

Appeared in Neurocomputing, 26-27 (1999) 865-874

Recruitment of binding and binding-error detector circuits via long-term potentiation

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: Bindings; Episodic Memory; Hippocampus; Synchrony; Recruitment.

Abstract

The memorization of events and situations (episodic memory) requires the rapid formation of neural circuits for detecting bindings and binding-errors. The formation of binding-error detectors, however, is problematic given their paradoxical behavior. A computational model is described that demonstrates how a transient pattern of activity representing an episode can lead to the rapid formation of circuits for detecting bindings and bindings-errors as a result of long-term potentiation within structures whose architecture and circuitry match those of the hippocampal formation, a neural structure known to be critical to episodic memory formation.

1 Introduction

Our ability to remember events in our daily life demonstrates our capacity to rapidly acquire new memories. Typically, such memories 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 often referred to as episodic memory [16], and there is a broad consensus that the hippocampal formation (HF) serves a critical role in its formation [7][14] [3][15].

The persistent encoding of an event or a situation (henceforth, an event) must satisfy several representational requirements. Consider the event *described* by “John gave Mary a book in the library on Tuesday”. This event cannot be encoded by simply forming associations between “John”, “Mary”, “a Book”, “Library”, “Tuesday” and “give” since such an encoding would be indistinguishable from that of the event described by “Mary gave John a book in the Library on Tuesday”. In order to make the necessary distinctions, the encoding 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 the event in question should specify the following *role-entity* bindings: ($\langle \text{giver}=\text{John} \rangle$, $\langle \text{recipient}=\text{Mary} \rangle$, $\langle \text{give-object}=\text{a-Book} \rangle$, $\langle \text{temporal-location}=\text{Tuesday} \rangle$, $\langle \text{location}=\text{Library} \rangle$).

As explained in [11], it is possible to evoke a fleshed out representation of an event by “retrieving” the bindings describing the event and activating the web of semantic and procedural knowledge with these bindings. It is argued in [11] that cortical circuits encoding 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 above bindings.

The memory trace of a relational instance must respond positively to partial cues, but at the same time reject a cue that specifies incompatible bindings — even if the cue is otherwise highly similar to the encoded instance. The need

to recognize partial cues and at same time reject similar but erroneous cues entails that the memory trace of an event must incorporate circuits for detecting binding-errors. For example, a memory trace of $(\langle r1=a \rangle, \langle r2=b \rangle, \langle r3=c \rangle)$ that detects binding matches, but not binding errors, will treat an erroneous cue such as $(\langle r1=a \rangle, \langle r2=b \rangle, \langle r3=d \rangle)$ on par with a partial but matching cue such as $(\langle r1=a \rangle, \langle r2=b \rangle)$, since both contain the same number of matching bindings (two). In view of the above, the memory trace of an event must also encode circuits for detecting binding errors.

The rapid formation of binding-error detector circuits is problematic given their paradoxical functional behavior. The crux of the problem is this: The formation of an error detector for the binding of a role r and an entity f must occur in response to the concurrent activity of r and f . But subsequent to its formation, the binding-error detector must *not* fire anymore in response to the concurrent activity of r and f — *the very activity that led to its formation*. Instead, it must fire in response to the firing of r without the coincident firing of f . A satisfactory explanation of how binding-error detector circuits may be formed rapidly has not been proposed thus far.

This article partially describes a computational model that demonstrates how a transient pattern of activity representing an event can lead to the rapid formation of circuits for detecting bindings as well as binding errors as a result of long-term potentiation (LTP) [2] within structures whose architecture and circuitry match those of the HF. The model also offers an alternative interpretation of the functional role of CA3 and predicts the nature of memory impairment that would result from focal damage to specific regions of the HF. A detailed description of this model appears in [11].

A binding-error detector circuit can perform the generic function of *coincidence error* detection, since such a circuit is formed when two patterns A and B occur concurrently, and once formed, it fires whenever A occurs without being accompanied by B . Moreover, the firing of such circuits can also signify a failure of expectation, and hence, such circuits can form the basis of a system for novelty detection, a function attributed to the HF [6].

1.1 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, \dots, 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} .

2 A structure for the formation of binding and binding-error detector circuits

The composite structure for the rapid formation of binding and binding-error detectors consists of four regions: ROLE, ENTITY, BIND and BED (see Figure 1(a)). Region ENTITY projects to region BIND, region ROLE projects to both regions BIND and BED, and region BIND projects to region BED. All these projections are diffuse and dense with the exception of the projection from BIND to BED which is diffuse but sparse. Each role and entity is encoded by a small ensemble of cells in the ROLE and ENTITY regions, respectively. Cells within an ensemble are dispersed within a region.

2.1 The transient representation of role-entity bindings

The model assumes that an experience construed as a relational instance is expressed as a transient pattern of rhythmic activity over distributed high-level cortical circuits (HLCCs). These HLCCs project to cells in ENTITY and ROLE

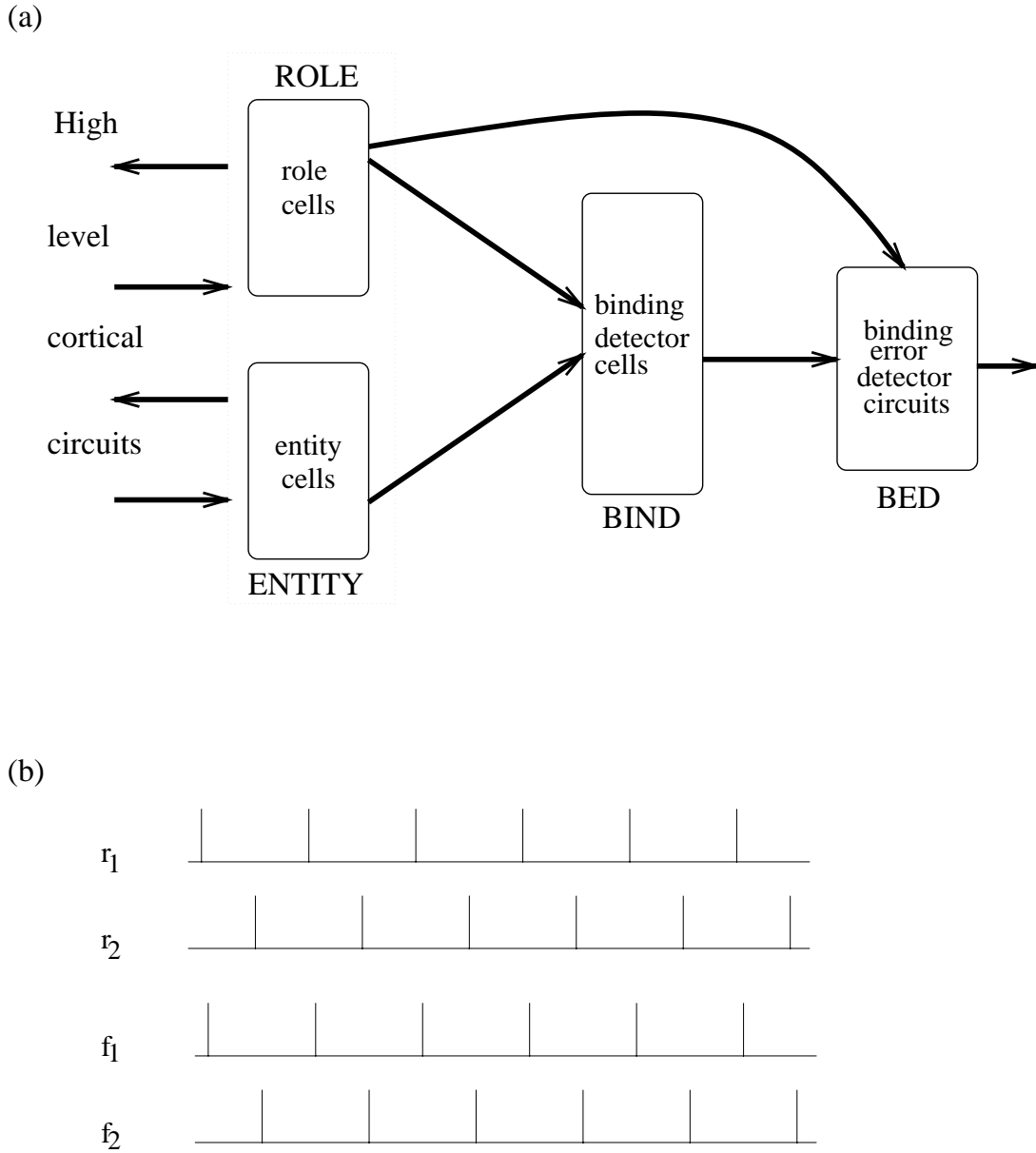


Figure 1: (a) The composite structure for the formation of binding and binding-error detectors. Arrows indicate projections. Each role and entity is encoded by a small ensemble of cells in the ROLE and ENTITY regions, respectively. Binding and binding-error detectors are formed in regions BIND and BED, respectively. (b) The transient encoding of a relational instance RI given by: $((r_1 = f_1), (r_2 = f_2))$. Here r_1 and r_2 are roles, and f_1 and f_2 are entities bound to r_1 and r_2 , respectively. Each spike in the illustration signifies the synchronous firing of a cell ensemble. Cells in the r_1 and f_1 ensembles fire in synchrony and so do cells in the r_2 and f_2 ensembles.

regions and, in turn, induce transient patterns of rhythmic activity within these regions. Figure 1(b) is an idealized depiction of the transient activity induced in ENTITY and ROLE regions by HLCCs to convey the relational instance $RI: (\langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$. In the above, r_1 and r_2 are roles, and f_1 and f_2 are entities bound to r_1 and r_2 , respectively. Each spike in the illustration signifies the quasi-synchronous firing of a cell ensemble. Cells in the r_1 and f_1 ensembles fire in synchrony and so do cells in the r_2 and f_2 ensembles. The firing of cells in the r_1 and f_1 ensembles, however, is desynchronized with the firing of cells in the r_2 and f_2 ensembles. Thus each role-entity binding is expressed as the synchronous firing of cell ensembles associated with the bound role and entity [19][12][13].

2.2 Internal structure of BIND and the formation of binding detector cells

BIND contains principal cells and Type-1 inhibitory interneurons. Each principal cell receives afferents from cells in ROLE and ENTITY regions and makes synaptic contacts on interneurons. The interneurons in turn make contacts on a number of principal cells, thereby forming feedback and feedforward inhibitory circuits within BIND.

The transient encoding of the relational instance RI (Figure 1(b)) leads to the following events in BIND. The synchronous firing of cells in the r_1 and f_1 ensembles (henceforth, r_1 and f_1 cells) leads to the associative LTP of active synapses at those principal cells that receive afferents from both r_1 and f_1 cells. This *recruits*¹ these principal cells to be the binding detector cells for the binding $\langle r_1 = f_1 \rangle$ and they will be referred to as $\langle r_1 = f_1 \rangle$ cells. Similar LTP events occur at the synapses of principal cells that receive coincident activity along afferents from r_2 and f_2 cells and lead to the recruitment of $\langle r_2 = f_2 \rangle$ cells.

The efficacy of naive synapses formed by afferents from ROLE and ENTITY is low and impulses arriving at these synapses only lead to sporadic principal cell activity. But after a principal cell is recruited, the coincident arrival of impulses at potentiated synapses from the associated role and entity cells produces robust firing of the cell in close temporal proximity of the presynaptic activity. Thus a $\langle r_1 = f_1 \rangle$ cell fires whenever r_1 and f_1 cells fire in synchrony and behaves as a binding detector cell for the role-entity binding $\langle r_1 = f_1 \rangle$. Similarly, a $\langle r_2 = f_2 \rangle$ cell fires whenever r_2 and f_2 cells fire in synchrony and behaves as a binding detector cell for the role-entity binding $\langle r_2 = f_2 \rangle$.

The coding of the binding $\langle r_1 = f_1 \rangle$ requires the existence of cells that receive afferents from both r_1 and f_1 cells. Given the quasi-random nature of connectivity this cannot be guaranteed, but the probability of not finding suitable connected cells can be extremely small if the size of the projective fields (PFs) of role and entity cells are large. This however, has the undesirable effect of making the expected number of BIND cells recruited to each role-entity binding very large, thereby increasing the expected overlap between binder cells for different bindings. The local inhibitory feedback and feedforward circuits formed by principal cells and inhibitory interneurons alleviate this problem by limiting the number of cells whose synapses undergo LTP (cf. [8]). Spurious (false-positive) binding detector cells for $\langle r_1 = f_1 \rangle$ may be formed due to the recruitment of cells receiving adequate number of afferents from r_1 cells alone or f_1 cells alone. The ratio of spurious to well-formed binder cells, however, is quite small (see Section 4).

2.3 The internal structure of BED and the encoding of binding-error detectors

BED contains principal cells and two types of inhibitory interneurons (Type-1 and Type-2). The principal cells and interneurons form two types of local circuits. The first involve Type-1 interneurons and perform the same function as that performed by local inhibitory circuits in BIND; they limit the number of principal cells whose synapses undergo LTP. The second type of circuits involve Type-2 interneurons and lead to the formation of binding-error detectors.

Each principal cell receives afferents from a number of cells in ROLE and BIND and sends collaterals to neighboring Type-2 interneurons. Type-2 interneurons in turn make contacts on neighboring principal cells. If a principal cell P receives an inhibitory contact from a Type-2 interneuron int , then the likelihood that P also sends a collateral to

¹The term recruitment learning was suggested by Feldman[4] for the rapid learning of conjunctive concepts in random networks. Also see [17].

int is high. Consequently, there exist a large number of feedback circuits consisting of a principal cell and a Type-2 interneuron (Figure 2(a)). In general, each principal cell and Type-2 interneuron may participate in many feedback circuits. The projection from BIND to BED is such that given an interconnected principal cell P and Type-2 interneuron int , if P receives an afferent from a cell b in BIND, then it is likely that int also receives an afferent from b . A principal cell does not fire unless it receives impulses at potentiated synapses from ROLE or BIND cells, and a Type-2 interneuron does not fire unless it receives impulses at potentiated synapses from BIND cells.

Figure 2(a) shows a principal cell P receiving afferents from r_1^* (r_1 cell in ROLE), and $\langle r_1 = f_1 \rangle^*$, a $\langle r_1 = f_1 \rangle$ cell in BIND. Given the dynamic encoding of RI (Figure 1(b)), these afferents convey synchronous activity and this leads to the associative LTP of P 's synapses receiving afferents from $\langle r_1 = f_1 \rangle^*$ and r_1^* , respectively. After the potentiation of its synapses, P fires upon receiving activation from r_1^* and/or $\langle r_1 = f_1 \rangle^*$ (Figure 2(b)).

Subsequent volleys of inputs from r_1^* and/or $\langle r_1 = f_1 \rangle^*$ cause P to fire and result in impulses arriving at the synapse between P and int . Now int is already receiving activation from $\langle r_1 = f_1 \rangle^*$ and the arrival of concurrent activity from P leads to the associative LTP of the synapse at which int receives activation from $\langle r_1 = f_1 \rangle^*$. After the LTP of this synapse, activation of $\langle r_1 = f_1 \rangle^*$ is sufficient to fire int and cause the inhibition of P (Figure 2(b)).

At the end of the above sequence of events, the circuit consisting of P and int becomes a binding-error detector circuit for the binding $\langle r_1 = f_1 \rangle$ and will be referred to as a $bed(\langle r_1 = f_1 \rangle)$ circuit. In future, P fires whenever the firing of r_1^* is *not* accompanied by the synchronous firing of $\langle r_1 = f_1 \rangle^*$. In other words, P fires whenever the activity in the ROLE and ENTITY regions binds the role r_1 to any entity other than f_1 . A similar recruitment process leads to the formation of $bed(\langle r_2 = f_2 \rangle)$ circuits that act as binding-error detectors for the binding $\langle r_2 = f_2 \rangle$. In general, many binding-error detector circuits are formed for each binding. Spurious (false-positive as well as false-negative) bed circuits may be formed due to the recruitment of ill-connected cells or due to the sharing of cells among circuits. The ratio of such spurious circuits to well-formed circuits, however, is relatively small (see Section 4).

3 Match between the hippocampal formation and the model structure

There is a direct correspondence between the model structure described above and the HF. Thus ROLE and ENTITY regions correspond to subregions of the entorhinal cortex (EC), the BIND region corresponds to the dentate gyrus (DG), and the BED region to field CA3 of the hippocampus. The projections from high-level cortical areas to ROLE and ENTITY correspond to the well known cortical projections to EC [18]. The dense and diffuse projections from ROLE and ENTITY to BIND and BED correspond to the dense and diffuse projections along the perforant path from EC to DG and CA3, respectively [1]. Similarly, the sparse but diffuse projection from BIND to BED corresponds to the sparse mossy fiber projection from DG to CA3 [1]. The internal circuitry of BIND and BED regions also matches the local circuitry of DG and CA3. The principal cells in BIND and BED correspond to DG granule cells and CA3 pyramidal cells, respectively, and the interneurons in these regions correspond to inhibitory interneurons in DG and CA3, respectively [5][10] [1][9]. The model posits that LTP occurs at a synapse formed by mossy fibers on a CA3 interneuron if the latter receives coincident activation from a CA3 pyramidal cell.

4 Quantitative considerations

The following quantities have been calculated analytically (see Table 1): (i) for a given binding, the probabilities that *no* cells with suitable connections will be found in BIND (DG) and BED (CA3) for recruitment as binding detector cell and binding-error detector circuit, respectively, (ii) the *expected number* of cells that will receive appropriate connections and will be candidates for recruitment for a binding, (iii) the estimated number of cells recruited per binding, and (iv) the expected number of recruited binding detector cells and binding-error detector circuits that will respond in a false

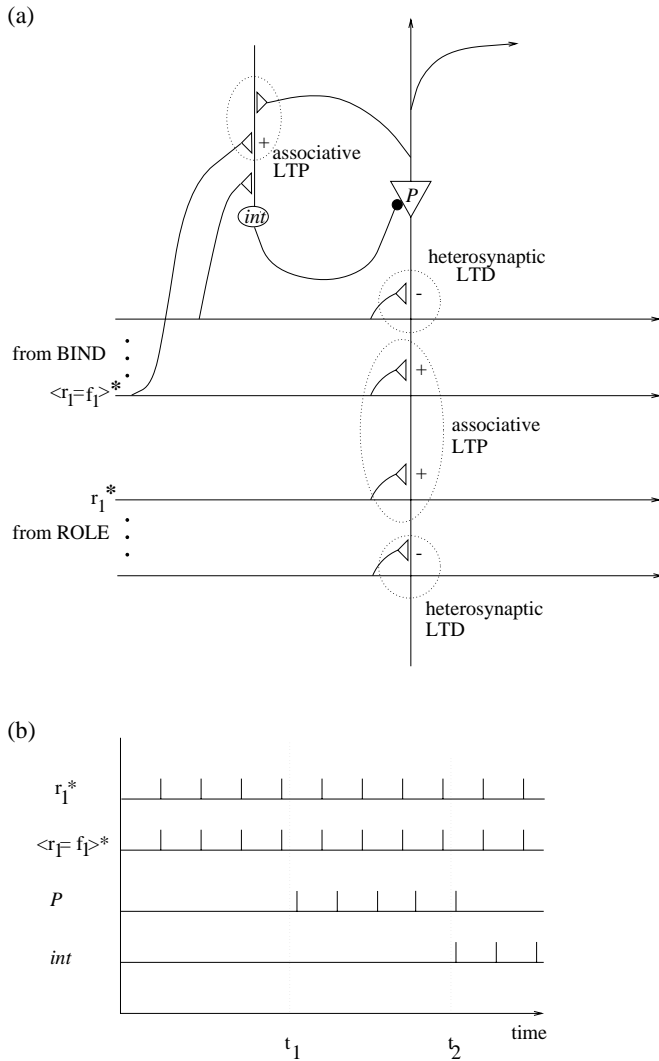


Figure 2: (a) A local inhibitory feedback circuit consisting of a principal cell P and a Type-2 interneuron int embedded within BED. Type-1 inhibitory interneurons are not shown. In general, each principal cell and inhibitory interneuron may participate in several feedback circuits. P receives afferents from r_1^* , a cell in the r_1 ensemble in ROLE, and $\langle r_1 = f_1 \rangle^*$, a binding detector cell in BIND for the binding $\langle r_1 = f_1 \rangle$. The latter also sends an afferent to int . The coincident activity arriving from r_1^* and $\langle r_1 = f_1 \rangle^*$ results in the formation of a binding-error detector circuit consisting of P and int for the binding $\langle r_1 = f_1 \rangle$ (see text). P 's response constitutes the response of the binding-error detector circuit. In general several circuits are recruited for each role-entity binding. (b) A schematic representation of the activity of P and int during the recruitment process. The LTP of P 's synapses has occurred by time t_1 , and the LTP of int 's synapse has occurred by time t_2 (κ is assumed to be 4).

Region	P_{fail}	$E\langle candidates \rangle$	$E\langle recruits \rangle$	$E\langle \text{False Positive} \rangle$	$E\langle \text{False Negative} \rangle$
BIND (DG)	$< 10^{-18}$	195.0	195.0	6.4	0
BED (CA3)	$< 10^{-18}$	412.7	16.3	0.5	2.4

Table 1: Failure probabilities and expected number of candidate cells and circuits for recruitment in BIND (DG) and BED (CA3), respectively. P_{fail} denotes the probability that no cells or circuits with suitable connections will be found for recruitment in a target region in response to a binding. $E\langle n \rangle$ denotes the expected number of cells that will be candidates for recruitment. $E\langle recruits \rangle$ specifies the average number of candidate cells and circuits recruited to each binding. Values are also computed for the expected number of binder cells and binding-error detector circuits that will respond erroneously to a cue in a false positive and false negative manner. The false positive response datum is for a binding-error detector circuit for the binding $\langle r_1 = f_1 \rangle$ responding to a cue containing the closely related bindings $\langle r_1 = f_2 \rangle$ and $\langle r_2 = f_1 \rangle$. These calculation assume that 200,000 distinct bindings have been memorized.

positive and a false negative manner, respectively, to a cue. The calculations assume that DG and CA3 contain 15 million and 2.7 million principal cells, respectively, and CA3 contains 270,000 Type-2 interneurons. The PF sizes for the projections from DG to CA3, EC to DG, and EC to CA3 are assumed to be 14, 17,000, and 6000, respectively. These choices are based in part on data provided in [1][20]. The PFs are assumed to be uniformly distributed over their respective target regions. Role and entity ensembles in the ROLE and ENTITY subregions of EC are assumed to contain 600 cells each. For a complete set of assumptions and LTP parameter values refer to [11].

5 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 destroy binding detector cells. Consequently, binding error detector circuits in CA3 will not be inhibited even when a cue specifies correct bindings. This will lead to erroneous “don’t know” responses. Behaviorally, this corresponds to forgetting. Large scale loss of CA3 pyramids or loss of perforant path inputs to CA3 will prevent the generation of binding-error signals, and hence, lead to false positive responses. Behaviorally, this corresponds to the existence of spurious memories. Large scale loss of mossy fiber inputs or loss of CA3 interneurons will prevent the blocking of binding-error signals, and hence, will lead to erroneous “don’t know” responses (forgetting). 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 even though they may be amnesic.

6 Discussion

This work demonstrates that a transient pattern of activity representing an event can be transformed rapidly into persistent circuits for detecting bindings and binding errors within a structure whose architecture and circuitry matches that of the HF. The work offers an alternative interpretation of the functional role of CA3 and its local inhibitory circuits. Most models of the HF 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. The structure for the formation of binding and binding-error detectors has been embedded within a detailed model of episodic memory formation, SMRITI [11].

Acknowledgments

This work was supported by the National Science Foundation grant SBR-9720398 and the Office of Naval Research grant N00014-93-1-1149 to the author.

References

- [1] Amaral, D.G. and Witter, M.P. Hippocampal Formation. In G. Paxinos ed. *The Rat Nervous System*, 2nd edn (Academic Press, London, 1995) 443–493.
- [2] Bliss, T.V.P. and Collingridge, G.L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993);
- [3] Cohen, N.J. and Eichenbaum, H. *Memory, Amnesia, and the Hippocampal System* (M.I.T. Press, Cambridge, Massachusetts, 1993).
- [4] Feldman, J. A. Dynamic connections in neural networks. *Bio-Cybernetics* **46**, 27–39 (1982).
- [5] Kawaguchi, Y. and Hama, K. Fast-spiking non-pyramidal cells in the hippocampal CA3 region, dentate gyrus and subiculum of rats. *Brain Res.* **425**, 351–355 (1987).
- [6] Knight, R.T. Contribution of human hippocampal region to novelty detection. *Nature* **382**, 256–259 (1996).
- [7] O’Keefe, J. and Nadel, L. *The hippocampus as a cognitive map*. (Clarendon Press, Oxford, 1978).
- [8] Marr, D. Simple memory: a theory for archicortex. *Phil. Trans. R. Soc. B* **262**, 23–81 (1971).
- [9] Miles, R. and Wong, R.K.S. Unitary inhibitory synaptic potentials in the guinea-pig hippocampus *in vitro*. *J. Physiol.* **356**, 97–113 (1984).
- [10] Schwartzkroin, P.A., Scharfman, H.E. and Sloviter, R.S. Similarities in circuitry between Ammon’s horn and dentate gyrus: local interactions and parallel processing. 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) 269–286.
- [11] Shastri, L. From transient patterns to persistent structures: a computational model of rapid memory formation in the hippocampal system. In preparation.
- [12] L. Shastri and V. Ajjanagadde. From simple associations to systematic reasoning. connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences* 16 (3) 417–494, 1993.
- [13] Singer, W. and Gray, C.M. Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.* **18**, 555–586 (1995).
- [14] Squire, L.R. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psych. Rev.* **99**, 195–231 (1992);
- [15] Treves, A and Rolls, E.T. Computational analysis of the role of the hippocampus in memory. *Hippocampus* **4**, 374–391 (1994).
- [16] Tulving, E. *Elements of Episodic Memory*. (Clarendon Press, Oxford, 1978).

- [17] Valiant, L. (1994). *Circuits of the mind*. (Oxford University Press, New York 1994).
- [18] Van Hoesen, G.W. The primate hippocampus gyrus: New insights regarding its cortical connections. *Trends Neurosci.* **5**, 345–350 (1982);
- [19] C. von der Malsburg. Am I thinking assemblies? In G. Palm and A. Aertsen eds. *Brain Theory*, (Springer-Verlag, Berlin, 1986).
- [20] West, M.J. 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.

Lokendra Shastri

Lokendra Shastri is a Member of the AI Group at the International Computer Science Institute (ICSI) Berkeley, CA. He is interested in neurally motivated computational models of learning, knowledge representation, and inference; spatio-temporal connectionist networks; the role of temporal synchrony in the expression of dynamic bindings and relational information processing; rapid memory formation in the hippocampal system; common-sense reasoning and its relation to abductive, deductive, and analogical reasoning. Before joining ICSI Berkeley, he was on the faculty of the University of Pennsylvania. He received a Bachelor of Engineering degree in Electronics with distinction from the Birla Institute of Technology and Science, India, an M.S. in Computer Science from the Indian Institute of Technology, Madras, and a Ph.D. in Computer Science from the University of Rochester in 1985.