avoiding a Tradeoff, Technical Report PDP.CNS.94.4, Carnegie Mellon University, Pittsburgh, PA, 1994.

[14] N.A. Schmajuk and J.J. DiCarlo, Stimulus configuration, classical conditioning, and hippocampal function. *Psych Rev.* 99 (1992) 268-305.

[15] L. Shastri, *Semantic Networks: An evidential formalization and its connectionist realization*, (Morgan Kaufamnn, Los Altos, CA, 1988).

[16] L. Shastri, Recruitment of binding and binding-error detector circuits via long-term potentiation, *Neurocomputing* 26-27 (1999) 865-874.

[17] L. Shastri, Biological grounding of recruitment learning and vicinal algorithms in long-term potentiation and depression. Technical Report TR-99-009, International Computer Science Institute, Berkeley, CA, 1999.

[18] L. Shastri, From transient patterns to persistent structures: a model of episodic memory formation via cortico-hippocampal interaction. Submitted.

[19] L. Shastri and V. Ajjanagadde. From simple associations to systematic reasoning. connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behav. Brain Sci.* 16 (1993) 417-494.

[20] W. Singer and C.M. Gray, Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.* 18 (1995) 555-586.

[21] L.R. Squire, Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psych. Rev.* 99 (1992) 195-231.

[22] A. Treves and E.T. Rolls, Computational analysis of the role of the hippocampus in memory. *Hippocampus* 4 (1994) 374-391.

[23] E. Tulving, *Elements of Episodic Memory* (Clarendon Press, Oxford, 1978).

[24] G.W. Van Hoesen, The primate hippocampus gyrus: New insights regarding its cortical connections. *Trends Neurosci.* 5 (1982) 345-350.

[25] C. von der Malsburg. Am I thinking assemblies? in: G. Palm and A. Aertsen, eds., *Brain Theory* (Springer-Verlag, Berlin, 1986).

[26] M.J. West, Stereological studies of the hippocampus: a comparison of the hippocampal subdivisions of diverse species including hedgehogs, laboratory rodents, wild mice and men, in:

J. Storm-Mathisen, J. Zimmer, and O.P. Ottersen, eds., *Progress in Brain Research: Understanding the brain through the hippocampus* (Elsevier Science, Amsterdam, 1990) 13-36.

Acknowledgments: This work was supported by NSF grants 9720398 and 9970890.