Submitted to *Behavioral and Brain Sciences*

From transient patterns to persistent structures: A model of episodic memory formation via cortico-hippocampal interactions

Lokendra Shastri
International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, CA 94704, USA
E-mail: shastri@icsi.berkeley.edu
http://www.icsi.berkeley.edu/∼shastri

SHORT ABSTRACT

We readily acquire memories of events and situations in our daily lives. There is a broad consensus that the hippocampal system (HS) plays a critical role in the encoding and retrieval of such "episodic" memories. But how the HS subserves this mnemonic function is not fully understood. This article presents a computational model, SMRITI, that demonstrates how a transient pattern of activity representing an event can be transformed rapidly into a persistent and robust memory trace as a result of long-term potentiation within structures whose architecture and circuitry resemble those of the HS. Predictions and implications of the model are discussed.

LONG ABSTRACT

We readily remember events and situations in our daily lives and rapidly acquire memories of specific events by watching a telecast or reading a newspaper. There is a broad consensus that the hippocampal system (HS), consisting of the hippocampal formation and neighboring cortical areas, plays a critical role in the encoding and retrieval of such "episodic" memories. But how the HS subserves this mnemonic function is not fully understood. This article presents a computational model, SMRITI, that demonstrates how a cortically expressed transient pattern of activity representing an event can be transformed rapidly into a persistent and robust memory trace as a result of long-term potentiation within structures whose architecture and circuitry resemble those of the HS. Memory traces formed by the model respond to partial cues, and at the same time, reject similar but erroneous cues. During retrieval these memory traces, acting in concert with cortical circuits encoding semantic, causal, and procedural knowledge, can recreate activation-based representations of memorized events. The model explicates the representational requirements of encoding episodic memories, and suggests that the idiosyncratic architecture of the HS is well matched to the representational problems it must solve in order to support the episodic memory function. The model predicts the nature of memory deficits that would result from insult to specific HS components and to cortical circuits projecting to the HS. It also identifies the sorts of memories that must remain encoded in the HS for the long-term, and helps delineate the semantic and episodic memory distinction.

Keywords: cortico-hippocampal interactions; declarative memory; entorhinal cortex; episodic memory; hippocampus; long-term potentiation; recruitment learning; relational memory.

1 Introduction

We readily remember events and situations in our daily lives and acquire memories of specific events by reading a newspaper, watching a newscast, or participating in a conversation. This ability to rapidly acquire "episodic" memories (Tulving 1983; 1995) has been the focus of considerable research in psychology and neuroscience, and there is a broad consensus that this form of memory is distinct, both in its functional properties and in its neural basis, from other forms of memories involving common sense knowledge, perceptual-motor skills, priming, and simple classical conditioning (for a review of relevant experimental findings see Squire 1992; Cohen & Eichenbaum 1995; Schacter 1996b).[1]

Since episodic memories record ongoing experience, they must be acquired rapidly and as a result of a single occurrence. 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 distributed neural circuits. This expression, however, is per force transient and changes continually as we interact with our environment. Hence, the neural expression of a memorable event or situation must be transformed rapidly from a *transient* pattern of activity into a *persistent* structural encoding, or else it will be lost.

A wide array of neuropsychological, neuroanatomical, imaging, and neurophysiological data suggests that the hippocampal system (HS) consisting of the hippocampal formation (HF) and neighboring cortical areas in the ventromedial temporal lobe plays a critical role in the encoding and recall of events and situations (for example, Scoville & Milner 1957; Squire 1992; Cohen & Eichenbaum 1995; Corkin, Amaral, Gonzales, Johnson & Hyman 1997; Lepage, Habib & Tulving 1998; Fernandez, Effern, Grunwald, Pezer, Lehnertz, Dumpelmann, Roost & Elger 1999; Dolan & Fletcher 1999; Schacter & Wagner 1999; Tesche & Karhu 2000). Behavioral data shows that human patients with bilateral damage to the HS suffer from severe amnesia. Not only do they lose the ability to acquire novel episodic memories (anterograde amnesia), they also forget past memories of specific events and situations acquired over a period spanning decades prior to the damage (retrograde amnesia) (see Kartsounis, Rudge & Stevens 1995; Rempel-Clower, Zola, Squire & Amaral 1996; Nadel & Moscovitch 1997; Stefanacci, Buffalo, Schmolck & Squire 2000). Such patients, however, retain previously acquired semantic and procedural knowledge[2], exhibit some recognition ability based on familiarity (Aggleton & Brown 1999; Yonelinas 1997), continue to produce and understand language, demonstrate priming effects (Shimamura 1986), and acquire novel categories (Knowlton & Squire 1993) and procedural skills (Corkin 1968). Even patients who have suffered bilateral trauma to the HS at a very early age remain profoundly amnesic through life, and are unable to acquire episodic memories, even though they are able to acquire language skills, literacy, and semantic knowledge in the low average to average range (Vargha-Khadem, Gadian, Watkins, Connelly, Paesschen & Mishkin 1997).[3]

Another significant link between the HS and memory is suggested by the neuropathology of Alzheimer's disease, a progressive brain disorder whose early stages are marked by a loss of memory and confusion about recent events and situations. The HS is one of the first areas affected by Alzheimer's disease, and also one of the areas most severely afflicted by neurofibrillary tangles and neuritic plaques that are characteristic of the disease (Hyman & Van Hoesen 1989; Van Hoesen & Hyman 1990; Gomez-Isla, Price, McKeel, Morris, Growdon & Hyman 1996).

A number of researchers have proposed models to explain how the HS subserves the episodic memory function. These models include macroscopic system-level models that attempt to describe the functional role of the HS (for example, O'Keefe & Nadel 1978; Olton, Becker & Handelmann 1979; Wickelgren 1979; Mishkin & Petri 1984; Halgren 1984; Rawlins 1985; Teyler & DiScenna 1986; Squire & Zola-Morgan 1991; Eichenbaum, Otto & Cohen 1994; Johnson & Chalfonte 1994; Moscowitch 1994; Cohen & Eichenbaum 1995; Kroll, Knight, Metcalfe & Wolf 1996; Morris & Frey 1997; Nadel & Moscovitch 1997; Yonelinas 1997; Aggleton & Brown 1999; Lisman 1999), as well as more detailed computational models that attempt to explicate *how* the HS might realize its putative function (for example, Marr 1971; McNaughton & Morris 1987; Lynch & Granger 1992; Schmajuk & DiCarlo 1992; Carpenter & Grossberg 1993; Gluck & Myers 1993; Metcalfe 1993; Alvarez & Squire 1994; O'Reilly & McClelland 1994; Treves & Rolls 1994; Hasselmo & Stern 1995; Granger, Wiebe, Taketani & Lynch 1996; Levy 1996; McClelland & Goddard 1996; Murre 1996; Treves, Skaggs & Barnes 1996; Moll & Miikkulainen 1997; Menschik & Finkel 1998;

[1] The sort of memory that is the focus of this work has also been described as *declarative memory* (Cohen & Eichenbaum 1995) and *explicit memory* (Graf & Schacter 1985).

[2] Semantic knowledge subsumes generic knowledge about concepts and categories (for example, Emus are birds), as well as causal knowledge capturing systematic relationships (for example, if you buy something you own it). Although semantic knowledge can pertain to individual entities (the Eiffel tower is in Paris), it often involves generalizations and abstractions stemming from a number of observations (for example, parents of young children often own a minivan). Furthermore, semantic knowledge is typically devoid of "source information" indicating where, when, or how the knowledge was obtained. In contrast, procedural knowledge refers to perceptual-motor skills such as riding a bicycle or playing tennis.

[3] Links between the role of the HS in humans and other animals are discussed in Section 12.1.

Nadel, Samsonovich, Ryan & Moscovitch 2000). While our understanding of the HS and its potential role in memory formation has been enhanced by this extensive body of work, several key representational problems associated with the encoding of specific events and situations 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 as argued in Section 4, this view of episodic memory is representationally inadequate for dealing with events and situations.

This paper presents a computational model, SMRITI[4], that addresses some of the unresolved representational issues concerning the encoding and retrieval of episodic memory. SMRITI explains how a *transient* pattern of cortical activity encoding an event or a situation may be transformed rapidly into a *persistent* and robust memory trace in the HS as a result of long-term synaptic potentiation (Malenka & Nicoll 1999).

Episodic memory traces formed by the model respond to partial cues, and at the same time, reject similar but erroneous cues. During retrieval, these memory traces, acting in concert with cortical circuits encoding semantic, causal, and procedural knowledge, can *recreate* activation-based representations of memorized events in high-level cortical circuits (HLCCs).

The episodic memory trace of an event formed in SMRITI is sparse, yet physically *dispersed* and highly *redundant*. While the sparseness of the encoding enables the model to memorize a large number of events with minimal cross-talk, the physically dispersed and redundant nature of the encoding makes the model robust against cell loss.

SMRITI's architecture provides a rationale for various components of the HS and their interactions, and suggests that the idiosyncratic architecture of the HS is well matched to the representational requirements of episodic memory function. The model offers an explanation for the existence of multiple pathways within the HS, backprojections from CA1 and the subiculum to EC, and extensive feedforward and feedback local inhibitory circuits found in the HS. In particular, SMRITI suggests that local inhibitory circuits serve as critical functional elements in an event's episodic memory trace.

The model helps delineate the distinction between semantic and episodic memory, and identifies the sorts of memories that must continue to be encoded in the HS and are not "transferred" to the cortex via a process of consolidation (Squire 1992; McClelland, McNaughton & O'Reilly 1995; Nadel & Moscovitch 1997; Bontempi, Laurent-Demir, Destrade & Jaffard 1999; Teng & Squire 1999).

Finally, SMRITI makes several predictions about the nature of encoding and retrieval deficits that would result from damage to various components and pathways of the HS and from cell loss in cortical circuits encoding semantic knowledge. In doing so, it also explains differences in encoding and retrieval deficits observed in hippocampal patients and those observed in semantic dementia patients (Hodges, Patterson, Oxbury & Funnel 1992).

1.1 A system-level description of SMRITI

At a macroscopic level, the functioning of the model may be described as follows (refer to Figure 1): Our cognitive apparatus construes our experiences as a stream of events and situations. These construals are the result of complex interactions between sensory, perceptual, categorical, linguistic, and inferential processes, and 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 in the HS. The resulting activity can be viewed as the "presentation" of an event to the HS by HLCCs for possible memorization. Alternately, HLCCs may present a "query" to the HS and expect a certain type of response if the query matches one of the events previously memorized by the HS, and a qualitatively different type of response if it does not. In case of a positive response, the HS would reinstate the matching event as a pattern of activity over HLCCs.[5]

The transient activity injected into EC by HLCCs propagates around the complex loop consisting of EC, dentate gyrus (DG), fields CA3, CA2, and CA1 of Ammon's horn, the subiculum and EC, and triggers a complex sequence of synaptic changes in these structures. The model demonstrates how such synaptic changes can transform the transient pattern of activity into a persistent structural encoding (an episodic memory trace) composed of the requisite functional circuits mentioned in Section 4.2. The activity arriving from CA1 and the subiculum into EC constitutes the response of the HS. This reentrant activity in EC propagates back to HLCCs and completes a cycle of cortico-hippocampal interaction.

[4] The name is an acronym for "**S**ystem for the **M**emorization of **R**elational **I**nstances from **T**ransient **I**mpulses".

[5] At a macroscopic level of description, the cortico-hippocampal interaction envisioned above is similar to that assumed by other models of the HS-based memory system (for example, see Marr 1971; Halgren 1984; Teyler & DiScenna 1986; Damasio 1989; Squire & Zola-Morgan 1991; Alvarez & Squire 1994; O'Reilly & McClelland 1994; Rudy & Sutherland 1994; Treves & Rolls 1994; Cohen & Eichenbaum 1995; Hasselmo & Stern 1995; Knight 1996; McClelland & Goddard 1996; Murre 1996; and Nadel & Moscovitch 1997).

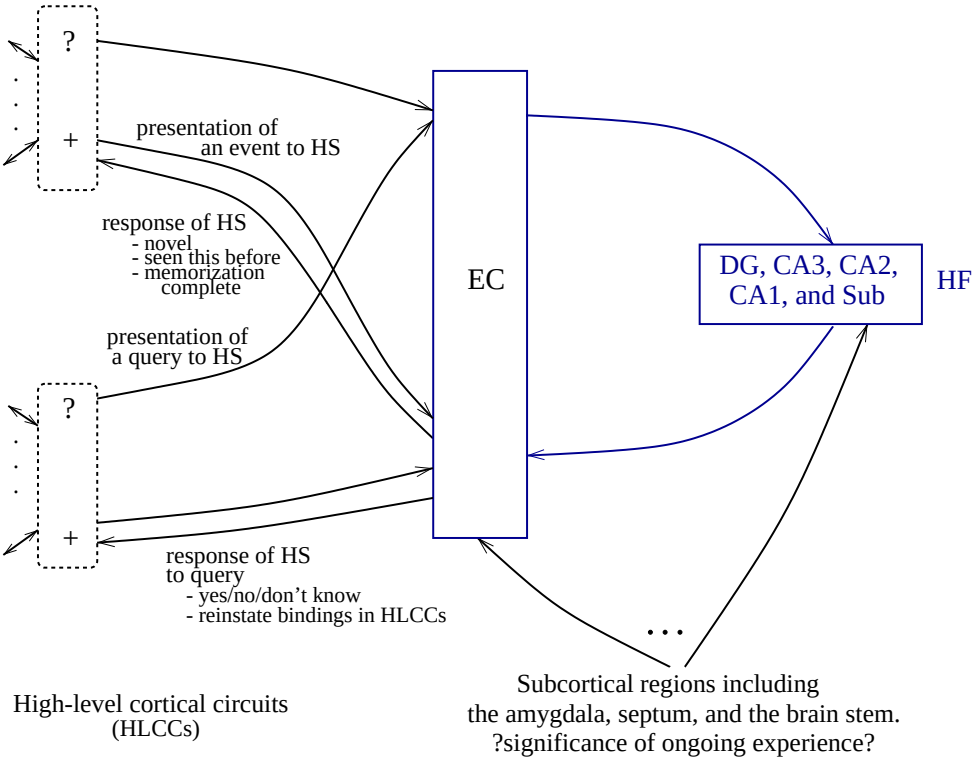

Figure 1: A schematic of the cortico-hippocampal interactions assumed by the model. See text for details.

The proper functioning of the HS depends on its interactions with neural circuits and subsystems concerned with emotion, motivation, arousal, and attention. SMRITI assumes that the presentation of an event to the HS is accompanied by additional signals that indicate how "significant" the event is to the organism. These signals are diffuse and have a graded (modulatory) impact on the number of cells recruited for encoding an event's memory trace. The HS receives a rich set of afferents from subcortical structures and it is likely that some of these inputs serve as "significance" signals (for example, Vertes & Kocsis 1997; Hasselmo, Wyble & Wallenstein 1996; McGaugh 2000).

1.2 Outline of the paper

Section 2 gives a brief description of the organization and circuitry of the HS. Section 3 discusses the phenomena of long-term potentiation, and specifies the computational abstraction of cells, synapses, and synaptic modification used in SMRITI. Section 4 discusses the representational requirements of the episodic memory system, and Section 5 reviews how an event may be expressed as a transient pattern of activity over cortical circuits. Section 6 provides a circuit-level description of the model and explains *how* structures with the requisite functional circuits emerge rapidly within the HS in response to a transient pattern of rhythmic activity. Issues of memory consolidation and forgetting are discussed in Section 7. A quantitative analysis of the model's memory capacity and its robustness against diffuse cell loss and cross-talk is presented in Section 8. Section 9 discusses interactions between cortical and hippocampal representations during retrieval and inference, and Section 10 considers the issue of information transfer from the HS-based episodic memory traces to cortical structures representing semantic and causal knowledge. Section 11 lists several predictions of the model, and Section 12 concludes with a discussion of open issues and future directions.

2 The hippocampal system: an overview

The hippocampal system (HS) refers to a heterogeneous collection of neural structures buried deep within the cerebral hemispheres. These structures include the entorhinal cortex (EC), the dentate gyrus (DG), Ammon's horn (or the hippocampus proper), and the subiculum. Ammon's horn consists of three distinct fields, namely, CA1, CA2, and

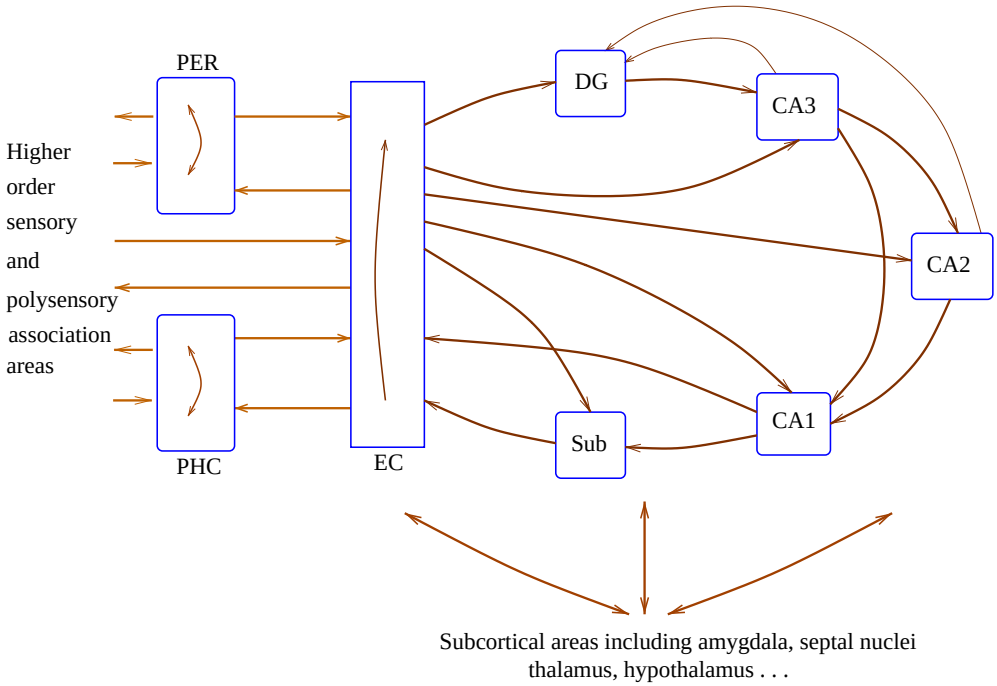

Figure 2: Summary of inputs to the hippocampal system (HS) and the major pathways interconnecting components of the HS. The pre- and parasubiculum regions are not shown. Abbreviations: PER, perirhinal cortex; PHC, parahippocampal cortex; EC, entorhinal cortex; DG, dentate gyrus; CA1, CA2, and CA3, fields of Ammon's horn; Sub, subiculum.

CA3. 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 (Braak 1974; Duvernay 1988). Other cortical areas intimately related to the HS are the perirhinal and parahippocampal cortices and the presubiculum and parasubiculum regions. The Ammon's horn and DG together form a distinctive sea-horse shaped structure that arches around the midbrain and is referred to as the hippocampus. The hippocampus together with the subiculum is referred to as the hippocampal formation (HF).

The following observations highlight some of the salient features of the HS and are largely drawn from (Braak 1974; Duvernay 1988; Amaral & Insausti 1990; Amaral & Witter 1995; Johnston & Amaral 1998). These observations relate primarily to the human HS, but include guarded inferences drawn from what is known about the monkey HS, and in some instances, the rat HS.[6]

2.1 Component regions of the HS and their interconnections

Figure 2 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, polymodal association areas, as well as supramodal association areas project to EC either directly or via the perirhinal and parahippocampal cortices (Van Hoesen 1982; Horel 1988; Insausti, Amaral & Cowan 1987; Suzuki & Amaral 1994).[7] Thus EC is the locus of converging activity resulting from considerable processing and integration of ongoing sensory experience, and it is plausible to assume that this activity encodes the agent's construal of its ongoing experience in terms of events and situations. In turn, upper layers of EC project to DG, CA3, CA2, CA1, and the subiculum; DG projects to CA3; CA3 projects to CA2 and CA1; CA2 projects to CA1; and CA1 projects to the subiculum. CA1 and the subiculum

[6] Inferences about the human HS based on data about the rat HS should be drawn with caution since there are significant differences between the two brains. The close homology of the primate and the human brain makes it more appropriate to draw guarded inferences about the latter based on the former. However, caution is warranted since there are subtle differences between the two brains.

[7] These cortical areas include the superior temporal gyrus, inferior temporal cortex, the dorsal (and to a lesser extent) the ventral bank of superior temporal sulcus, cingulate cortex (including retrosplenial cortex), insular cortex, parietal area 7a and the lateral intraparietal area (LIP), orbitofrontal cortex, dorsolateral prefrontal cortex, and medial frontal cortex.

project back to the deeper layers of EC which in turn projects back to high-level cortical areas that project to it. These backprojections are either direct, or via the perirhinal and parahippocampal cortices.

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 the high-level cortical regions from where it originated.[8]

In addition to the HS 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 a number of subcortical regions[9] that mediate arousal and other autonomic, emotional, and motivational aspects of behavior including the perception of fear and pleasure/reward (Amaral & Cowan 1980; Insausti, Amaral & Cowan 1987; Horel 1988). The HS in turn projects back to many of the subcortical structures that project to it. These subcortical inputs are capable of communicating to the HS the internal state of the agent (organism) as well as the *affective significance* of the agent's on going experience, and hence, play an important modulatory role in episodic memory formation (for example, Vertes & Kocsis 1997; Hasselmo, Wyble & Wallenstein 1996; McGaugh 2000). Basal forebrain lesions disconnecting cholinergic pathways to the HS may lead to anterograde amnesia in humans (Abe, Inokawa, Kashiwagi and Yanagihara 1998). It is also known that in the case of emotionally arousing material, there is a good correlation between the degree of amygdala activation *at the time of encoding* and how well the material is recalled subsequently (Cahill, Haier, Fallon, Alkire, Tang, Keator, Wu & McGaugh 1998; Canli, Zhao, Brewer, Gabrieli & Cahill 2000).[10]

EC has six layers (I–VI). Of these, layers II, III, V, and VI have a high density of cells. Cortical inputs to EC reach the superficial (II and III) as well as the deep (IV and V) layers of EC (Witter et al. 2000). Layers II and III are the primary source of projections from EC to the rest of the HS, with stellate cells in layer II projecting to DG, CA3, CA2, and the subiculum, and pyramidal cells in layer III projecting to CA1 and the subiculum. Cells in CA1 and the subiculum project back to the deeper layers of EC, which in turn project back to the very cortical areas that project to EC. Moreover, cells in deeper layers of EC send axonal collaterals to cells in the upper layers of EC. Various layers of EC also contain a variety of inhibitory interneurons that participate in local inhibitory circuits. In addition to its laminar structure, EC is organized into several distinct cytoarchitectonic subfields. For example, Insausti, Tunon, Sobreviela, Insausti & Gonzalo (1995) have suggested a division of the human EC into eight subfields.

DG is composed of the molecular, granular, and polymorphic layers. The granular cell layer contains granule cells that are the principal cells of DG, and basket cells that are inhibitory interneurons. The molecular layer is relatively cell free, but contains some inhibitory interneurons (for example, axo-axonic chandelier cells). The polymorphic layer contains (excitatory) mossy cells and several types of inhibitory interneurons. The dendrites of granule cells extend into the molecular layer where they receive afferents from EC via the so called *perforant path* projection.

Granule cell axons – the mossy fibers — make synapses with mossy cells and inhibitory interneurons in the polymorphic layer and then continue on to make synaptic contacts with pyramidal cells and inhibitory interneurons in CA3. Mossy cell axons make widespread contacts with granule cells and with inhibitory interneurons that form synapses on granule cells. Thus granule cells, mossy cells, and inhibitory interneurons form numerous excitatory as well as inhibitory feedback circuits within DG (Schwartzkroin, Scharfman & Sloviter 1990; Buckmaster & Schwartzkroin 1995; Jackson & Scharfman 1996).

Ammon's horn is composed of the alveus, oriens, pyramidal, radiatum, lacunosum, and molecular layers (Duvernay 1988). The lacunosum and molecular layers are jointly referred to by many authors as the lacunosum-molecular layer. CA3 also has an additional layer — stratum lucidum — situated between the pyramidal and radiatum layers. Mossy fibers originating in DG travel through this layer making synaptic contacts with CA3 pyramidal cells and interneurons. Pyramidal cells located in the pyramidal layer are the principal cell type of Ammon's horn. The axons of these cells travel to the alveus and project to subcortical regions via the fimbria and fornix. Pyramidal cell axons also give off numerous collaterals that make contacts with other pyramidal cells and interneurons in the Ammon's horn.

Ammon's horn contains several types of inhibitory interneurons in large numbers (Lacaille, Kunkel & Schwartzkroin 1989; Freund & Buzsaki 1996). These include basket cells, axo-axonic cells, O/A interneurons (located at the junction

[8] The EC mediated reciprocal flow of cortico-hippocampal activity is complemented by direct projections from the CA1 and the subiculum to perirhinal and parahippocampal cortices, and vice versa (Witter, Naber, van Haeften, Machielsen, Rombouts, Barkhof, Scheltens, Lopes da Silva 2000; Lavenex & Amaral 2000). CA1 and the subiculum also project directly to the medial prefrontal and orbitofrontal cortices (Laroche, Davis & Jay 2000).

[9] These subcortical regions include the amygdaloid complex, claustrum, septum, substantia innominata, thalamus, hypothalamus, and several brainstem structures.

[10] In addition to subcortical regions, the retrosplenial cortex may also be involved in linking emotions with memory (Maddock 1999).

of oriens and alveus layers), and L-M interneurons (located in the lacunosum and molecular layers). Different types of interneurons receive different types of inputs depending on the location of their dendrites, and different types of interneurons target different parts of pyramidal cells. Synapses formed by mossy fibers on CA3 inhibitory interneurons, by CA3 associational collaterals on inhibitory interneurons, and by inhibitory interneurons on pyramidal cells, together give rise to a large number of feedforward and feedback inhibitory circuits.

The CA3 $\rightarrow$ DG mossy cell $\rightarrow$ DG granule cell pathway is *overtly* expressed only under conditions that are not related to the acquisition of new memory (Scharfman 1994; Penttonen, Kamondi, Sik, Acsady & Buzsaki 1997). This suggests that the *functional* role of the CA3 to DG feedback may not be to directly cause granule cells to fire. Rather, it may be to provide a subthreshold *bias* signal that modulates the firing of granule cells and the induction of LTP at granule cell synapses, in response to perforant path activity.

The subiculum is populated by pyramidal cells and inhibitory interneurons. The subiculum receives major inputs from CA1 and the superficial layers of EC. It also receives afferents from several subcortical regions. The subiculum sends a major projection to the deep layers of EC. Together with the projection from CA1 to EC, this projection constitutes the major response of the HS. The subiculum also gives rise to projections to pre- and parasubiculum, and several prominent projections to subcortical areas and to cortical regions including the medial prefrontal, retrosplenial, and perirhinal cortices.

2.2 Some quantitative findings pertaining to the HS

West (1990) estimates the number of neurons in regions of the human HS to be as follows: DG (not including the hilus/polymorphic layer) — 15.4 million; hilus/polymorphic layer — 1.98 million; *regio inferior* (CA3/CA2)— 2.70 million, *regio superior* (CA1) — 16.4 million, and the subiculum 4.51 million. West and Slomianka (1998a,b) estimate the number of neurons in the upper layers (II and III) and lower layers (V and VI) of EC to be 4.22 million and 3.78 million, respectively. According to Olbrich and Braak (1985), about 9.4% of CA1 cells are interneurons, and the remaining are pyramidal cells.

The projections between components of the HS have widely different densities and strengths. For example, the projection from EC to DG is extremely dense — a stellate cell in EC may connect to about 17,000 granule cells in DG (Amaral, Ishizuka & Claiborne 1990), but the projection from DG to CA3 is extremely sparse — a granule cell in DG may make contacts with only 14 pyramidal cells in CA3 (Amaral & Witter 1995). In contrast, the strength of synapses formed by the EC to DG projection is quite modest, but the strength of DG to CA3 synapses is extremely high. Thus while synchronous activity in several EC to DG fibers is required to discharge a granule cell, a CA3 pyramidal cell can fire upon receiving input from a single granule cell, perhaps with some contribution from subcortical and perforant path inputs (Acsady, Kamondi, Sik, Freund & Buzsaki 1998).

2.3 The *theta* rhythm and the HS

As a rat explores its environment, its EEG shows a prominent and nearly sinusoidal oscillation of 4-10 Hz in the HS. A similar oscillation is observed during rapid eye movement sleep (see Vertes & Kocsis 1997, for a review). During such *theta* activity, subsets of hippocampal pyramidal cells fire complex spike bursts that are phase-locked to *theta* activity. In addition to the hippocampi of rat and other mammals, *theta* activity related to memory function is also observed in the human hippocampi (for example, see Klimesch 1999; Tesche & Karhu 2000). It has been proposed that *theta* activity is induced in the HF by the coordinated firing of cholinergic and GABAergic cells in the medial septum and the diagonal band[11], and by *theta* activity generated in EC.

The induction of LTP in the HF is most effective when a suitable stimulation is delivered during the positive phase of a *theta* cycle (in particular, at the peak of the cycle) (Pavlides, Greenstein, Grudman & Winson 1988; Huerta & Lisman 1995; Vertes & Kocsis 1997).

In agreement with Vertes & Kocsis (1997) we speculate that for an event to be memorized by the HS, the communication of the event's pattern of activity to the HS by HLCCs should co-occur with *theta* activity induced in the HS by subcortical inputs. In fact, subcortically induced *theta* activity might be the primary component of a graded significance signal that modulates the mass of cells recruited for encoding an event's memory trace (see Section 6.5).

[11] These cholinergic and GABAergic cells are in turn driven by activity in the reticular formation and the supramammillary nucleus.

3 Long-term potentiation and depression

In SMRITI, the encoding of an event's memory trace involves the rapid formation of circuits with specific functionalities. These circuits are carved out of existing networks of excitatory cells and inhibitory interneurons as a result of changes in synaptic strengths induced by the propagation of coherent activity through the HS. Thus the functioning of SMRITI requires a mechanism for activity dependent synaptic modification having the following general properties: (i) the arrival of coincident activity at a cell should lead to changes in synaptic strengths, (ii) these changes should be induced rapidly in response to transient activity lasting no more than a few seconds, (iii) once induced, these changes should persist for a long time, and (iv) these changes should be synapse specific so as to allow the formation of circuits with specific functional properties.

The cellular mechanisms of long-term potentiation (LTP) (Bliss & Lomo 1973; Malenka & Nicoll 1999) and long-term depression (LTD) (Lynch, Dunwiddie & Gribkoff 1977; Linden 1994) possess all of the above properties, and almost certainly play a direct causal role in learning and memory formation (for example, Tang, Shimizu, Dube, Rampon, Kerchner, Zhuo, Liu & Tsien 1999; Rioult-Pedotti, Friedman & Donoghue 2000; but see Shors & Matzel 1997). Consequently, learning in SMRITI is based on LTP and LTD. LTP was first observed in the rabbit HF, and has since been observed in synapses along all excitatory pathways in the mammalian HF as well as along many excitatory pathways in the mammalian brain, including cortico-hippocampal pathways (for example, see Ivanco & Racine 2000).

LTP involves a long-term increase in synaptic efficacy resulting from the pairing of presynaptic activity with postsynaptic depolarization. The most extensively studied form of LTP involves the unusual receptor NMDA (*N*-methyl-D-aspartate) which is activated by the excitatory neurotransmitter glutamate, but only if the postsynaptic membrane in which the receptor is embedded is sufficiently depolarized.[12] In the absence of adequate depolarization, NMDA receptor-gated channels remain blocked by magnesium ions in spite of glutamate being bound to the receptor. Adequate depolarization of the postsynaptic membrane, however, expels the magnesium ions and unblocks the channels. Once the channels are unblocked, calcium ions flood into the dendritic spine of the postsynaptic cell and trigger a complex series of biochemical changes that result in the induction of LTP.

The two conditions required for the activation of NMDA receptor, namely, presynaptic activity and strong postsynaptic depolarization, together entail that the LTP of a synapse requires the concurrent arrival of activity at several other synapses of the postsynaptic cell. This is referred to as the *cooperativity* property of LTP (McNaughton, Douglas & Goddard 1978; Levy & Steward 1979). The cooperativity property of LTP makes it an ideal mechanism for transforming a *transient* expression of a relationship between two items (encoded as the coherent activity of two ensembles) into a *persistent* expression of this relationship (encoded via increased efficacy of synapses linking the two ensembles).

LTP resulting from the arrival of coincident activity along one set of afferent fibers is referred to as *homosynaptic* LTP. If the arrival of coincident activity along two sets of fibers, A and B, leads to the LTP of synapses formed by fibers of A, but the arrival of activity along fibers of A alone does not, then the LTP of synapses formed by fibers of A is referred to as *associative* LTP (Levy & Steward 1979; Brown, Kairiss & Keenan 1990).

In addition to LTP, synapses along key excitatory pathways in the mammalian HF have been shown to undergo LTD. The absence of presynaptic activity in the presence of strong postsynaptic activity can lead to *heterosynaptic* LTD of a synapse (Lynch, Dunwiddie & Gribkoff 1977). Finally, prolonged low frequency stimulation of a synapse can lead to its homosynaptic LTD (Dudek & Bear 1992).

Experimentalists have compiled a catalog of stimulus conditions that induce LTP. However, only some of these conditions relate well to patterns of activity recorded in the hippocampus of behaving animals engaged in exploration and learning. One such condition is the *theta burst stimulation* consisting of a brief but high-frequency burst of pulses (for example, 4 pulses at 100Hz) repeated with an interburst interval of ca. 200 msec. (Larson, Wong & Lynch 1986; Staubli & Lynch 1987). Such an interburst interval corresponds to a rhythmic activity in the *theta* band (see Section 2.3); hence, the descriptor "theta burst stimulation." Furthermore, as stated in Section 2.3, stimulation delivered during the positive phase of a *theta* cycle, and in particular, at the positive peak of the cycle, is most effective in inducing LTP.

SMRITI uses a computational abstraction of LTP and LTD (Shastri 2001a). This abstraction is described in brief below.

[12] Not all forms of LTP are NMDA receptor-dependent. The LTP of synapses formed by mossy-fibers on CA3 pyramidal cells is a case in point (Nicoll & Malenka 1995).

3.1 A computational abstraction of LTP and LTD

The computational abstraction of LTP and LTD used in SMRITI is a highly simplified idealization of the complex biophysical processes underlying the induction and expression of LTP and LTD. The abstraction, nevertheless, captures key temporal and cooperative properties of LTP and LTD, and at the same time, makes it possible to carry out quantitative analyses and efficient computer simulations of large neuronal networks.

The abstraction of LTP and LTD is based on an abstraction of cells as integrate-and-fire neurons. The spatio-temporal integration of activity arriving at a cell is modeled as follows:

Let $a_i(t)$ be a measure of presynaptic activity occurring at synapse s_i of the cell at time t. In biophysical terms, $a_i(t)$ may correspond to the number of spikes arriving at s_i within a unit time interval anchored at t. Let $w_i(t)$ refer to the weight of synapse s_i at time t. The postsynaptic potential, $psp_i(t\,|a_i(t_0))$, resulting from the presynaptic activity at s_i at time t_0 is modeled as a piecewise linear function consisting of a ramp-up segment, a flat segment, and a decay segment; the height of the ramp being $a_i(t_o) * w_i(t_o)$. This postsynaptic potential is fully characterized by a small number of parameters, one of them being ω_{int}, the temporal extent of $psp_i(t\,|a_i(t_0))$ (in other words, ω_{int} is the temporal window over which two incident activities can summate).

$psp_i(t)$, the postsynaptic potential at time t attributable to s_i, equals: $\sum_{(0\leq\tau<\omega_{int})} psp_i(t\,|a_i(t-\tau))$, and $pot(t)$, the cell's potential at time t resulting from the combined effect of presynaptic activity at all its synapses, equals: $\sum_i psp_i(t)$, (the sum being taken over all synapses of the cell).

A cell has a firing threshold, $thresh_f(t)$, with a resting value of θ_f. A cell fires at time t if $pot(t) \geq thresh_f(t)$. The resulting spike arrives at downstream synapses after a variable propagation delay. After firing, a cell enters a refractory state for a duration ω_{ref} wherein it does not fire, irrespective of the inputs.

A cell can have two firing modes: *supra-active* and *normal*. These modes are associated with different firing thresholds and output levels. The *supra-active* mode corresponds to a high-frequency burst response such as the complex spike burst response generated by hippocampal pyramidal cells, and the *normal* mode corresponds to a simple spike response consisting of isolated spikes. The proposed abstraction of the distinction between a complex spike burst response and a simple spike response is a gross simplification, but for suitable choice of parameter values this abstraction offers a computationally inexpensive yet functionally adequate means of modeling these two firing modes.

A synapse can be in any one of the following three states: *naive*, *potentiated*, or *depressed*. The state of a synapse signifies its strength (weight). The induction of LTP is governed by the following parameters: the *potentiation threshold* θ_p, the *weight increment* Δw_{ltp}, the *repetition factor* κ, and the *maximum inter-activity interval* τ_{iai}.

Consider a set of neighboring synapses $s_1, \ldots, s_n$ sharing the same postsynaptic cell. Convergent presynaptic activity at $s_1, \ldots, s_n$ can lead to LTP of naive s_is and increase their weights by Δw_{ltp} if the following conditions hold:

- $\sum_{1\leq i\leq n} psp_i(t) \geq \theta_p$, that is, the presynaptic activity arriving at neighboring synapses must be "synchronous" (the lead/lag in incident activity at any pair of synapses should be $\leq \omega_{int}$),
- such synchronous presynaptic activity should repeat $\geq \kappa$ times, and
- the interval between two *successive* volleys of presynaptic activity at a synapse should be $\leq \tau_{iai}$ apart.

LTD is modeled in an analogous manner (see Shastri 2001a).

The effect of a neuromodulator is modeled as a region-wide bias signal that modifies the firing thresholds (θ_f and θ_{sf}) of cells and potentiation thresholds (θ_p) of synapses, respectively. For example, consider the effect of subcortical cholinergic/GABAergic inputs to the HS. Recall that these inputs are believed to contribute to hippocampal *theta* activity and may serve as a form of significance signal (Section 2.3), and LTP is facilitated when presynaptic activity occurs in the positive phase (especially, at the positive peak) of a *theta* cycle. The effect of these subcortical inputs can be modeled as a bias signal whose amplitude varies with the phase of the *theta* cycle, and whose *peak* amplitude is proportional to the significance of the event being experienced. The more positive the bias, the lower the effective value of θ_p, and the greater the likelihood that a synapse undergoes LTP. Similarly, the more positive the bias, the greater the likelihood that a cell fires/bursts.

3.2 Emergence of cells and circuits with specific functionalities: recruitment learning

LTP and LTD can transform a loosely organized network of neurons into a network of cells and circuits tuned to specific functionalities. As discussed in (Shastri 2001a), the formation of functional structures within loosely orga-

nized networks via LTP and LTD provides a biological basis for "recruitment learning" (Wickelgren 1979; Feldman 1982; Shastri 1988; Valiant 1994; Diederich & Hogan 1997; Page 2000). Recruitment learning can be described informally as follows: Learning occurs within a partially structured network containing a large number of randomly interconnected nodes. *Recruited* nodes in such a network are nodes that have acquired distinct functionality by virtue of their *strong* interconnections to other recruited nodes and/or other sensorimotor (input/output) nodes. Nodes not yet recruited are *free* nodes. These nodes are connected via weak links to a large number of free, recruited, and/or sensorimotor nodes. Free nodes form a primordial network from which suitably connected nodes may be recruited for representing new concepts.

4 Episodic memory and its representational requirements

Our cognitive apparatus construes the bulk of our experience in terms of events and situations, and our episodic memories are a partial record of these construals. Typically, episodic memories record who did what to whom where and when (for example, John gave Mary a book in the library on Tuesday). Alternately, they may describe a state of affairs wherein multiple entities occur in a particular configuration (for example, Bill sat next to Tom at the dinner table), or they may record the state of an entity (for example, the pie I ate after dinner was hot). The term entity is being used here in a broad sense and includes, among other things, specific instances/individuals (for example, Charles Darwin, my house), non-specific instances of categories (for example, a person, a house), and also individual categories when viewed as a whole.

An episodic memory may be acquired by directly experiencing an event; by seeing a video recording, a photograph, or an illustration of the event; or by reading or hearing its verbal description. In the proposed model, a key characteristic of episodic memories is that they are about specific events and situations located in a particular spatio-temporal context.[13]

Let us work toward identifying some basic representational requirements of encoding an event or a situation (henceforth, simply an event) in episodic memory. Consider the event where you see John give Mary a book in the library on Tuesday. This event is an instance of a specific sort of interaction involving John, Mary and a book that occurs in a particular location (the library) and at a particular time (Tuesday). John and Mary are performing specific "roles" in this interaction; John is the one who is doing the giving, and Mary is the one who is doing the receiving. Moreover, a book is the object being given by John to Mary. Clearly, this event cannot be expressed in the mind/brain as a mere association between John, Mary, a book, the library, and Tuesday. *At a bare minimum*, the memory trace of this event must encode some sort of *relational* structure wherein the role *giver* is bound to (the concept) John, the role *recipient* is bound to Mary, the role *object* is bound to a book, the role *location* is bound to the library, and the role *temporal-location* is bound to Tuesday. This relational information can be expressed succinctly as follows:[14]

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

Note that the above encoding of an event involves **two** levels of bindings: (1) `entities` occurring in the event are bound to the respective *roles* they fill in the event, and (2) all of the role-entity bindings pertaining to the event are grouped together in order to distinguish them from role-entity bindings pertaining to other events. Such an encoding is more complex than one that only chunks together, or forms a *conjunctive* representation of, the concepts involved in the event.

In certain cases, the encoding of an event may require the specification of *parameter values* pertaining to relevant sensorimotor schemas. For example, the encoding of an event such as "John hit the ball hard" may require the specification of parameter-value bindings such as ⟨*force-magnitude*=`high`⟩ in addition to role-entity bindings such as ⟨*hitter*=`John`⟩ and ⟨*hit-object*=`a Ball`⟩. In the context of episodic memory trace formation, however, the encoding of a role filler, a parameter value, a state-variable value, or an object location pose similar problems since all of these involve the binding of two items. Therefore, we will simplify matters and typically refer to all such bindings as role-entity bindings.

[13] Additionally episodic memory might also encode inferred events and planned events. As an example of an inferred event, consider hearing that streets in Boston were flooded in the morning. Upon hearing this, one might infer that it rained heavily in Boston in the morning. Depending on its significance, this inferred event may get encoded in episodic memory. As an example of a planned event, consider the specific memory I may have of my plan to attend my child's soccer game next Saturday afternoon at 4pm in Cedar Park.

[14] Examples used here and elsewhere in the paper are meant to be illustrative and do not lay claim to the actual choice of relational schemas, roles, and entities in our conceptual apparatus.

4.1 Do bindings suffice for encoding events?

The above observations suggest that the memorization of an event in episodic memory requires — *at the very least* — an encoding consisting of role-entity bindings. But is a representation consisting of a few role-entity bindings sufficient to capture the rich memory of an event? Events are not static snap-shots. They extend over time and space, and involve complex actions and reactions. Events can be sensorially rich and emotionally charged, and remembering them can evoke vivid images and arouse strong emotions. In view of this, it would seem that the memory trace of an event should involve much more than the memorization of a few bindings.

Indeed, events are dynamic and complex objects, but as argued below, it is possible to reconstruct a vivid representation of an event in the mind/brain by activating the web of semantic, procedural, and sensorimotor knowledge with the relevant role-entity bindings. In particular, it is possible to reconstruct in the mind/brain an event where John gave Mary a book in the library on Tuesday by (i) activating cortically expressed schemas and sensorimotor programs pertaining to the action *give*, (ii) activating cortical circuits embodying knowledge about John, Mary, books, the library, and Tuesday, and (iii) communicating to these schemas and circuits the bindings ⟨*giver*=`John`⟩, ⟨*recipient*=`Mary`⟩, ⟨*give-object*=`a Book`⟩, ⟨*location*=`Library`⟩, and ⟨*temporal-location*=`Tuesday`⟩.

Direct and irrefutable evidence that the mind/brain can reconstruct the gestalt and details pertaining to an event from a small number of bindings comes from the phenomena of language understanding. Consider the simple sentence "John bought a Rolls Royce." Upon hearing this sentence we can effortlessly understand the implied transaction which may involve John visiting a car showroom, selecting a car, making a payment, and obtaining ownership of the car. Additionally, we may also surmise that since Rolls Royce is an expensive car, John is likely to be a well-heeled individual. As illustrated by the above example, our mind/brain can *construct* a relatively complex event given the four word sentence "John bought a Rolls Royce," even though **the only explicit information contained in the sentence** is that (i) ⟨*buyer=John*⟩, (ii) ⟨*buy-object=a Rolls Royce*⟩, and (iii) the described event occurred in the past.

How is it that an impoverished four word "input" whose informational content is equivalent to the specification of two bindings leads to an understanding of a complex event such as someone buying a car? A plausible answer seems to be that an elaborate understanding of the event emerges when the bindings specified in the sentence tap into, and activate, the complex web of conceptual knowledge encoded in our mind/brain. This web of knowledge includes, besides other things, semantic knowledge about different sorts of entities and their attributes, and sub- and super-ordinate relationships among categories, causal knowledge about the relationship between actions and their effects, and schematized and embodied representations of generic actions. The activation of this rich web of knowledge by a sparse "input" containing only a few bindings is sufficient to produce the necessary elaboration of the buy event and trigger appropriate inferences.

The above conception of language understanding strongly resonates with the proposal that language understanding involves *embodied mental simulations* (Bailey, Chang, Feldman & Narayanan 1998; Barsalou 1999; Lakoff & Johnson 1999; MacWhinney 1999), and reflexive inferences to establish referential and causal coherence (Shastri & Ajjanagadde 1993; Shastri & Wendelken 2000).

In traditional as well as cognitive linguistics it has long been argued that events can be characterized by relational structures composed of role-entity bindings (for example, Fillmore 1968; Jackendoff 1990; Langacker 1986; Pinker 1989). Such structures have been variably referred to as frames, schemas, and scripts. This static conception of events has shortcomings that have been addressed in recent work on the modeling of action verbs using process-based representations such as X-schemas (for example, see Narayanan 1997; Bailey et al. 1998; also see Arbib 1994). This work suggests that not only static, but also dynamic aspects of events can be encoded adequately by specifying a small number of role and parameter bindings. In particular, the detailed temporal structure and dynamics of an event can be reconstructed by binding the roles and parameters of an appropriate action (or event) schema to suitable entities and values, and "executing" the schema. Such a schema execution can recreate the event's time-course, infer its consequences, and given a bodily grounding, even evoke the sensory, motoric, and somatosensory "feel" associated with the event.

It has also been shown that (i) action schemas can be realized as neurally plausible networks and integrated with neurally plausible representations of other sorts of conceptual knowledge such as beliefs, semantic facts, causal models, categories, and entities, and (ii) action schemas can "execute" via the propagation of activity over distributed neural networks (Shastri, Grannes, Narayanan & Feldman 2001). Furthermore, it has been demonstrated that inferences required for establishing causal and referential coherence across events can be drawn rapidly by instantiating a small number of bindings via patterns of activity within network structures that capture semantic knowledge and systematic causal relationships among event types (Shastri & Ajjanagadde 1993; Shastri 1999a; Shastri & Wendelken

2000).

The language understanding analogy helps explain how a sparse encoding involving a small set of bindings suffices to represent an event in episodic memory: As in the case of language understanding, a fleshed out representation of an event is reconstructed during memory recall by retrieving a small set of bindings pertaining to the event and activating the web of semantic and procedural knowledge with these bindings. What is different in the two cases is the *source* of bindings. In the case of language understanding, bindings are obtained from verbal input. In the case of remembering, they are retrieved from the event's *episodic memory trace* in the HS.[15]

The idea that memory recall involves a constructive process is very old (for example, see Bartlett 1932) and has received considerable support from psychologists (for example, see Neisser 1968; Schacter 1996a,b), and more recently, from imaging studies (for example, see Nyberg, Habib, McIntosh,& Tulving 2000; Wheeler, Petersen & Buckner 2000). But key additional insights obtained from the above analysis are as follows:

1. The *seed* underlying the reconstruction of a specific event is a small number of role-entity bindings.
2. Such role-entity bindings — together with ancillary functional circuits enumerated in Section 4.2 — are encoded in the HS during the memorization of an event. These bindings and ancillary functional circuits together constitute an event's *episodic memory trace*.
3. During recall, the episodic memory trace of an event becomes active and reinstates the bindings associated with the event within cortical circuits.
4. Upon being activated with the appropriate bindings, cortical circuits encoding action schemas and sensorimotor programs act in concert with other cortical circuits encoding generic knowledge about entities, and reconstruct the necessary gestalt and details about an event

4.2 Additional representational requirements of encoding events in episodic memory

Since role-entity bindings are critical for memorizing and reconstructing an event, the episodic memory trace of an event should include functional circuits that (i) encode role-entity bindings pertaining to the event, (ii) detect a match between the encoded bindings and those specified in a cue, and (iii) in response to a matching cue, reinstate the bindings pertaining to the event within cortical circuits. In addition to the above, the properties of episodic memory impose several other representational requirements on the episodic memory trace of an event. We consider these requirements below.

The episodic memory trace of an event must be capable of recognizing and responding positively to highly partial cues. For example, the memory trace of the event "John gave Mary a book in the library on Tuesday" should respond positively to a partial cue such as "Did John give Mary a book?".

The memory trace of an event should not match a cue that specifies an incompatible binding even if the cue contains a number of other bindings that match the memorized instance. For example, the memory trace of the event "John gave Mary a book in the library on Tuesday" should not match a cue such as "Did John give Susan a book in the library on Tuesday?" even though the latter is highly similar to the memorized event. The integrity of episodic memory depends critically on its ability to support this strong form of pattern separation.

Given a cue, we may be *reminded* of a memorized event that has an incompatible binding, but which is otherwise very similar to the cue. Being reminded of a similar, but distinct event, however, is different from erroneously matching a cue to a similar, but distinct, event. In case of such reminding, *we are explicitly aware* that there is a mismatch between the cue and the evoked memory. For example, if we have memorized the event "John gave Mary a book in the library on Tuesday" we may be reminded of this event when asked "Did John give *Susan* a book in the library on Tuesday?". But at the same time, we would also become aware that the cue specifies an incorrect recipient. Thus, reminding further highlights the strong pattern separation property of episodic memory.[16]

[15] To steal a metaphor from Harnad (1996), episodic memories, like language, allow us to (re)create an event in our mind/brain without the honest toil of perception and categorization.

[16] While an event's episodic memory trace should reject a cue containing an incompatible binding, it must be more forgiving of a cue that specifies extraneous bindings. Assume that you know "John met Tom" and are asked "Did John meet Tom on Tuesday?" Since John did meet Tom and since in the absence of any information to the contrary, it is *possible* that the meeting took place on Tuesday, a reasonable response would be: "possibly, yes" or "John did meet Tom, but I don't know if he did so on Tuesday." Thus the memory trace of an event should respond positively to a cue that contains matching bindings, even if it contains some additional bindings for roles that were left unspecified in the memorized instance.

In view of the above, the episodic memory trace of an event must respond positively to cues that are highly partial with respect to the memorized event, and at the same time, it must reject any cue that specifies an incompatible binding, even though the cue may contain a large number of bindings that match the memorized event. Taken together, these two requirements entail that the episodic memory trace of an event must be capable of detecting *binding errors* as well as *binding matches*. Note that any memory trace that can detect only binding matches cannot satisfy these requirements since it cannot distinguish between an unspecified binding and an incorrect binding. For example, a memory trace of (R_1 : ⟨*r1*=a⟩, ⟨*r2*=b⟩, ⟨*r3*=c⟩) that detects binding matches, but not binding errors, will treat an erroneous cue such as (R_1: ⟨*r1*=a⟩, ⟨*r2*=b⟩, ⟨*r3*=d⟩) on par with a partial but matching cue such as (R_1: ⟨*r1*=a⟩, ⟨*r2*=b⟩) since both cues contain the same number of matching bindings (two).

To summarize, the functional requirements discussed above suggest that the episodic memory trace of an event should incorporate neural circuits capable of:

- memorizing bindings and detecting a match between memorized bindings and those specified in a cue
- detecting a mismatch (or error) between memorized bindings and bindings specified in a cue
- detecting a match between a cue and the memorized event based on the activity of the abovementioned circuits, and communicating this match to cortical circuits
- reinstating the bindings associated with the memorized event within cortical circuits if the cue matches the event.

The above discussion identifies some of the representational properties that must be satisfied by an episodic memory trace. Any complex physical system, especially one that has been shaped by evolutionary forces, would be expected to deviate from the desired behavior in interesting ways. A model of memory should be capable of explaining such errors as collateral — though perhaps, inescapable — attributes of the system, while still explaining how the system embodies and exhibits the desirable functional behavior.

5 Cortical representations

Before describing the recruitment of an episodic memory trace, let us examine how an event (that is, a relational instance) might be expressed as a transient pattern of activity over HLCCs. Our goal here is rather limited; it is to outline an idealized and minimal description of cortical representations that suffices to illustrate how HLCCs and the HS might interact to give rise to episodic memory traces in the HS.

5.1 Encoding of generic relational schemas involves focal-clusters

Figure 3 depicts a candidate structure for expressing relational information in HLCCs (Shastri & Ajjanagadde 1993; Shastri 1999a). This depiction is highly idealized and shows only some of the essential components of the structure.

Each relational schema (or a relational frame) has an associated *focal-cluster*. Such a focal-cluster for the relation *give* is enclosed within the dotted ellipse labeled GIVE in Figure 3. For the purpose of this example, it is assumed that *give* has only three roles: *giver*, *recipient*, and *give-object*. Each role is encoded by an ensemble of cells, and these are labeled *giver*, *recip*, and *g-obj*, respectively. The focal-cluster also includes an *enabler* ensemble, ?*:give*, and two *collector* ensembles, *+:give* and *-:give*.

Note that each *label* in a focal-cluster denotes an *ensemble of cells*, and a connection from label *A* to label *B* corresponds to several connections from cells in ensemble *A* to cells in ensemble *B*. Although cell ensembles comprising a focal-cluster are grouped together and enclosed within a dotted ellipse to highlight their functional cohesiveness, their depicted proximity does not imply physical proximity. For example, cells within an ensemble would be physically dispersed within a cortical region, and cells in two different ensembles (for example, *+:give* and *?:give*) may be situated in different cortical regions or different cortical layers.

The focal-cluster associated with a relational schema acts as an anchor for encoding and attaching various kinds of knowledge about the relation. This includes motor and perceptual schemas associated with the relational schema, causal connections between this relational schema and other relational schemas, lexical and naming information, and episodic and semantic facts involving this generic relation. Information pertaining to a relation converges on its focal-cluster, and this information can be accessed by fanning out from the focal-cluster. It has been argued in Shastri & Ajjanagadde (1993) that it may be essential to associate such a focal-cluster with each relational schema in order to

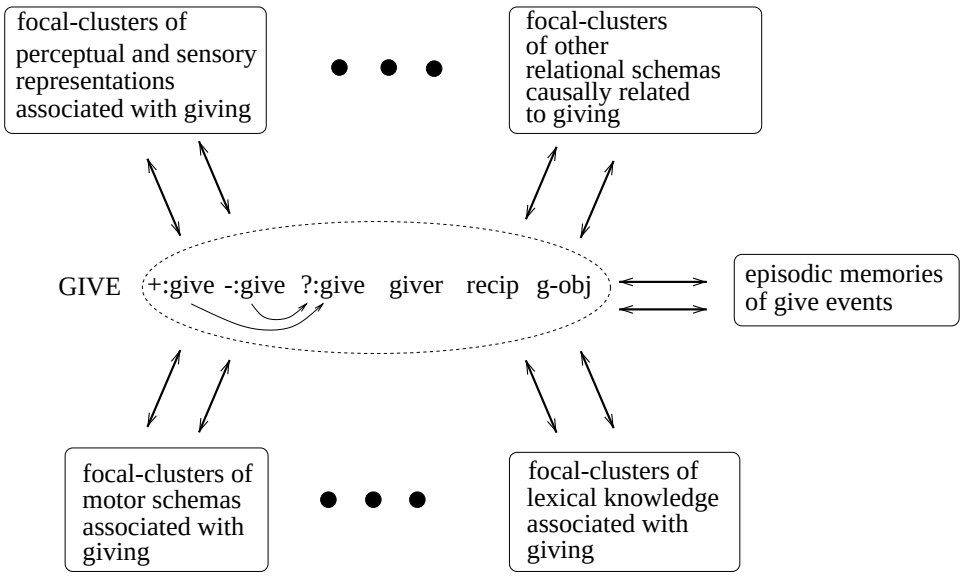

Figure 3: An idealized depiction of the focal-cluster of the relational schema *give*. The focal-cluster is enclosed within the dotted ellipse. Each label within the ellipse denotes a small but physically dispersed ensemble of cells. The relation *give* is assumed to have three roles. These are encoded by ensembles *giver*, *recip* and *g-obj*. The focal-cluster also includes an *enabler* ensemble (*?:give*) and two *collector* ensembles (*+:give* and *–:give*).

process relational information without cross-talk and at speeds required by cognitive processing. This representation of a relational schema is consistent with, but more refined than, the notion of "convergence zones" (Damasio 1989).

In what follows, we will use an ensemble label to refer collectively to cells within the ensemble. Thus "+:$give$ is active" would mean that cells in +:$give$, the collector ensemble for give, are active. Also, we will use "+:$give$ cell" to refer to an individual cell in the ensemble +:$give$.

5.2 Enabler and collector ensembles and cortico-hippocampal interactions

Assume that the roles *giver*, *recipient* and *give-object* are dynamically bound (we will see how, shortly) to John, Mary, and a book, respectively, then the activation of *?:give* means that some HLCC is "asking" the HS whether the event described by "John gave Mary a book" matches one of the events memorized by the HS. In contrast, the activation of *+:give* with the same role bindings means that some HLCC is asserting the event "John gave Mary a book". If *–:give* is active instead of *+:give*, it means that some HLCC is explicitly asserting that the event "John gave Mary a book" did *not* occur.

In response to a query about *give*, the HS activates *+:give* if the currently active instance of *give* matches one of the positive events memorized by the HS. Similarly, the HS activates *–:give* if the currently active instance of *give* matches one of the explicitly negated events memorized by the HS. If neither *+:give* nor *–:give* is activated by the HS, it means that the currently active instance of *give* does not match any event (positive or negative) memorized by the HS.

Note that a matching episodic memory trace provides *closure* between the enabler and collector ensembles of a focal-cluster, and allows activity in the enabler ensemble to propagate to the collector ensemble.

Henceforth, we will only refer to positive events, and hence, only to positive collectors of relational schemas. The treatment of negated events is analogous to that of positive instances and involves negative collectors instead of positive ones.

The significance of enabler and collector ensembles extends beyond cortico-hippocampal interactions. In general, the activation of *?:P* means that some cognitive process is seeking an explanation for (or trying to find support for) the currently active instance of *P*. The HS-based memory system is just one possible source of support; other sources being direct perception, inference, and *taxon-facts* (see Section 10.1). In contrast, the activation of *+:P* means that some cognitive process (for example, a perceptual, memory, inferential, or linguistic process) is affirming the currently active dynamic instance of *P*.

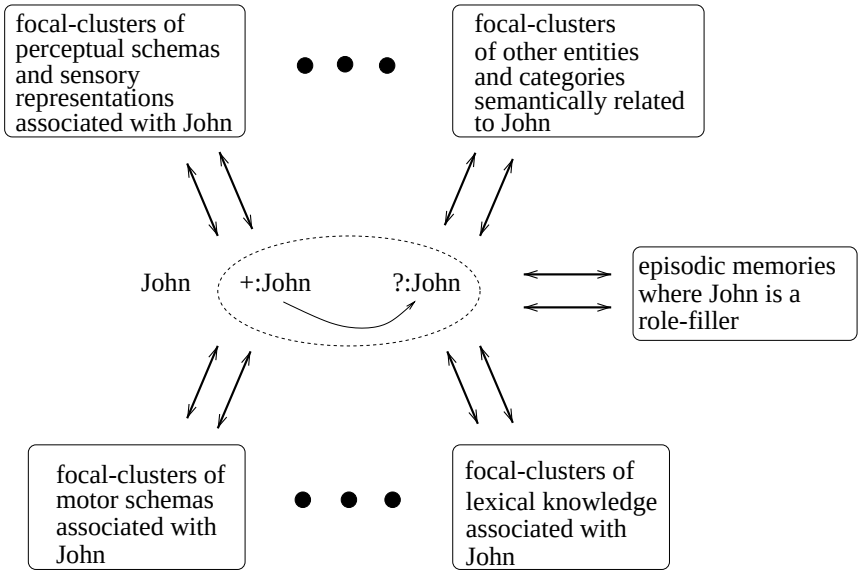

Figure 4: An idealized depiction of the focal-cluster of entity *John*. The focal-cluster is enclosed within the dotted ellipse. Each label within the ellipse denotes a small but physically dispersed ensemble of cells.

The link from the collector to the enabler ensemble of a relational schema converts a dynamic assertion about the relation into a query about this dynamic assertion. Thus HLCCs can seek an explanation of incoming knowledge in the context of existing knowledge. The collector to enabler link also creates positive feedback loops of activation. Assume that an HLCC is seeking an explanation about the currently active instance of *give*, and therefore, *?:give* is active. If the cognitive apparatus (this includes cortical and hippocampal circuits) finds support for this instance of *give* it would activate *+:give*. This would create a feedback loop — or a stable coalition — consisting of *?:give*, other ensembles participating in the explanation, *+:give*, and *?:give*.

Note that the focal-cluster associated with the *give* relational schema can be viewed as a functionally cohesive ensemble of *supra* "mirror" neurons (Gallese, Fadiga, Fogassi & Rizollati 1996; Rizzolatti & Arbib 1998) that fire whenever the mind/brain is actively representing a perception, an action, a thought, or a memory involving giving.

5.3 Focal-cluster for entities

The focal-cluster for an entity, say John, consists of an enabler ensemble *?:John* and a collector ensemble *+:John* (see Figure 4). Persistent information about various perceptual and semantic features of John, his relationship with other concepts, and the roles he fills in various events are encoded via links between the focal-cluster of John and appropriate circuits and focal-clusters representing sensory, perceptual, and semantic "knowledge" distributed across various neural structures and regions. If *?:John* is active it means that John fills a role in a query being posed by some HLCC. If *+:John* is active it means that John fills a role in an assertion being made by some HLCC. The HS activates *+:John* in response to a query involving John, if the query matches one of the memorized events.

An episodic memory trace need not always be grounded in high-level focal-clusters of an entity. It may also be grounded in focal-clusters encoding perceptual (for example, visual) features of the entity. This is consistent with the observation that the HS receives inputs from supra-modal, multi-modal, as well as high-level unimodal areas.

5.4 Dynamic (activity-based) representation of bindings

The dynamic representation of an event requires the dynamic expression of role-entity bindings. The event "John gave Mary a book" cannot be represented by simply activating the roles *giver*, *recipient*, and *give-object*, and the entities John, Mary, and a Book. Such a representation would be indistinguishable from that of "Mary gave John a Book". We assume that the brain expresses dynamic bindings by the *transient* synchronization of appropriate cells (see Ajjanagadde & Shastri 1991; Shastri & Ajjanagadde 1993; von der Malsburg 1986; Singer 1993; Hummel & Holyoak 1997). Thus the dynamic encoding of "John gave Mary a book" corresponds to the rhythmic pattern of activity shown

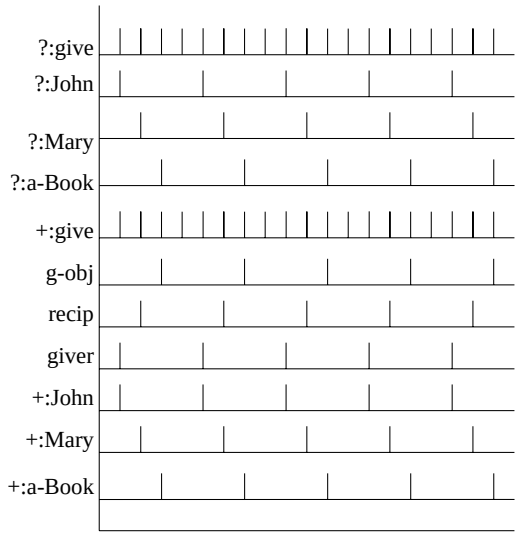

Figure 5: The rhythmic pattern of activation representing the dynamic bindings (⟨*giver = John*⟩, ⟨*recipient = Mary*⟩, ⟨*give-object = a Ball*⟩). Bindings are expressed by the synchronous activity of bound role and entity ensembles. Each spike in the illustration signifies the synchronous firing of cells in the appropriate ensemble.

in Figure 5 wherein the collectors *+:John*, *+:Mary* and *+:a-Book* are firing in distinct phases, but *+:John* and *giver* are firing in synchrony, *+:Mary* and *recip* are firing in synchrony, and *+:a-Book* and *g-obj* are firing in synchrony. Since *+:give* is also firing, the system is essentially making an assertion. As a result of the connections from collector to enabler ensembles, the enabler ensembles *?:give*, *?:John*, *?:Mary*, and *?:a-Book* will start firing soon after. The dynamic representation of the query "Did John give Mary a book" would be similar except that only the enabler ensemble would be active; the collector ensembles would remain inactive. As speculated in (Shastri & Ajjanagadde 1993), the activity of role and entity cells engaged in dynamic bindings might correspond to *gamma* band activity (ca. 30-60 Hz).

5.5 Localization of focal-clusters

It is proposed that cell assemblies making up focal-clusters of entities are located in the inferotemporal cortex[17], in particular, the perirhinal cortex. This proposal is inspired by the anatomical location of the perirhinal cortex which makes it a region of convergence of multi-modal knowledge about entities, and is consistent with growing evidence that in non-human primates and humans, the perirhinal cortex plays an important role in encoding and accessing knowledge about entities/objects (for example, Meunier, Bachevalier, Mishkin & Murray 1993; Gaffan & Parker 1996; Buckley, Gaffan & Murray 1997; Murray & Mishkin 1998; for a review see Murray & Bussey 1999; also Simons, Graham & Hodges 1999). In particular, insult to the perirhinal cortex of monkeys affects both object recognition and object-object associations across multiple modalities. Moreover, in humans, insult to the perirhinal cortex and neighboring cortical regions in the inferotemporal cortex leads to a loss of semantic knowledge about entities (semantic dementia).

It is also proposed that functional cell assemblies making up focal-clusters of relational schemas are located in (i) parahippocampal cortex which receives input from key cortical areas in the parietal and prefrontal areas, including prefrontal association areas 9 and 46 that are the regions of confluence for sensory and motor representations (Fuster 1995) and (ii) some of the cortical areas projecting directly to EC (see Section 2). Additionally, some focal-clusters of entities and relations may also be located in EC itself.

6 A model of episodic memory trace formation in the HS

This section describes how a transient encoding of an event in the form of a rhythmic pattern of activity in HLCCs can be transformed rapidly into a persistent episodic memory trace as a result of LTP within structures whose architecture

[17] This includes the temporal pole and the middle and inferior temporal gyri

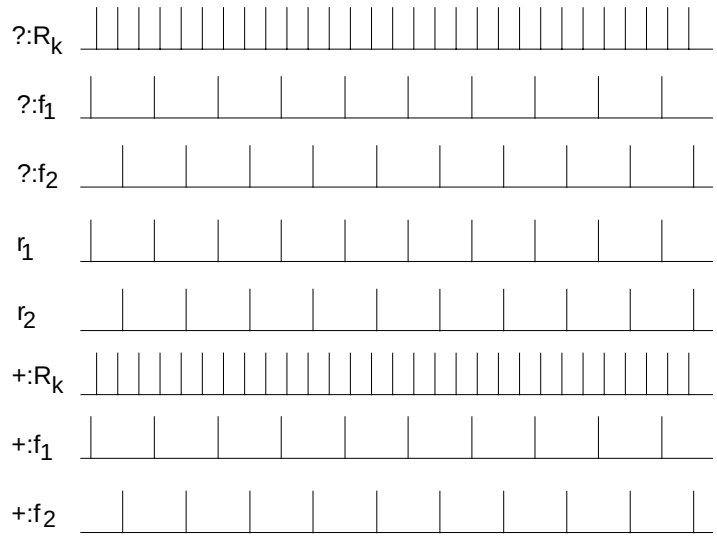

Figure 6: The transient (activity-based) representation of the event *RI* given by $(R_k : \langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$. Each spike in the illustration signifies the synchronous firing of a cell ensemble in HLCCs.

and circuitry parallel that of the HS.

6.1 Stepping through the memory acquisition process

To illustrate the process of memory acquisition let us consider the event RI given by:

$(R_k : \langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$

Typically, an event would contain several bindings, but we will work with an example involving only two bindings in order to keep the discussion and circuit diagrams simple.

The transient (activity-based) representation of event RI in HLCCs is depicted in Figure 6. As explained in Section 5.4, this activity involves the focal-clusters of relational schema R_k and entities f_1 and f_2. It is assumed that the HS is driven by both, *gamma* activity that encodes dynamic bindings between the roles and entities comprising an event, and *theta* activity that conveys the significance of an event as a whole (cf. Buzsaki & Chrobak 1995; Vertes & Kocsis 1997; Chrobak & Buzsaki 1998; Tesche & Karhu 2000).

Figure 7 depicts a simplified schematic of the persistent episodic memory trace of *RI* recruited in the model HS as a result of LTP (and LTD), when the transient pattern of activation shown in Figure 6 is presented to the model HS. Each circle in the schematic refers to one or more *cells* recruited during the memorization of RI, and each square refers to one or more *local circuits* formed during the memorization of RI. Moreover, each edge in the schematic refers to links whose synapses undergo LTP during the memorization of RI. Note that

- *Linking* cells for connecting HLCC-based focal-clusters of R_k, r_1, r_2, f_1, and f_2 to the HS are recruited in EC.
- *Binding-detector* cells (or *bind* cells) for role-entity bindings specified in *RI* are recruited in DG. After recruitment, a *bind* cell for the binding $\langle r_1 = f_1 \rangle$ will fire whenever any cue specifies this binding. *Bind* cells for $\langle r_2 = f_2 \rangle$ will behave in an analogous manner.
- Circuits that serve as *binding-error-detectors* (or *bed* circuits) for the bindings in *RI* are recruited in CA3. After recruitment, a *bed* circuit for the binding $\langle r_1 = f_1 \rangle$ will fire whenever a cue specifies a binding $\langle r_1 = f_j \rangle$ such that f_j is not the same as f_1. *Bed* circuits for $\langle r_2 = f_2 \rangle$ will behave in an analogous manner.
- Cells that integrate the outputs of *bed* circuits pertaining to *RI* are recruited in CA2. After recruitment, such *binding-error-integrator* cells (or *bei* cells) will fire whenever a cue specifies an erroneous binding with respect to *RI*.

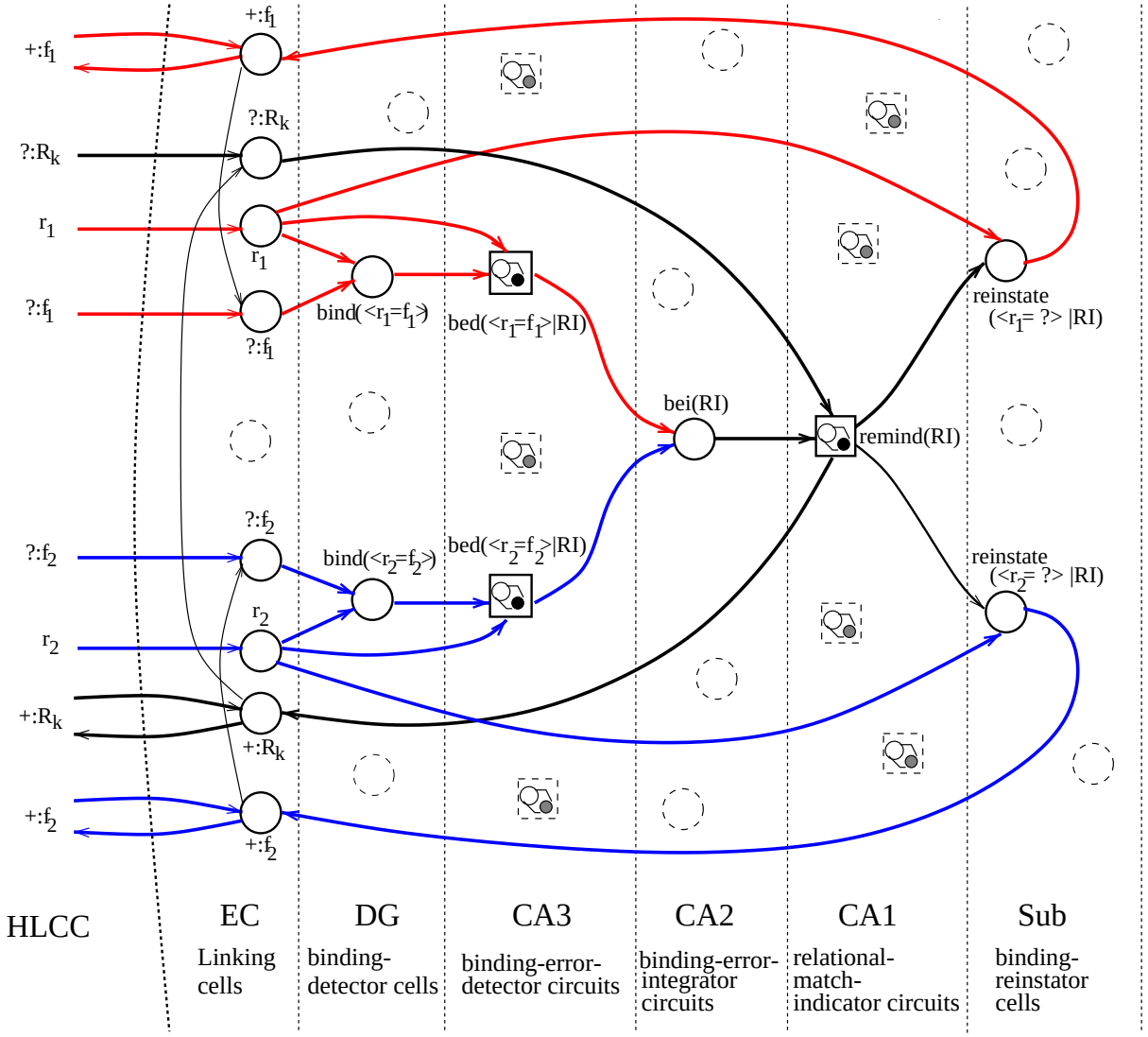

Figure 7: A simplified schematic of the distributed circuit recruited in SMRITI during the memorization of the event *RI* given by $(R_k : \langle r_1 = f_1 \rangle, \langle r_2 = f_2 \rangle)$. These circuits are formed as a result of LTP (and LTD). See text for details.

- CA1 provides the locus for the formation of *relational-match-indicator* circuits (or *remind* circuits). for *RI*. After recruitment, these *remind* circuits will fire in response to a cue whenever *none* of the bindings in the cue are erroneous with respect to *RI*. Thus the firing of *remind* circuits for *RI* will indicate that the cue matches *RI*.
- *Remind* circuits recruited for *RI* are linked to +:R_k linking cells in EC. Thereafter, the firing of these *remind* circuits will lead to the firing +:R_k linking cells in EC, and hence, to the firing of +:R_k cells in HLCCs.
- *Binding-reinstator* cells (or *reinstate* cells) are recruited in the subiculum. In response to a matching cue, a recruited *reinstate* cell for the binding $\langle r_1 = f_1 \rangle$ of RI will fire in phase with r_1 cells. *Reinstate* cells for $\langle r_2 = f_2 \rangle$ will behave in an analogous manner.
- *Reinstate* cells recruited for the bindings $\langle r_1 = f_1 \rangle$ and $\langle r_2 = f_2 \rangle$ of RI are linked to +:f_1 and +:f_2 linking cells, respectively, in EC. Thereafter, the firing of these *reinstate* cells will lead to the firing of +:f_1 and +:f_2 linking cells in EC, and hence, to the firing of +:f_1 +:f_2 cells in HLCCs.

Several "copies" of each functional cell and circuit mentioned above are recruited during the memorization of an event. Furthermore, the multiple copies recruited for each functional unit are physically dispersed. The redundancy and physical dispersion are crucial for the proper functioning of the episodic memory system, and for its robustness in the face of diffuse cell loss.

The following sections describe the recruitment of (1) *linking* cells in EC, (2) *binding-detector* cells in DG, and (3) *binding-error-detector* cells in CA3. Due to limited space, the recruitment of *binding-error-integrator* cells in CA2, *relational-match-indicator* circuits in CA1, *binding-reinstator* cells in the subiculum, and the synaptic modifications associated with the linking of *relational-match-indicator* circuits and *binding-reinstator* cells to EC are not described. A complete circuit-level description of the recruitment of an episodic memory trace appears in (Shastri 2001b).

The process described below by which cells and circuits get recruited to form functional units is susceptible to several problems. For example, in each target region there should exist cells that receive afferents from appropriate cells situated in regions upstream from the target region. The existence of such cells cannot be guaranteed. Furthermore, a cell in a target region may get recruited as part of multiple function units. Excessive sharing of cells can lead to cross-talk and interference during encoding and retrieval. However, results of the statistical analysis presented in Section 8 will show that for a plausible choice of system parameters, the probability of not finding any candidates for recruitment is extremely small, and the impact of spurious activity resulting from shared cells and ill-formed functional units is minimal.

6.2 EC: Linking the model HS to relations, entities and roles in HLCC

The model EC contains two cell types: principal cells and Type-1 inhibitory interneurons. The latter form local inhibitory circuits with principal cells and thereby regulate the extent of excitatory activity and the induction of LTP in these cells. The upper layers of model EC are divided into three distinct regions and the deeper layers into two[18] (see Figure 8). The regions in the upper layers are referred to as ECee, ECer, and ECro and those in the deeper layers as EEce and EEcr. The name of a region in EC reflects the function performed by the HLCC cells that get linked to cells in this region. For example, ECer refers to the region whose cells get linked to **e**nablers of **r**elational schemas in HLCCs, ECcr refers to the region whose cells get linked to **c**ollectors of **r**elational schemas in HLCCs, and ECro refers to the region whose cells get linked to **ro**les of relational schemas in HLCCs.

Enabler cells of entities and relational schemas located in HLCCs project to the regions ECee and ECer, respectively. Role cells in HLCCs project to the region ECro. Collector cells of entities and relational schemas located in HLCCs project to the regions ECce and ECcr, respectively. Moreover, cells in ECee and ECer also receive afferents from cells in ECce and ECcr, respectively.

The presentation of the event *RI* by HLCCs to EC leads to the following sequence of events in EC (refer to Figures 7 and 8).

Linking of cells in ECcr, ECce and ECro to HLCCs: Some cells in ECcr become linked to +:R_k cells. At the same time, some cells in ECce become linked to +:f_1 cells and some to +:f_2 cells. Furthermore, some cells in ECro

[18] It is known that EC is divided into several distinct regions (see Insausti et. al 1995). For ease of modeling we are assuming that enablers for all relational schemas are grouped together in one region of EC. Similarly, we are assuming that enablers of all entities are grouped together in a single region of EC. But it is quite possible that different types of relations, for example, spatial relations and social relations may occupy different regions. Similarly, it is possible that the enablers of different types of entities such as persons and tools, may occupy different regions.

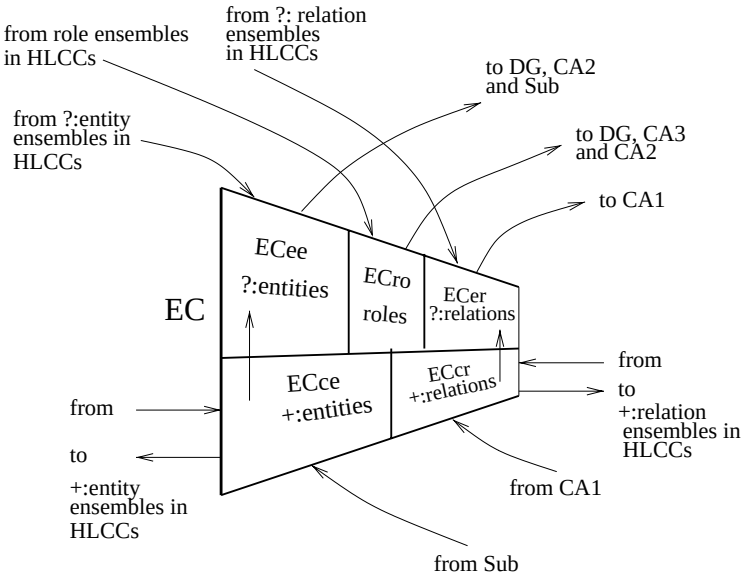

Figure 8: A schematic illustration of the functional division of the model EC into multiple regions and the projections between these regions and the rest of the model HS. Links within regions of the model EC are also shown.

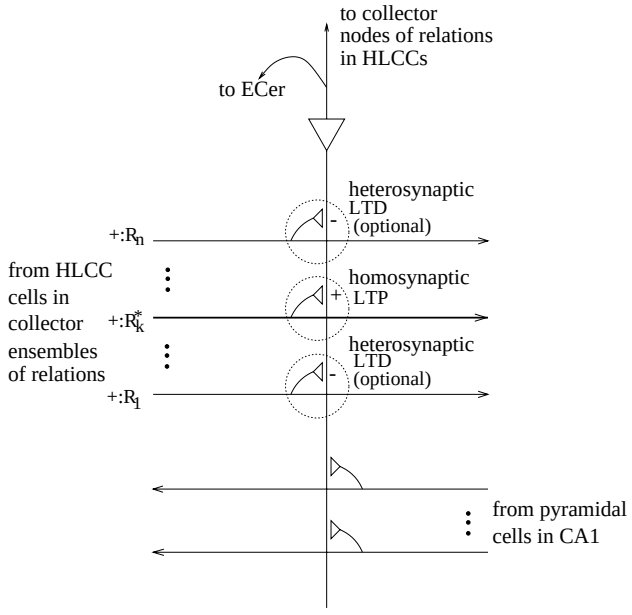

Figure 9: Linking of a cell in ECcr to $+:R_k$, the collector of relational schema R_k in a HLCC.

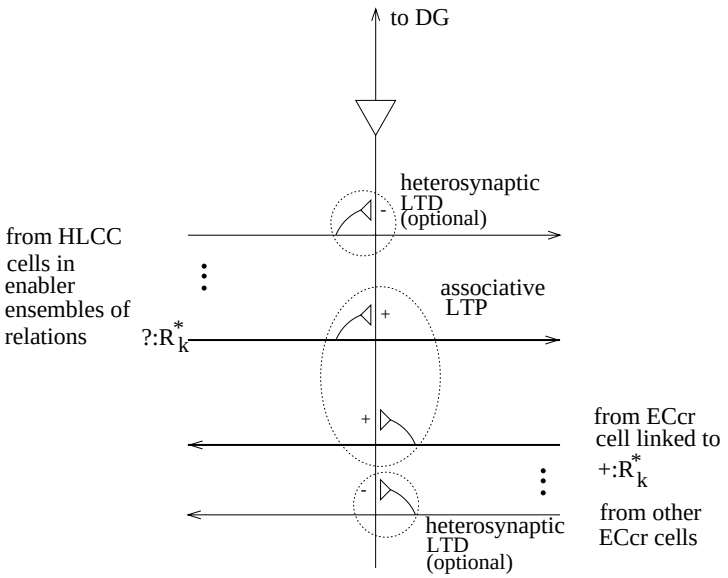

Figure 10: Linking of an ECer cell to ?:R_k, the enabler of relational schema R_k in HLCCs.

become linked to r_1 cells and some to r_2 cells. The linking of cells in ECcr is the result of homosynaptic LTP of synapses receiving afferents from +:R_k cells, and optionally, the heterosynaptic LTD of synapses formed on these cells by afferents from inactive collector cells of relational schemas other than +:R_k (refer to Figure 9). The linking of cells in ECce and ECro occurs in an analogous manner as a result of impulses arriving at synapses formed on ECce cells by afferents from +:f_1 and +:f_2 cells, and impulses arriving at synapses formed on ECro cells by afferents from +:r_1 and +:r_2 cells, respectively.

Potentiation of backprojections from ECcr and ECce cells to HLCCs: Once ECce cells get linked to +:f_1 and +:f_2 cells, and ECcr cells get linked to +:R_k cells, these linked cells start firing in synchrony with the respective collector ensembles to which they are linked. This leads to the potentiation of synapses formed by efferents emanating from these linked cells and impinging on collector cells in HLCCs to which they are linked.

Linking of ECer and ECee cells to HLCCs: As a result of the firing of ECcr cells linked to +:R_k, some cells in ECer receive convergent activity from these linked ECcr cells and ?:R_k cells. Consequently, active synapses of such ECer cells undergo associative LTP. As a result, these cells get linked to ?:R_k cells and to ECcr cells linked to +:R_k cells (refer to Figure 10).[19] The optional heterosynaptic LTD of some of the inactive synapses of these ECer cells further increases the selectivity of these cells. In a similar manner, some ECee cells get linked to ?:f_1 cells and to ECce cells linked to +:f_1 cells, and some other ECee cells get linked to ?:f_2 cells and to ECce cells linked to +:f_2 cells.

To summarize, the above neural events lead to the recruitment within EC of small ensembles of cells linked to each of the following HLCC ensembles: +:R_k, ?:R_k, r_1, r_2, +:f_1, +:f_2, ?:f_1 and ?:f_2. Furthermore, cells in the collector ensembles recruited in ECcr for +:R_k, and in ECce for +:f_1 and +:f_2, connect back to the respective collector ensembles in HLCCs via potentiated links. The recruitment of linked cells occurs the first time a relational schema, entity, or role participates in an event presented to the HS.

> For convenience we will refer to an ensemble of linked cells in EC by the name of the HLCC ensemble to which it is linked. For example, we will refer to the cell ensemble in ECee that is linked to the ensemble ?:f_1 in HLCCs as ?:f_1. Also, we will refer to a cell within this ensemble as a ?:f_1 cell.

6.3 DG: The recruitment of binding-detector cells

Model DG contains two kinds of cells: principal cells and Type-1 inhibitory interneurons. Principal cells receive afferents from cells in ECee and ECro regions. In turn, principal cells make synaptic contacts with Type-1 interneurons

[19] The key synaptic modification at ECer cells required for the proper functioning of the model is the associative LTP of synapses receiving afferents from ?:R_k in the presence of concurrent activity arriving from ECcr cells linked to +:R_k. The LTP of synapses receiving afferents from ECcr cells linked to +:R_k is not critical.

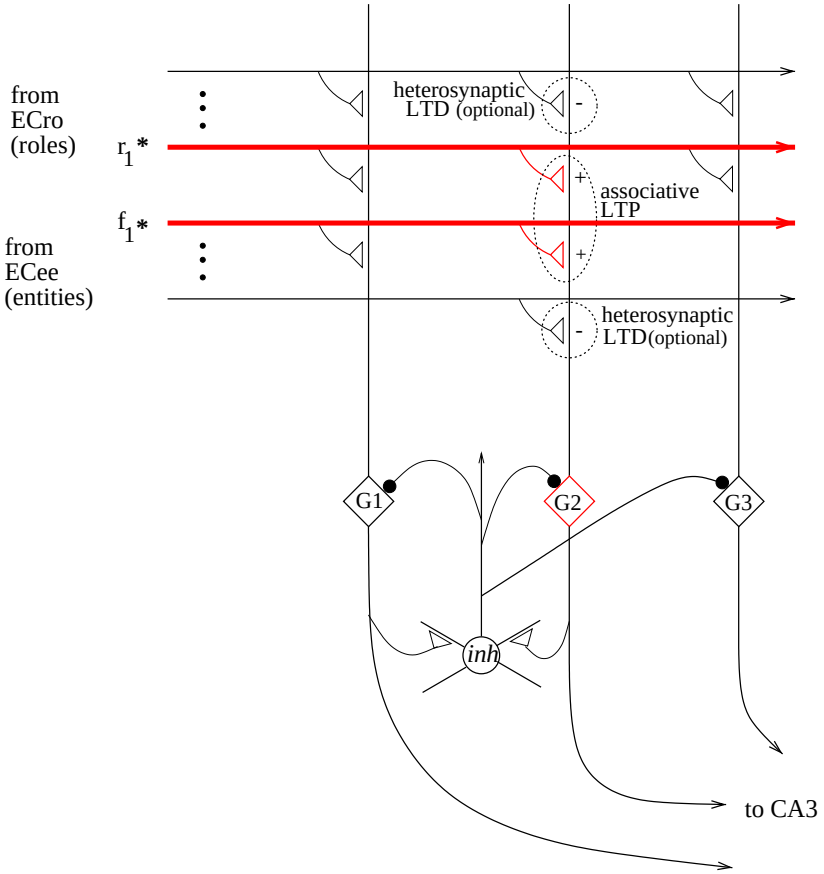

Figure 11: Internal structure of model DG. The region consists of principal cells and inhibitory interneurons (Type-1) and receives dense and diffuse afferents from ECro and ECee regions. Principal cells and Type-1 interneurons form feedback and feedforward inhibitory circuits that limit the number of cells whose synapses undergo LTP. G1—G3 are principal cells and inh is a Type-1 interneuron. Afferents labeled r_1* and f_1* are from cells in the ensembles for role r_1 and entity f_1, respectively. Synapses of G1 and G2 receive sufficient synchronous activity along afferents from r_1 and f_1 cells, and hence, are *candidates* for becoming *binding-detector* cells for the binding $\langle r_1 = f_1 \rangle$. While the inhibition from inh prevents the LTP of G1's synapses, G2's synapses undergo LTP, and hence, G2 gets recruited as a *binding-detector* cell for $\langle r_1 = f_1 \rangle$. Synapses undergoing LTP are marked with a '+', those undergoing LTD are marked with a '-'.

and project to model CA3. Type-1 interneurons form inhibitory synapses with principal cells giving rise to feedback inhibitory circuits in model DG (see Figure 11).

The weight of naive synapses formed by afferents from ECro and ECee is assumed to be 100, and the weight of potentiated synapses is assumed to be 200. The potentiation and firing thresholds (θ_p and θ_f) of principal cells are assumed to be 850 and 1700, respectively.[20] Since θ_p of principal cells is sufficiently high, LTP of a synapse occurs only if multiple synapses of the postsynaptic cell receive coincident presynaptic activity. Moreover, since θ_f of principal cells is sufficiently high, a principal cell does not fire unless it receives impulses at multiple potentiated synapses.

Subsequent to the linking of cells in EC, the presentation of RI (see Figure 6) leads to the following events in DG (refer to Figures 11 and 7). As a result of the synchronous firing of r_1 and ?:f_1 cells in ECro and ECee, respectively, certain DG principal cells receive sufficient synchronous inputs and their active synapses undergo associative LTP. At the same time, some of the inactive naive synapses of these principal cells (optionally) undergo heterogeneous LTD. We will refer to such cells as $bind(\langle r_1 = f_1 \rangle)$ cells.

Since θ_f of principal cells is such that a cell does not fire unless it receives impulses at multiple potentiated synapses, impulses arriving at numerous naive synapses do not lead to the firing of a $bind(\langle r_1 = f_1 \rangle)$ cell. Even activity arriving at a few potentiated synapses does not lead to the firing of such a cell. Only the coincident arrival of impulses at many potentiated synapses (for example, from r_1 *and* ?:f_1 cells) satisfies θ_f and causes such a cell to fire in close temporal proximity of presynaptic activity. Thus most $bind(\langle r_1 = f_1 \rangle)$ cells fire when r_1 cells in ECro fire in synchrony with ?:f_1 cells in ECee, and hence, behave as *binding-detector* cells for the role-entity binding $\langle r_1 = f_1 \rangle$.

Similar LTP and LTD events occur at the synapses of principal cells that receive coincident activity along afferents from r_2 cells in ECro and ?:f_2 cells in ECee, and lead to their recruitment as $bind(\langle r_2 = f_2 \rangle)$ cells.

Local inhibitory circuits involving Type-1 interneurons serve as soft-winner-take-all networks (soft-WTA) and allow synapses of only a limited number of cells to undergo LTP (cf. Marr 1971; McNaughton & Morris 1987) (Section 8.1.3). Nevertheless, numerous DG principal cells are recruited as *bind* cells for each binding (see Section 8.2) and a vast majority of these behave in the desired manner (see Section 8.3).

6.4 CA3: The recruitment of binding-error-detector circuits

Given their unusual behavior, the recruitment of functional units responsive to binding-errors is more complex than the recruitment of *binding-detector* cells. Consider a *binding-error-detector* for the binding $\langle r_1 = f_1 \rangle$. This functional unit must be recruited in response to the concurrent activation of r_1 and f_1, but subsequent to its recruitment, it must *not* fire anymore in response to the concurrent activity of r_1 and f_1 — *the very activity that led to its formation*. This section explains how such circuits can get recruited in a model region whose local circuitry and afferent connections are similar to those of CA3.

Model CA3 contains principal cells and two types of inhibitory interneurons — Type-1 and Type-2. The principal cells and interneurons form two kinds of local circuits. The first kind of circuits involve Type-1 interneurons and perform the same function as that performed by Type-1 local inhibitory circuits in DG; they limit the number of principal cells whose synapses undergo LTP. The second kind of circuits involve Type-2 interneurons and lead to the recruitment of *binding-error-detector* circuits.

Each principal cell in CA3 receives afferents from a number of cells in ECro and DG, and sends collaterals to neighboring Type-2 interneurons. Type-2 interneurons in turn make contacts on neighboring principal cells. If a principal cell receives an inhibitory contact from a Type-2 interneuron, then the likelihood that the principal cell also sends a collateral back to the same interneuron is high. Consequently, there exist a large number of feedback circuits consisting of a principal cell and a Type-2 interneuron. One such feedback circuit consisting of principal cell P and Type-2 interneuron int is depicted in Figure 12(a). As a matter of convention, we will refer to int as a *satellite* of P. Typically, each principal cell will have several satellites and each Type-2 interneuron will be a satellite of numerous principal cells.

The projection from DG to CA3 is such that given a principal cell P and one of its satellite int, if P receives an afferent from a cell b in DG, then it is likely that int also receives an afferent from b. This sort of connectivity could arise naturally since DG granule cells make numerous contacts on inhibitory interneurons in CA3 (Acsady et al. 1998).

The potentiation threshold, θ_p, and weight parameters of principal cells and Type-2 interneurons are such that LTP of synapses of a principal cell occurs only if multiple synapses of the cell receive coincident presynaptic activity.

[20] These and other threshold values may be interpreted to be the effective threshold values near positive peaks of the *theta* cycle.

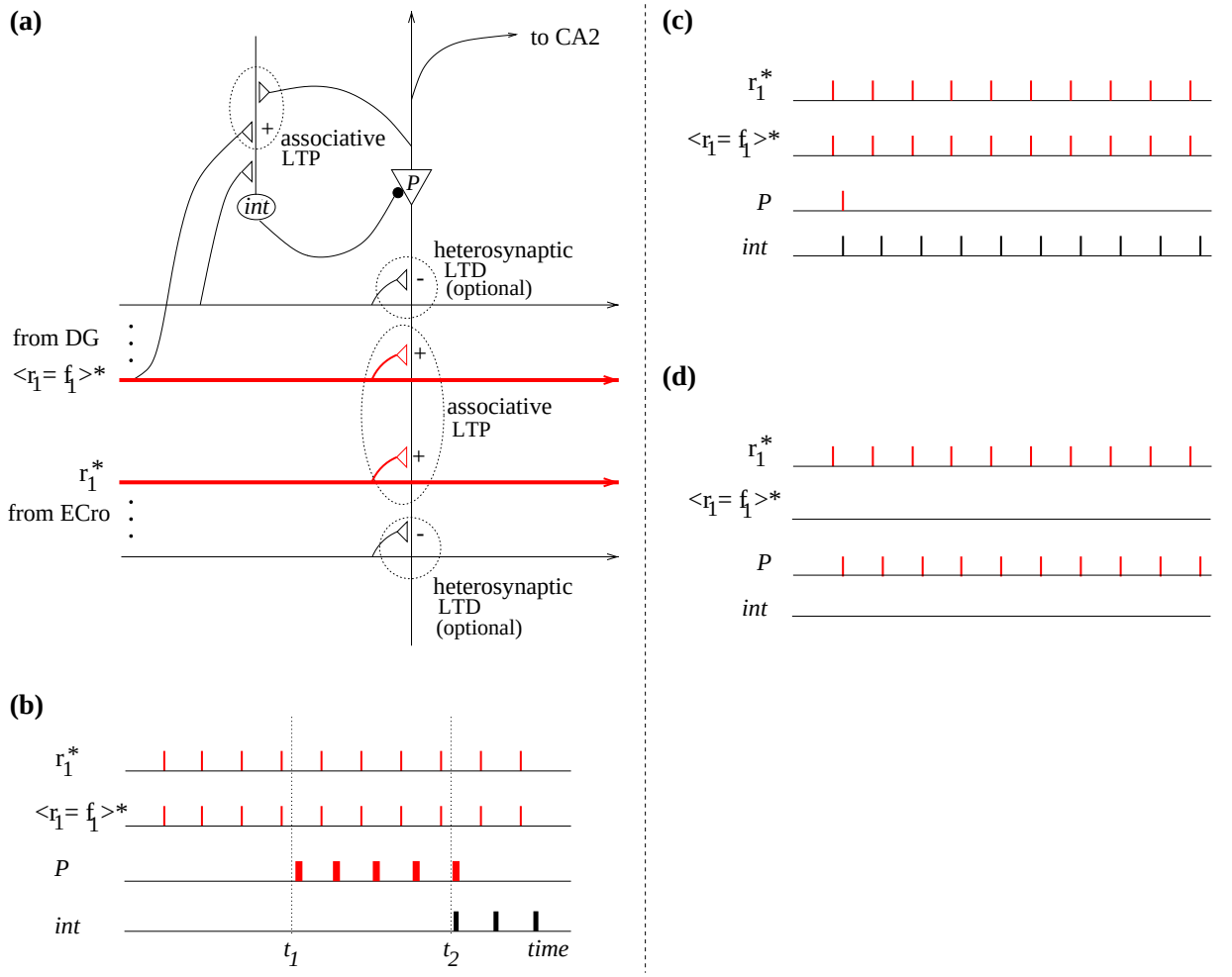

Figure 12: (a) A local inhibitory feedback circuit in CA3 consisting of a principal cell P and a Type-2 inhibitory interneuron int. Type-1 interneurons are not shown. The input labeled r_1* refers to afferents from r_1 cells in ECro, and that labeled $bind(\langle r_1 = f_1 \rangle)$* refers to afferents from $bind(\langle r_1 = f_1 \rangle)$ cells in DG. The arrival of coincident activity along r_1* and $bind(\langle r_1 = f_1 \rangle)$* causes LTP, and optionally, LTD of synapses, and this results in the recruitment of a *binding-error-detector* circuit consisting of P and int, for the binding $\langle r_1 = f_1 \rangle$ in RI (see text for details). Synapses undergoing LTP are marked with a '+' and those undergoing LTD are marked with a '-'. (b) A schematic representation of the activity of P and int during the above process. The LTP and LTD of P's synapses has occurred by time t_1, and the LTP of int's synapse has occurred by time t_2. The supra-active response of P upon receiving simultaneous inputs at potentiated synapses from r_1* and $bind(\langle r_1 = f_1 \rangle)$* is shown as a thick line. The response of P and int, subsequent to their recruitment as a *binding-error-detector* circuit, to a cue specifying the binding $\langle r_1 = f_1 \rangle$ is shown in (c) and to a cue specifying a binding of r_1 with an entity other than f_1 is shown in (d). The sustained firing of P signals a binding-error.

Cell	θ_p	θ_f	θ_{sf}
CA3 (princ)	1050	850	3200
CA3 (Type-2 int)	2600	2300	n.a.

Table 1: Potentiation and firing thresholds of principal cells and Type-2 interneurons in model CA3.

Similarly, LTP of synapses formed by afferents from DG cells on Type-2 interneurons occurs only if the interneuron receives coincident activity at multiple synapses. The synapses formed on Type-2 interneurons by collaterals from principal cells are not required to undergo LTP.

Principal cells in CA3 have two activity (or firing) modes: *normal* and *supra-active*. These modes are associated with different firing thresholds, namely, θ_f and θ_{sf}, and different firing levels. An inhibitory input from a Type-2 interneuron is sufficient to block a principal cell from firing — irrespective of any excitatory inputs received by the latter. Finally, θ_f of Type-2 interneurons is such that it is unlikely to fire unless it receives impulses at potentiated synapses from DG cells. A set of plausible values for θ_p, θ_f, and θ_{sf}, and synaptic weights of naive and potentiated synapses are given in Tables 1 and 2.

The recruitment of a binding-error-detector circuit involves the following steps (refer to Figure 12(a)).

1. The principal cell P in Figure 12(a) receives afferents from r_1 cells in ECro, and a $bind(\langle r_1 = f_1 \rangle)$ cell in DG. Given the dynamic encoding of RI (Figure 6), these afferents convey synchronous activity and this leads to the associative LTP of P's synapses receiving afferents from $bind(\langle r_1 = f_1 \rangle)$ and r_1 cells.[21] At this time, some of the inactive naive synapses of P receiving afferents from DG cells and ECro cells may (optionally) undergo heterosynaptic LTD.

2. The firing thresholds of CA3 principal cells are such that a cell receiving synchronous impulses at potentiated synapses from adequate number of ECro and DG cells fires in the *supra-active* mode. Hence, subsequent volleys of inputs from $bind(\langle r_1 = f_1 \rangle)$ and r_1 cells cause P to fire in the supra-active mode (Figure 12(b)).

3. The arrival of (supra-level) activity from P leads to the associative LTP of the synapse at which int is receiving concurrent activity from a $bind(\langle r_1 = f_1 \rangle)$ cell. After the potentiation of this synapse, activation arriving from the $bind(\langle r_1 = f_1 \rangle)$ cell is sufficient to fire int and cause the inhibition of P.

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 $\langle r_1 = f_1 \rangle$. We will refer to this circuit as a $bed(\langle r_1 = f_1 \rangle | RI)$ circuit and to P as a $bed(\langle r_1 = f_1 \rangle | RI)$ cell. During retrieval, P will *not* fire if role r_1 is bound to entity f_1 in the cue presented to EC because the synchronous firing of r_1 and f_1 cells will activate $bind(\langle r_1 = f_1 \rangle)$ cells in DG, which will activate int, which will in turn inhibit P (Figure 12(c)). P however, *will* fire in the normal mode whenever the firing of r_1 cells is *not* accompanied by the synchronous firing of $bind(\langle r_1 = f_1 \rangle)$ cells. In other words, P will fire whenever the cue currently active in EC binds r_1 to any entity other than f_1 (Figure 12(d)). While the same set of $bind(\langle r_1 = f_1 \rangle)$ cells is shared among episodic memory traces of different events containing the binding $\langle r_1 = f_1 \rangle$, a distinct set of bed circuits is formed for the binding $\langle r_1 = f_1 \rangle$ *each* time an event containing this binding is memorized. Hence, the names of *bed* circuits and cells are qualified by the name of the event that lead to their recruitment.

A process similar to the one described above leads to the recruitment of $bed(\langle r_2 = f_2 \rangle | RI)$ circuits that act as *bed* circuits for the binding $\langle r_2 = f_2 \rangle$.

Several *bed* circuits are recruited for each binding in the event being memorized (see Section 8.2), and a vast majority of these behave as desired (see Section 8.3).

6.4.1 Functional significance of projections from CA3 and CA2 to DG

Bed circuits in CA3 are a critical resource since a new set of *bed* circuits must be recruited for a binding *each time* an event containing this binding is memorized. Type-1 inhibitory interneurons conserve this limited resource by allowing only a small number of candidate *bed* circuits to be recruited for each occurrence of a binding. SMRITI posits that in

[21] The associative LTP of mossy fiber → CA3 synapses in response to persistent activity along mossy fibers coupled with activity along EC → CA3 fibers, is analogous to the associative LTP of mossy fiber → CA3 synapses reported in (Derrick & Martinez 1996).

Projection	naive weights	potentiated weights
DG → CA3 (princ)	800	2400
ECro → CA3 (princ)	100	300
DG → CA3 (inh)	800	2400
CA3 (princ) → CA3 (inh)	500	n.a.

Table 2: Naive and potentiated weights of synapses in model CA3. "Princ" refers to principal cells and "inh" refers to Type-2 interneurons

.

addition to this gross control exerted by Type-1 interneurons, the projections from CA3 and CA2 to DG (Section 2.1) can provide an effective feedback control mechanism for regulating the allocation of this limited resource.

As the number of memorized events pertaining to a particular relational schema, say *give*, increases, the pool of candidate cells remaining in CA3 that can serve in *bed* circuits of subsequent *give* events gets depleted. But at the same time, the number of *bed* and *bei* circuits that fire in response to a novel *give* event increases (since all previously memorized *give* events mismatch a novel *give* event). Given the feedback connections from CA3 and CA2 to DG, this increase in the level of CA3 and CA2 activity effectively lowers the potentiation threshold of DG granule cells (see Section 2.1), and thereby, increases the number of DG granule cells recruited as *bind* cells for bindings pertaining to *give* events. The increase in the number of recruited *bind* cells directly enlarges the pool of cells in CA3 that are candidates for recruitment as *bed* circuits for bindings pertaining to additional *give* events.

6.5 Mass of an episodic memory trace

Let us refer to the number of cells involved in an event's episodic memory trace as its neural mass. *The greater the neural mass of a memory trace, the greater its redundancy, and hence, the greater its robustness.*

The neural mass recruited for an episodic memory trace depends on at least six factors: (1) architectural parameters such as region and projective field (PF) sizes[22], (2) physiological parameters governing LTP and LTD, and potentiation and firing thresholds, (3) the neural mass of HLCC-based focal-clusters of the relational schema, entities, and roles pertaining to the event, (4) the significance of the event being coded, (5) the degree of repetition, and (6) the degree to which the HS is already loaded with other memories at the time of memorization. While the first two factors are independent of the event being memorized, the last four can vary from one event to another.

The numbers of cells in the HLCC-based ensembles of relational schemas, roles, and entities directly impact the number of cells and circuits recruited for an event's memory trace; the greater the HLCC-based ensemble size, the greater the expected number of cells in the HS recruited as functional units, and hence, the greater the expected neural mass of the episodic memory trace. Consequently, *all else being equal*, it would be easier to memorize an event involving entities with high cortical mass (that is, familiar and well-known entities) than it would be to memorize an event involving entities with low cortical mass (that is, unfamiliar entities).

The greater the strength of the significance signal impinging on the HS, the greater the likelihood that a cell will be recruited. Consequently, all else being equal, the greater the significance of an event, the greater the mass of its memory trace.[23]

The complete recruitment of an event's memory trace requires a certain minimum number of cycles (see Section 6.6 below). Upon the completion of recruitment, appropriate linking cells in ECcr and ECce become active, and convey this activity back to collector cells in HLCCs. These HLCCs would then cease "presenting" the event to the HS (this would be akin to an attentional shift mediated, perhaps, by prefrontal circuits). If however, the HLCCs continue to present the event to the HS (for example, if the event draws extended attention, or is rehearsed), another round of recruitment would occur and additional functional circuits would get recruited to encode the same event. Thus, all else being equal, rehearsal and sustained focus of attention will lead to a heavier memory trace.

[22] For a given projection, the collection of cells in the target region receiving afferents from a cell c in the source region is referred to as the *projective field* (PF) of c. The *size* of a PF refers to the number of cells in the PF.

[23] Although, the effect of significance is presumably realized in biological systems via neuromodulators, we model its effect computationally in the following simple manner: The significance signal acts as a region-wide bias signal that effectively lowers the potentiation (and firing) thresholds, and partially offsets the effect of Type-1 interneurons. Hence, greater the significance level, greater the likelihood that a cell's synapse will undergo LTP and lead to its recruitment.

Finally, as the number of events retained within the HS-based memory increases, more and more inhibitory interneurons get incorporated into bed and $remind$ circuits. This gradually increases the number of inhibitory interneurons firing during memorization and retrieval, and hence, the number of candidate cells blocked during memorization and retrieval. Additionally, the number of synapses undergoing LTD also increase with memory load. Consequently, all else being equal, an event memorized by a heavily loaded HS will have a smaller mass compared to an event memorized by a lightly loaded HS. We will see how memory load affects recruitment and response properties of episodic memory traces in Sections 8.2 and 8.3.

6.6 Encoding and retrieval times

SMRITI takes 20 cycles (*gamma* cycles) to recruit the memory trace of an event (in other words, the cortical activity representing an event's bindings must persist for 20 cycles). Since the period of a *gamma* cycle is about 25 msec., the model explains how the episodic memory trace of an event can be recruited in about one second. SMRITI takes eight cycles (about 200 msec.) to respond yes/no to a query (that is, to activate +:*give* in response to the query "Did John give Mary a book?") and ten cycles (about 250 msec.) to reinstate the bindings of a matching event (that is, to activate +:John in response to "Who gave Mary a book?"). Note that the retrieval process in SMRITI corresponds to a parallel search wherein the query (cue) is matched simultaneously against all the memory traces encoded in the HS. This time course of memory acquisition and retrieval is consistent with the findings reported in (Fernandez et al. 1999).

6.7 Learning sequence of events

If principal cells A and B in CA1 fire in the burst mode at times t_0 and t_1, respectively, where t_1 occurs soon after t_0, then synapses formed by afferents from A to B would undergo LTP. This will have the following consequence: If an event $E2$ occurs soon after an event $E1$, then the cluster of *remind* cells recruited for memorizing $E1$ would tend to get linked to the cluster of *remind* cells recruited for memorizing $E2$. As a result of this synaptic potentiation, any future retrieval of $E1$ would lead to the activation of $E2$. This would make it possible to link together sequences of events in episodic memory. In particular, this would make it easier to remember stories consisting of a sequence of causally coherent events.

7 Episodic memory consolidation and forgetting

The sequence of synaptic changes described in Section 6 specifies how the episodic memory trace of an event is rapidly *acquired* by the model HS. Biological and psychological data suggests that the rapid acquisition of a memory trace in the HS is followed by a much slower *consolidation* process that lasts hours, days, or perhaps, even weeks and makes the memory trace less prone to disruption and forgetting (Shimizu, Tang, Rampon & Tsien 2000; Wickelgren 1977). But what is the nature of episodic memory consolidation?

A prevailing view is that episodic memory consolidation involves a recoding process that transfers the episodic memory trace of an event from the HS to the neocortex (Marr 1971; Squire 1992; Murre 1996; Bontempi, et al. 1999; McClelland, McNaughton & O'Reilly 1995). Under this view of consolidation, the HS serves as a temporary buffer for new and recently acquired memory traces. Over time, these memory traces are transferred from the HS to cortical circuits where they reside as long-term (permanent) memories. Although this view of consolidation is widely held in the neuropsychological literature, it has been challenged by other researchers on the grounds that it does not offer a satisfactory account of experimental data (for example, Nadel & Moscovitch 1997; Murray & Bussey 2001; also see Wickelgren 1977).

A key problem with the "consolidation as transfer" hypothesis is the finding that retrograde amnesia in hippocampal patients often extends to events occurring several decades prior to the insult to their HS (see Kartsounis, Rudge & Stevens 1995; Rempel-Clower, Zola, Squire & Amaral 1996; Nadel & Moscovitch 1997; Stefanacci et al. 2000). Under the "consolidation as transfer" hypothesis, the temporal extent of retrograde amnesia in hippocampal patients should be no more than the time it takes to transfer an episodic memory trace from the HS to the neocortex. Hence, under the "consolidation as transfer" hypothesis, the temporal extent of retrograde amnesia in hippocampal patients implies that the transfer of an episodic memory trace from the HS to the neocortex occurs over several decades. However, such a long-lasting consolidation/transfer process seems unmotivated and unnecessary on computational as well as biological grounds.

Based on computational and architectural considerations SMRITI predicts that the error-sensitive episodic memory trace of an event must continue to be encoded in the HS for as long as the event is remembered as a specific episode situated in a spatio-temporal context. This prediction follows from the representational requirements of an episodic memory trace (cf. Section 4.2) and the match between the specialized neural circuit required to support these representational requirements on the one hand, and the idiosyncratic architecture of the HS on the other (cf. Section 6). Since the episodic memory trace of a memorized event must remain dependent on the HS, SMRITI predicts that memory consolidation does not involve a transfer of memory traces from the HS to the neocortex; instead it involves the stabilization of LTP at synapses underlying episodic memory traces encoded in the HS. Support for this prediction can be found in recent studies using CA1-specific knockout mice. These studies suggest that NMDA receptors in CA1 continue to play a critical role in memory consolidation for a period spanning more than a week *after* the initial acquisition of memory (Shimizu et al. 2000). This suggests that synapse specific plasticity *within* the HS may be crucial for episodic memory consolidation, and hence, the consolidation of a memory trace may involve the consolidation of the memory trace within the HS.

The prediction that error-sensitive episodic memory traces of memorable events remain in the HS, however, does not preclude a transfer of certain types of information from the HS to similarity-based semantic and causal representations in cortical circuits (see Section 10.2).

7.1 Sleep, consolidation and forgetting

It is speculated that the consolidation of episodic memory traces within the HS occurs, in part, as a result of an automatic reactivation of memory traces during sleep. It is also speculated that such a process of consolidation is accompanied by a complementary process of forgetting (cf. Winson 1985; Maquet, Peters, Aerts, Delfiore, Degueldre, Luxen & Franck 1996; Stickgold 1998; but for a contrary view see Vertes & Eastman 2000).

The time available for processing and memorizing an event is extremely limited, since the organism is often forced by a dynamic environment to continually shift its attention to ongoing experiences and actions. In particular, the time available is usually insufficient for fully evaluating an event's significance. Hence, the episodic memory system seems to have adopted a promiscuous strategy; during alert wakefulness, it memorizes almost any experience that is construed as an event or a situation by the cognitive apparatus.

While such a strategy ensures that everything of possible significance would be memorized by the HS, it also loads the HS rather quickly with a large number of memories, and increases the potential for cross-talk among these memories. Consequently, it is important that the episodic memory system resort to active forgetting. It is speculated that during sleep, episodic memory traces are activated at random and evaluated for their significance. The activation of an event's memory trace in the HS reinstates its bindings in cortical circuits and leads to the reconstruction of the event (cf. Section 4.1). This reconstruction enables the evaluation of the event by cortical and subcortical structures. If the event is found to be significant, its episodic memory trace is consolidated by stabilizing the LTP of synapses underlying the event's episodic memory trace in the HS (and perhaps, also by the recruitment of additional cells). But those events found not to be significant are *actively* forgotten by inducing a reversal of synaptic potentiation.[24] This cycle of memorization and selective consolidation/forgetting repeats itself with every cycle of wakefulness and sleep. As a result of this process, *events deemed significant persist in memory, but the probability that an insignificant event remains in memory decreases with each passing cycle.*

As discussed in Section 10.2, the reactivation of events during sleep also enables the transfer of certain types of information from episodic memory traces to semantic and causal knowledge structures in the cortex.

To make the process of consolidation and forgetting concrete, let us consider a simplified numerical example. Let us assume that a person is awake 16 hours a day, and *on an average*, acquires one episodic memory *every* ten seconds during wakefulness. This implies that, on an average, a person acquires 5,760 episodic memories per day. Let us also assume that during an average night's sleep, the person's mind/brain randomly samples and evaluates 9,000 memories. Finally, let us assume that from age 2 onwards a person experiences — on an average — one memorable event every day of one's life. Under this set of assumptions, at age 75, a person's episodic memory would contain 74,637 events. Of these, 26,645 would be significant and the remaining 47,992 would be insignificant. Furthermore, the number of *insignificant* memories acquired on any given day will gradually decrease over time as follows: Of the 5,759 insignificant memories acquired on that day, 5,069 will survive after one day, 2355 will survive after one week, and only 125 will stay intact after one month.

[24] The temporal extent of an event's significance may vary widely. Some memories may have significance for only a few minutes, while others may remain significant for hours, days, years, or even a lifetime.

Region	ECer	ECee	ECro	ECcr	ECce	DG	CA3	CA2	CA1	Sub
# cells (million)	2.75	0.75	0.75	2.75	1	15	2.7	0.8*	15	4.5

Table 3: Number of cells in each model region. These region sizes have been informed by (West 1990; West & Slomianka 1998a,b). Values marked with an asterisk are plausible values.

Support for the plausibility of the consolidation-and-forgetting-during-sleep hypothesis can be found in animal studies showing that "memory traces" of events occurring during wakefulness are reactivated spontaneously during sleep (Wilson & McNaughton 1994; Shen & McNaughton 1996; Skaggs & McNaughton 1996; Qin, McNaughton, Skaggs & Barnes 1997; Louie & Wilson 2001).

8 A quantitative analysis of SMRITI

The proposed model of episodic memory formation has been simulated based on the computational abstraction of cells, synapses, and LTP described in Section 3. These simulations confirm that functional units required for encoding an event get recruited as expected. These simulations, however, are on a small-scale (several hundred cells per region) and do not explicate the statistical properties of the model arising from the large scale of the HS. Consequently, a partial quantitative analysis of the full-scale model has been carried out. This section presents some results of this analysis; a more detailed discussion appears in (Shastri 2001b).

8.1 Parameter values and assumptions underlying the quantitative analysis

The quantitative analysis requires the specification of values for system parameters pertaining to the conceptual structure, typical load of episodic memory, architecture of the HS, and the behavior of synapses, cells, and LTP. While many of these parameter values have been determined from available empirical data, others have been assigned plausible values based on indirect evidence or computational considerations. The *precise* numerical value of system parameters, however, are not too critical since the primary objective of this analysis is to evaluate the plausibility and overall characteristics of the model. Some key system parameter values used in the analysis are discussed below.

8.1.1 Relational schemas and entities underlying episodic memories

It is assumed that an adult's conceptual structure has a repertoire of about 16,000 "high-level" relational schemas (or frames), any one of which can form the basis for encoding an episode. Examples of such relational schemas are throw, lob, chuck, toss, walk, run, sprint, etc. Since there does not exist any direct means of determining the actual number of such relational schemas in the mind/brain, an indirect estimate of this number is obtained by turning to language. The number 16,000 corresponds to the approximate number of distinct verbs (or verb *lemmas*) used and comprehended by the collective English speaking population.[25] Furthermore, it is assumed that 50,000 entities are represented in an adult's conceptual structure, and any of these can serve as role-fillers in events.

8.1.2 Anatomical assumptions

The number of cells in various regions and the size of various projective fields (PFs) are as shown in Tables 3 and 4. The numbers are based, in part, on data provided in (Amaral, Ishizuka & Claiborne 1990; Amaral & Witter 1995; West 1990; West & Slomianka 1998a,b). Plausible values were chosen for region and PF sizes if good estimates were not known. The proportion of principal cells to Type-2 interneurons in CA3 and CA1, and that of principal cells to Type-3 interneurons in CA2 is assumed to be 10:1 (see, Olbrich & Braak 1985). Finally, in order to make the analyses tractable, it is assumed that (i) PFs are uniformly distributed over their respective target regions and (ii) the postsynaptic potential is a square pulse.

[25] The estimate of 16,000 is obtained by dividing 64,000 by 4, where 64,000 equals the approximate number of distinct word forms identified as verbs in the British National Corpus (Burnard 1995; http://info.ox.ac.uk/bnc) a comprehensive 100 million word collection representing the current usage of written and spoken English, and 4 is the average number of variants of each verb (for example, the verb lemma "walk" can appear as walk, walked, walking, and walks).

Projection	PF size	Projection	PF size	Projection	PF size
ECee → DG	17,000	ECee → CA2	2000*	CA3 → CA2	8000*
ECro → DG	17,000	ECer → CA1	8500*	CA2 → CA1	800*
ECro → CA3	6000*	ECro → Sub	5000*	CA1 → Sub	500*
ECro → CA2	2000*	DG → CA3	14		

Table 4: Size of projective fields (PFs) used in the quantitative analysis. Values for EC → DG projection and the DG → CA3 projection are based on (Amaral, Ishizuka & Claiborne 1990; Amaral & Witter 1995; West 1990). Values marked with an asterisk are plausible values.

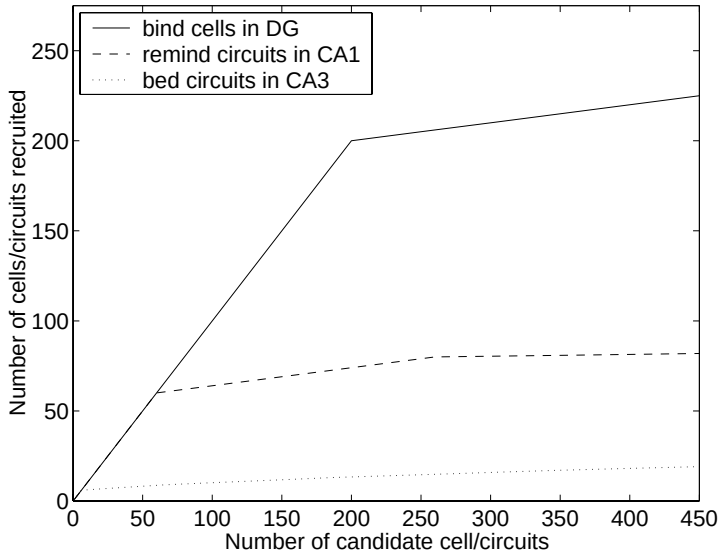

Figure 13: Soft-WTA functions governing the ratio of the number of candidate cells/circuits to the number of recruited cells/circuits in regions DG, CA3, and CA1.

To initialize the analysis, it is assumed that 600 linking cells are recruited for each role, +:entity, and ?:entity ensemble in regions ECro, ECce, and ECee, respectively. It is also assumed that 1200 linking cells are recruited for each ?:relation, and +:relation ensemble in regions ECer and ECcr, respectively. These linking cells are assumed to be uniformly distributed within their respective regions (note that a cell may be part of multiple ensembles).

8.1.3 Modeling of the soft-WTA effect of Type-1 interneurons

As mentioned in Section 6, soft-WTA networks formed by Type-1 inhibitory interneurons limit the number of cells that are actually recruited from the pool of adequately connected candidate cells. The soft-WTA function is modeled as a piecewise linear function. This enables a more plausible modeling of the inhibitory effect of Type-1 interneurons than that afforded by a function that simply chooses a fixed number of candidate cells. The soft-WTA functions for regions DG, CA3, and CA1 are shown in Figure 13. A relatively low value of the candidate-to-recruit multiple in CA3 contributes to the proper functioning of the model. The soft-WTA functions for CA2 and the subiculum are similar to those for CA3 and CA1, respectively.

8.2 Quantitative analysis of memorization

Let us consider the memorization of an event $E1$ described by "John gave Mary a book in the library on Tuesday," under the assumptions discussed in Section 8.1. Furthermore, let us assume that at the time of memorization, the state of episodic memory is as follows:

Statistic	DG [*bind*]	CA3 [*bed*]	CA2 [*bei*]	CA1 [*remind*]	Sub [*reinstate*]
P_{fail}	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$	$< 10^{-18}$
E$\langle candidates \rangle(\sigma)$	195.0 (14.0)	412.7 (20.3)	56.9 (7.5)	56.6 (7.5)	159.9 (12.7)
E$\langle recruits \rangle(\sigma)$	195.0 (14.0)	16.0 (4.0)	13.2 (3.6)	51.4 (7.2)	56.0 (7.5)

Table 5: P_{fail} denotes the probability that cells or circuits with suitable connections will *not* be found in a target region for recruitment as a functional unit during the memorization of the event *E1* (John gave Mary a book in the library on Tuesday). E$\langle candidate \rangle$ denotes the expected number of cells or circuits that receive adequate connections, and hence, will be candidates for recruitment during the memorization of *E1*. E$\langle recruits \rangle$ specifies the expected number of candidate cells or circuits that will be recruited for each functional unit. Thus E$\langle recruits \rangle$ specifies the expected number of "copies" of each functional unit in the episodic memory trace of E1. The quantities in parentheses are standard deviations.

- 75,000 events involving a total of 300,000 bindings are encoded in episodic memory (see Section 7.1),
- of these, 75 events are *give* events,
- each role-entity binding in $E1$ also occurs three or more times in previously memorized *give* events (for example, John is the *giver* in at least three previously memorized *give* events), and
- each entity filling a role in $E1$ also fills other roles in previously memorized *give* events (for example, there are at least four *give* events in memory where "John" is the recipient), and
- each entity filling a role in $E1$ occurs in at least 25 previously memorized events involving relations other than *give*.

The results of the quantitative analysis of memorization are displayed in Table 5.[26]

The quantitative analysis indicates that (i) the failure probability is practically zero and (ii) multiple copies are recruited for each functional unit. Thus the HS is capable of forming redundant memory traces of events presented to it for memorization. Let us now examine how recruited functional units respond to cues during retrieval.

8.3 Quantitative analysis of retrieval

We begin the analysis of retrieval by examining the behavior of $remind$ circuits. These circuits lie at the apex of an event's memory trace and their firing in response to a cue signals a match between the cue and the event encoded by the memory trace. Recall that the response of *remind* circuits, together with that of *reinstate* cells, constitutes the primary "output" of the HS-based episodic memory system.

Figure 14 shows the *expected* number of $remind(E1)$ circuits firing in response to retrieval cues containing different numbers of binding errors, including matching cues containing *zero* binding errors (the response to a *partially specified* cue matching $E1$ is discussed in Section 8.4). The error-bars in Figure 14 (and subsequent figures) indicate the standard deviation of the response.

Ideally, *remind(E1)* circuits should fire in response to a matching cue, and not fire in response to a cue containing one or more binding-errors. As shown in Figure 14, the response of *remind(E1)* circuits comes close to the ideal behavior. The expected number of *remind(E1)* circuits firing in the zero binding-error (match) condition is sharply higher than that in any of the binding-error conditions. Moreover, the expected number of $remind(E1)$ circuits responding to cues involving relations other than *give* is essentially zero (< 0.004). Thus the episodic memory traces formed in the model exhibit a strong form of pattern separation.

The *signal-to-noise* ratio of the *remind* circuit response, that is, the ratio of the expected number of *remind* circuits firing in response to a matching cue to the expected number of *remind* circuits firing in response to a cue with one binding-error, is more than 5.8. The high signal-to-noise ratio together with the relatively small standard deviation provides a robust basis for discriminating between a matching cue and an erroneous cue.

[26] In this and subsequent analysis, all standard deviations are calculated by approximating the response of upstream regions by the expected value of their response.

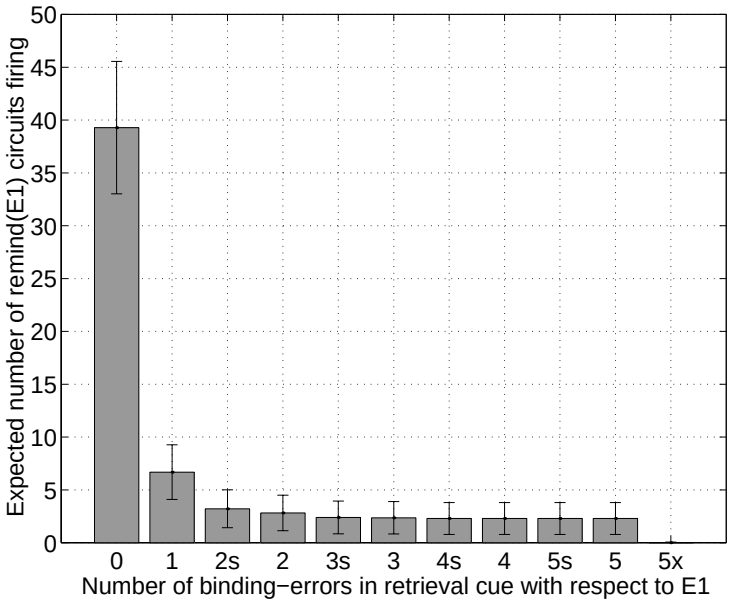

Figure 14: Response of *remind(E1)* circuits in CA1 as a function of the match between the retrieval cue and $E1$. Suffix "s" indicates that 2 of the binding-errors in the cue arise from swapping role-fillers in $E1$. Condition X refers to a retrieval cue involving a different type of event (for example, walking).

Since $remind(E1)$ circuits sit at the apex of $E1$'s memory trace, their response is affected not only by interference among *remind* circuits, but also by various types of errors and interference among preceding *bind*, *bed*, and *bei* functional units. In view of this, the robust response of $remind(E1)$ circuits is significant.

Figures 15 and 16 depict the response of *reinstate*(⟨ *giver* = John ⟩| *E1*) cells and *bei(E1)* cells, respectively. Recall that while *reinstate*(⟨ *giver* = John ⟩| *E1*) cells should fire in response to any cue that matches $E1$, *bei(E1)* cells should *not* fire in response to any maching cue. Instead, the latter should fire in response to any cue containing one or more binding-errors.

Figure 17 shows the expected number of $bed(\langle giver = John \rangle | E1)$ circuits that will fire in response to various types of bindings in a retrieval cue. Recall that $bed(\langle giver = John \rangle | E1)$ circuits should fire in response to any cue containing a binding of the form $\langle giver = fx \rangle$, where fx is an entity other than "John." They should not fire otherwise.

Figure 18 depicts the response of $bind(\langle giver = John \rangle)$ cells to various types of bindings in a retrieval cue. Recall that $bind(\langle giver = John \rangle)$ cells should fire in response to the binding $\langle giver = John \rangle$, but not otherwise.

As shown in Figures 15 through 18, the response of various functional units is close to the ideal behavior.

8.4 Effect of cue size on response of memory trace

An interesting property of SMRITI is that the memory trace of an event responds more vigorously to a partial cue matching the memorized event than to a fully specified cue matching the memorized event. For example, $E1$'s memory trace produces a more emphatic "yes" response to the cue "Did John give Mary a book?" than to the cue "Did John give Mary a book in the library on Tuesday?" The difference in the strength of response is caused by reduced interference as a result of fewer *bed* circuits and *bei* cells being active in response to the smaller cue. The signal-to-noise ratio of *remind* circuits improves from 5.9 for a cue with five bindings to 7.8 for a cue with three bindings.

8.5 Robustness of recruitment and response properties of functional units

The number of functional units recruited during the encoding of an event, as well as the fidelity of their response to retrieval cues are affected by (i) the number of other episodic memories involving the *same* relation as the encoded event and (ii) the number of other episodic memories involving some of the *same* bindings as the encoded event. Hence

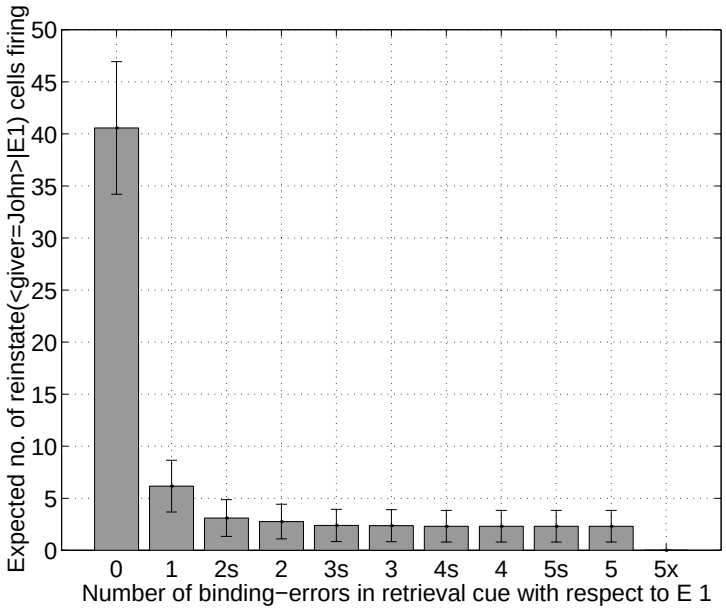

Figure 15: Response of *reinstate*(⟨ *giver* = John ⟩| *E1*) cells in the subiculum as a function of the match between the retrieval cue and $E1$.

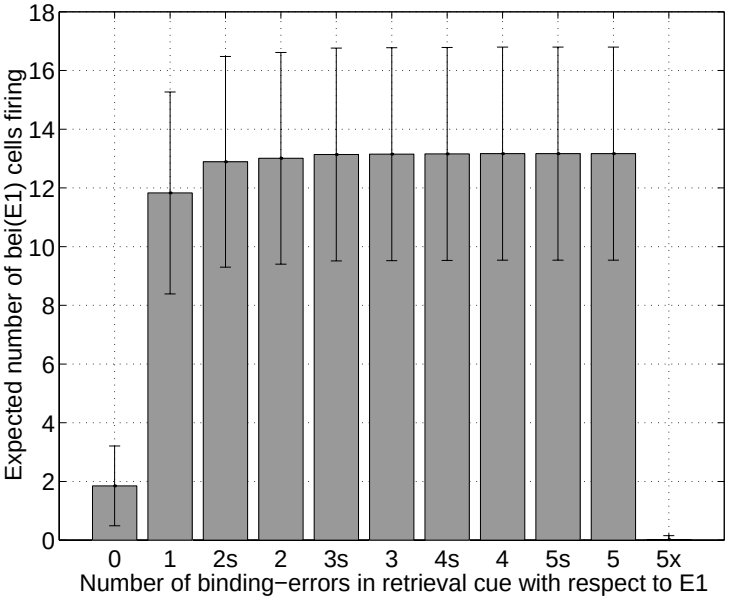

Figure 16: Response of *bei(E1)* cells in CA2 as a function of the match between the retrieval cue and $E1$.

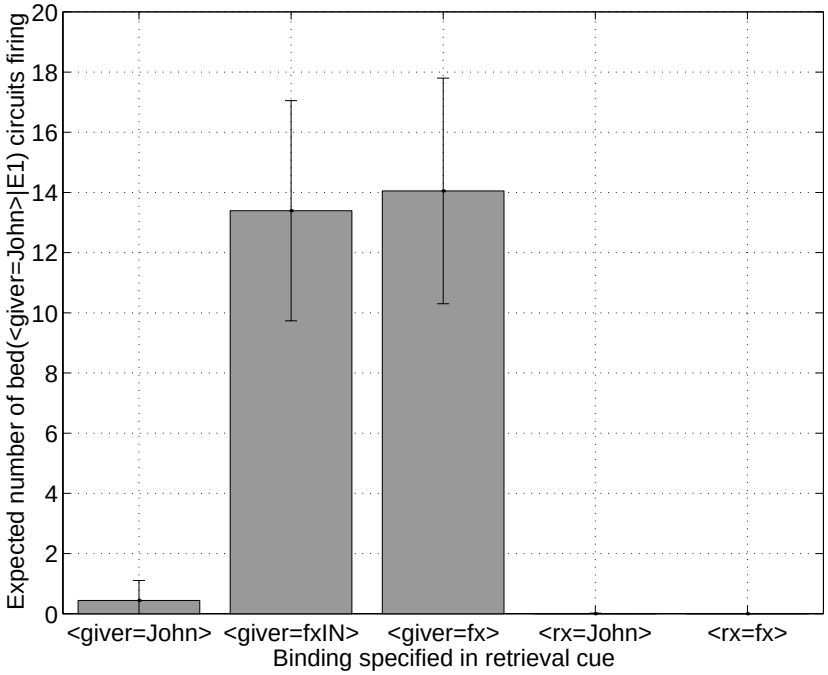

Figure 17: Response of *bed(*$\langle giver = John \rangle | E1$*)* circuits in CA3 to bindings in a retrieval cue. Here fx refers to an entity other than John and rx refers to a role other than *giver*. The condition $\langle giver = fxIN \rangle$ refers to a cue with a binding where John occurs in the cue as a filler of some role other than *giver*.

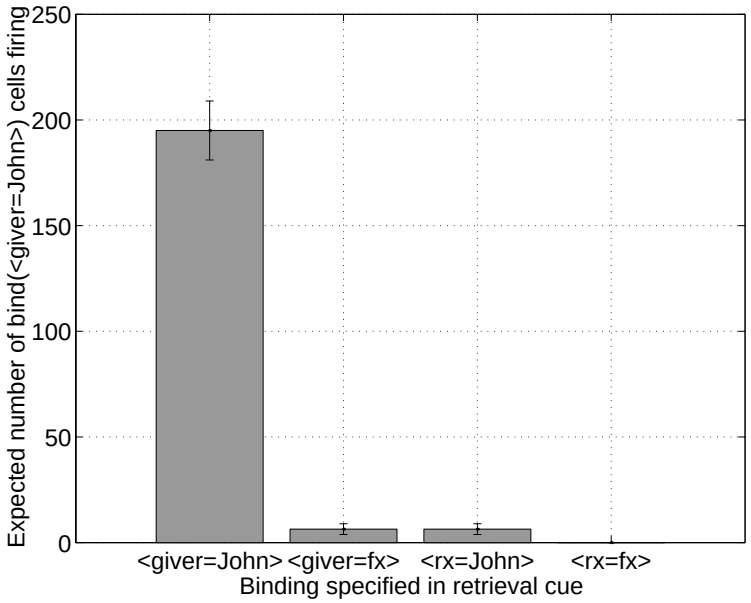

Figure 18: Response of $bind(\langle giver = John \rangle)$ cells in DG to bindings in a retrieval cue. Here fx refers to an entity other than John and rx refers to a role other than *giver*.

the recruitment and response properties of the episodic memory trace of $E1$ were analyzed by varying (i) the number of *prior* memories involving the relation *give* from 0 to 125 and (ii) the number of times each binding in $E1$ occurs in previously memorized *give* events from 4 to 15.

The number of functional units recruited for *E1* decreases as the number of memories involving *give* increases. A similar decrease is observed as the number of occurrences of the same binding in previously memorized events increases. This decrease in the number of units recruited is accompanied by a decrease in the number of units responding to a cue, and a decrease in the signal-to-noise ratios of the response. The signal-to-noise ratio of *remind* circuits, however, remains above 4.0 in all of the above conditions and continues to provide a sufficient basis for discriminating between a matching cue and an erroneous cue.

8.6 Effect of event arity

The arity of an event has a significant impact on the recruitment and response of *bei* cells, and in turn, on the recruitment and response of *remind* circuits and *reinstate* cells. Recall that *bei* cells integrate the response of *bed* circuits in CA3, and hence, a CA2 cell must receive afferents from a larger number of CA3 cells in order to be a candidate *bei* cell for a higher arity event. This reduces the expected number of recruited *bei* cells. However, the overall response of *bei* and *remind* functional units remains robust even when the arity of the memorized event is increased to six. In particular, the signal-to-noise ratio of *remind* circuits remains above 5.4.

8.7 Effect of cell loss

The effect of cell loss on the recruitment and response properties of an episodic memory trace was also evaluated. In this analysis, the loss of a cell also entails the loss of all its outgoing and incoming links. All other parameters values are assumed to be the same as those used in calculating the results presented in Sections 8.2 and 8.3.

Figure 19 shows the effect of cell loss in EC alone on the recruitment and the subsequent response of *remind* circuits. Since EC forms the first stage of the memory trace, any loss of cells in EC impacts the recruitment of each functional unit in the memory trace. Nevertheless, the quantitative analysis shows that the memory trace of an event remains viable even after a 10% cell loss in EC. Though the matching response reduces in absolute terms, it is sufficient to enable downstream circuits to distinguish between a match and a non-match response, since the signal-to-noise ratio is still 7.8. The recruitment and response of *remind* cells, however, falls to an unacceptable level when the cell loss in EC increases to 15%.

Changes in the recruitment and response properties of *remind* circuits resulting from cell loss in every region of the HF (DG, CA3, CA2, CA1 and the subiculum) and every region of the HS (EC + HF) were also analyzed. The analysis shows that the memory trace of an event remains viable for up to ca. 25% cell loss in every region of the HF (see Figure 20), and for up to ca. 10% cell loss in every region of the HS (see Figure 21). The effects of focal lesions in individual regions of the HF are discussed in Section 11.4.

8.8 Significance of quantitative analysis results and their relation to experimental data

The quantitative analysis explicates the robustness of a memory trace with respect to cell loss. This robustness stems from the physically dispersed and redundant nature of a memory trace in SMRITI. Given that *multiple* copies are recruited for each functional unit, and given that these copies are distributed within a region, the probability that limited amounts of cell loss in a region will destroy *several* copies of a functional unit for any given memory trace is small. The response of a memory trace is essentially unaffected by limited amounts of cell loss (for example, less than 5%). As the degree of cell loss increases, the system's response degrades gradually at first, and then undergoes catastrophic failure.

The strong impact of EC cell loss on episodic memory is consistent with findings that damage to EC is strongly correlated with memory loss. This result is also significant from the perspective of understanding the effect of Alzheimer's disease on episodic memory function. One of the first brain regions to be affected by Alzheimer's disease is EC (Hyman & Van Hoesen 1989; Van Hoesen & Hyman 1990; Gomez-Isla, et. al 1996). The results of the quantitative analysis offer a computational explanation for these observations by showing that the episodic memory function is particularly sensitive to loss of cells in EC, and that a cell loss in EC approaching 15% can have a profound impact on episodic memory function. At the same time, the analysis shows that low levels of cell loss (less than 10%) have a minimal impact on the performance of the HS-based episodic memory system. This is consistent with the observation

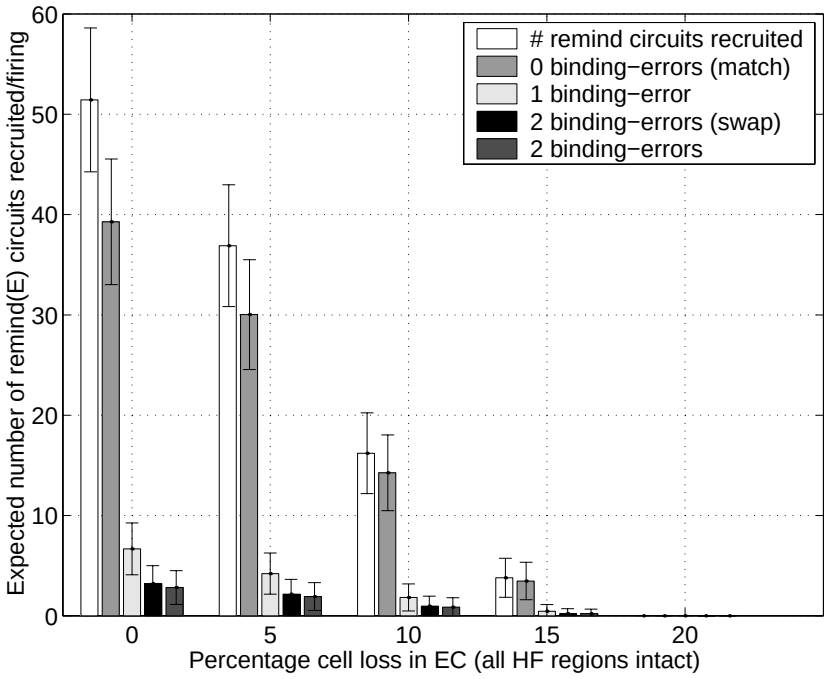

Figure 19: The effect of cell loss in EC (HF intact) on the recruitment and response properties of *remind(E)* circuits. Since the response properties are fully revealed in the way *remind* circuits respond to cues with zero, one, or two binding-errors, results for cues with more than two binding-errors are not displayed.

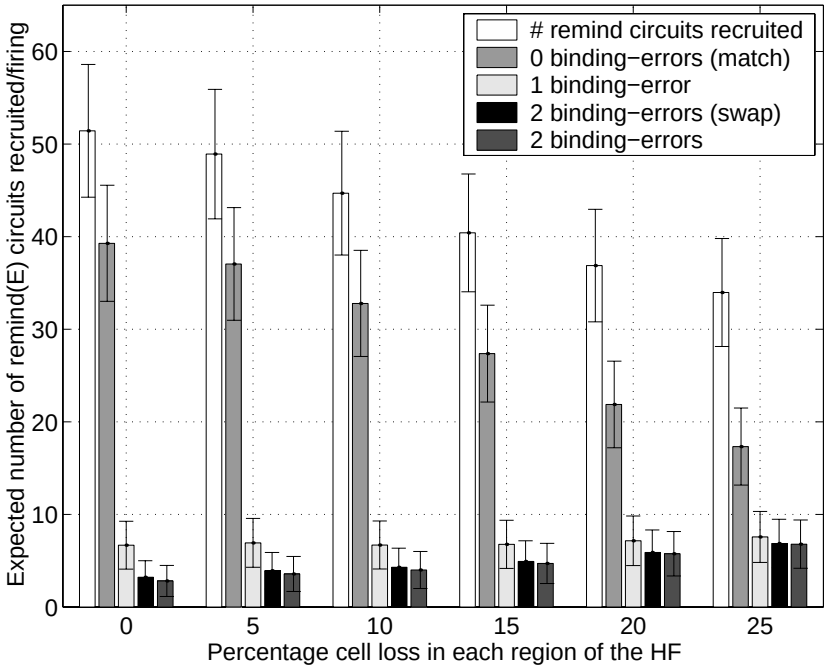

Figure 20: The effect of cell loss in every region of the HF (EC intact) on the recruitment and response properties of *remind(E)* circuits.

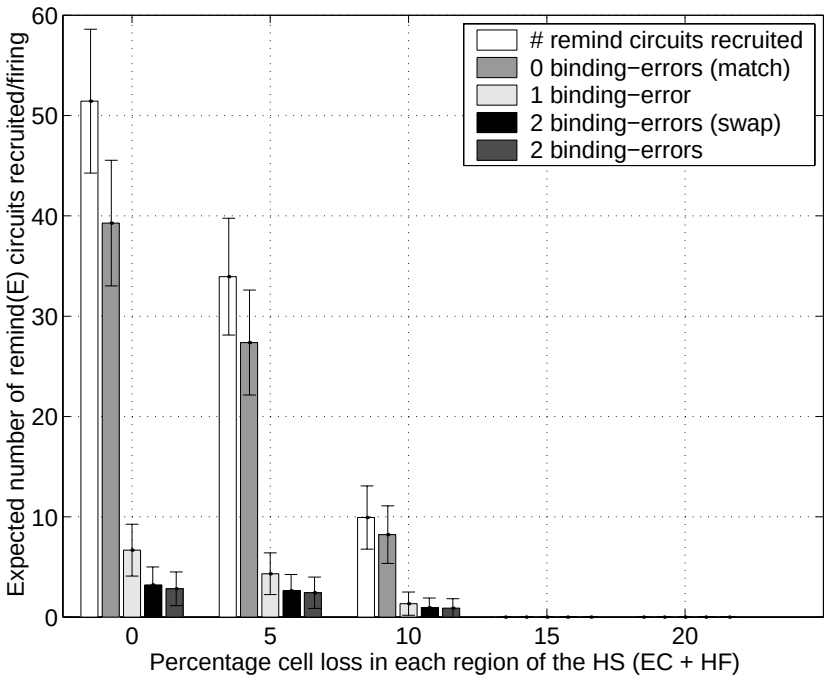

Figure 21: The effect of cell loss in every region of the HS (EC + HF) on the recruitment and response properties of *remind(E)* circuits.

that episodic memories generally remain intact well into old age in spite of low levels of cell loss that might occur as part of the normal aging process. Furthermore, the minimal impact of low-level cell loss on memory acquisition and retrieval suggests that very early stages of Alzheimer's disease may not manifest any measurable memory loss, and hence, go by undetected.

The quantitative analysis of SMRITI also helps explain certain experimental findings in psychological studies of memory function. For example, it shows that as the number of memorized events involving a particular role-entity binding increases, both the neural mass recruited for encoding events involving that binding and the signal-to-noise ratio of the response generated by memory traces of events involving that binding decrease. If one makes the reasonable assumption that the signal-to-noise ratio of the response is inversely related to retrieval time, it follows that as the number of memorized events involving a particular role-entity binding increases, the response time for retrieving a particular event involving that binding increases. Thus SMRITI can offer a biologically grounded explanation of the fan effect (Anderson 1974; Radvansky, Spieler & Zacks 1993; Anderson & Reder 1999; Radvansky 1999) wherein greater the number of facts memorized about a particular entity, the longer it takes to retrieve a particular fact about that entity.[27]

Since the neural mass recruited for an episodic memory trace depends on the neural mass of collector and enabler ensembles of entities participating in the memorized event, SMRITI predicts that the degree of fan effect will vary as a function of the type of entities involved in memorized events. In particular, SMRITI predicts that all else being equal, entities having a high cortical mass will have a greater number of candidate cells for recruitment as functional units than entities with a low cortical mass (see Section 6.5). Hence, entities with a high cortical mass will be less susceptible to the fan effect than entities with a low cortical mass. Moreover, since the match and mismatch between bindings in a cue and those in memorized events is computed in parallel, retrieval times and the fan effect for mismatching probes (foils) will not be significantly higher than that for matching probes (targets). A detailed accounting of various aspects of the fan effect using SMRITI is a topic of future research.

[27] In the basic fan effect experiments (Anderson 1974), the fan of a concept was essentially the number of times the concept filled a specific role in the studied facts. For example, facts in the study list involved the relational schema "entity X is in location Y" and a fact such as "A hippie is in the park" corresponded to the bindings ($\langle$ entity = a hippie $\rangle$, $\langle$ location = the park $\rangle$). Thus the occurrence of "A hippie" in n facts resulted in n occurrences of the binding $\langle$ entity = a hippie $\rangle$ in memorized facts.

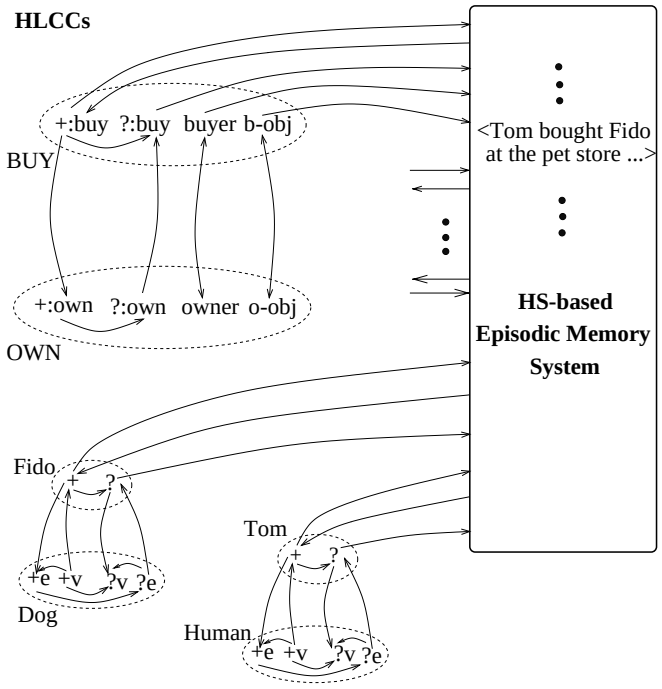

Figure 22: A schematic of the interaction between the HS-based episodic memory system and high-level cortical circuits encoding causal and semantic knowledge (HLCCs). The HLCCs are depicted on the left and the HS system is depicted on the right. The only links shown between HLCCs and the HS are the ones pertaining to the event "Tom bought Fido at the pet store in January." See text for details.

9 Interactions between the HS-based episodic memory and cortically expressed semantic and conceptual knowledge

An example of interaction between the HS-based episodic memory system and HLCCs encoding causal and semantic knowledge is shown in Figure 22. The HLCCs depicted on the left encode the following knowledge: (1) If you buy something you own it. (2) Tom is a human. (3) Fido is a dog. It is assumed that one of the events encoded in the HS is "Tom bought Fido at the pet store in January".

Relational schemas *buy* and *own* are encoded by focal-clusters labeled BUY and OWN, respectively. Entities Tom and Fido, and categories human and dog are encoded by focal-clusters labeled *Tom* and *Fido*, and *Human* and *Dog*, respectively. In order to keep network diagrams and activation traces tractable, we have shown buy and sell relations to comprise of only two roles each. In actuality, these relations involve other roles including the location and time of occurrence.

9.1 Expression of semantic and causal knowledge via interconnections between focal-clusters

We discussed the representational significance of entity and relational focal-clusters in Section 5.1. The representational significance of category focal-clusters is as follows: The focal-cluster of a category such as *Human* consists of a pair of enabler ensembles: *?e:Human* and *?v:Human*, and a pair of collector ensembles: *+e:Human* and *+v:Human* (recall that each label within a focal-cluster denotes an ensemble of cells). While the ensembles *+v:Human* and *?v:Human* participate in the expression of knowledge (episodic and semantic facts) involving the category human as a whole, the nodes *+e:Human* and *?e:Human* participate in the expression of knowledge involving particular (non-specific) instances of Human. The complete significance of the interconnections *within* entity and category focal-clusters and *between* entity and category focal-clusters is described in (Shastri 2000). For now, it is only important to note that links such as the one from *?e:Dog* to *?:Fido* cause any query about an unspecified instance of a category to lead to analogous queries about specific instances of that category (the query "Is there a furry dog?" leads to the query "Is Fido furry?").

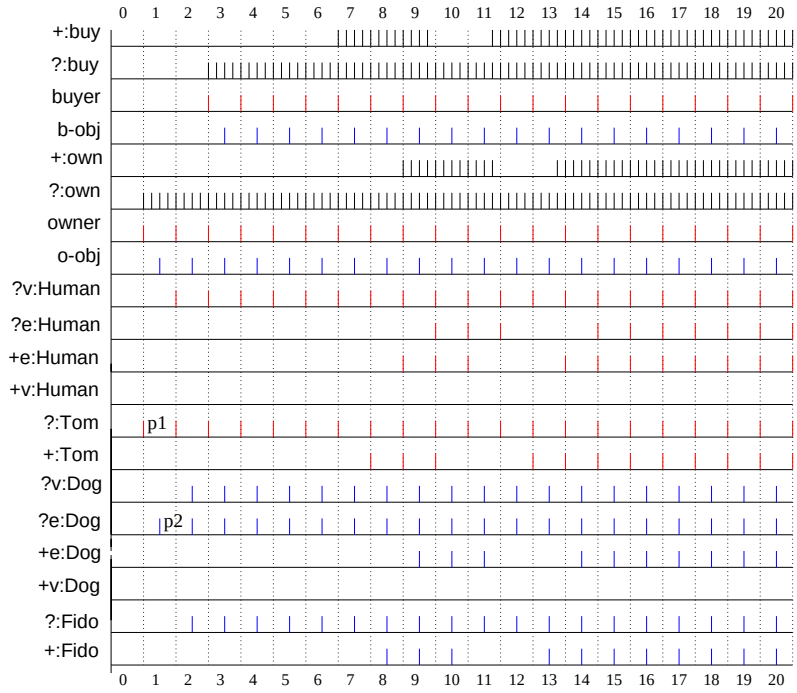

Figure 23: A schematic of the activity in HLCCs subsequent to the query "Does Tom own a dog?" posed in cycle 1. By cycle 3, this query gives rise to several queries including "Did Tom buy Fido?". The latter matches the event "Tom bought Fido at the pet store in January" in episodic memory. The match is conveyed to *+:buy* by cycle 12. By cycle 13, *+:Tom, +e:Fido* also receive activation. The onset of stable activity in *+:own* and *+e:Dog* in cycle 14 marks an affirmative answer to the query.

The interconnections between the OWN and BUY focal-clusters shown in Figure 22 encode the causal knowledge "if you buy something, you own it" (Shastri 1999a). This is expressed by:

- A link from *?:own* to *?:buy*. This link causes a query or a search for an explanation about owning to lead to a query about buying (a possible explanation of A owning B is that A bought B).
- A link from *+:buy* to *+:own*. This link causes an assertion about buying to lead to an assertion about owning (a likely consequence of A buying B is that A owns B).
- The reciprocal links between the *buyer* and *owner* ensembles, and between the *b-obj* and *o-obj* ensembles (here *o-obj* is an abbreviation of *own-object*). Such links between role ensembles cause interconnected role ensembles to fire in synchrony. Since dynamic role-entity bindings are expressed via synchronous firing of role and entity nodes, this leads to the propagation of dynamic bindings across connected focal-clusters.

9.2 Flow of activity between HLCCs and the HS

Now assume that the query "Does Tom own a dog?" arises in HLCCs. This will appear as a pattern of activity shown in cycle 1 of Figure 23. Soon thereafter (cycle 3 in Figure 23), the propagation of activity from the OWN to the BUY focal-cluster and along the entity and category structures will lead to a pattern of activity in HLCCs corresponding to several queries including "Did Tom buy a dog?" and "Did Tom buy Fido?". Note the activation of the *?:buy* ensemble, and the synchronous firing of the *buyer* and *?:Tom* ensembles and the *b-obj*, *?:Fido*, and *?e:Dog* ensembles.

The activity of ensembles in the BUY cluster as well as the activity of the *?:Tom* and *?:Fido* clusters will propagate to the HS where the memory trace of the event "Tom bought Fido at the pet store in January" will match and propagate activity back to *+:buy*, *+:Tom*, and *+:Fido*. The propagation of activity along the links from the *+:buy* ensemble to the *+:own* ensemble will lead to the activation of the *+:own* ensemble. The activity of the *own* ensemble together with

the synchronous firing of the *owner* and *+:Tom* ensembles and the *o-obj* and *+e:Dog* ensembles will lead to a dynamic representation of the assertion "Tom owns a dog." This will constitute an affirmative answer to the initial query "Does Tom own a dog?".

9.3 Stepping through the retrieval process within the HS

The response of the HS-based episodic memory system to the query "Did Tom buy Fido?"is depicted in Figure 24. The activity of *?:Tom* and *?:Fido* ensembles in HLCCs propagates to their respective linking cells in ECee, the activity of *buyer* and *b-obj* role ensembles propagates to their respective linking cells in ECro, and the activity of *?:buy* ensemble propagates to its linking cells in ECer. The firing of these cells leads to the following activity in the HS (the reader may wish to refer to Figure 7 and map *buy* to R_k, *buyer* and *b-obj* to r_1 and r_2, respectively, and Tom and Fido to f_1 and f_2, respectively):

1. The $bind(\langle$ *owner=Tom* $\rangle)$ cells and $bind(\langle$ *o-obj = Fido* $\rangle)$ cells in DG fire in distinct phases.
2. At the same time, bed cells pertaining to bindings involving roles *owner* and *b-obj* fire due to the arrival of activity from *owner* and *o-obj* linking cells in ECro, respectively. Similarly, $remind$ circuits pertaining to events involving the relational schema *own* fire due to the arrival of activity from *?:buy* linking cells in ECer.
3. The firing of $bind(\langle$ *owner=Tom* $\rangle)$ and $bind(\langle$ *o-obj = Fido* $\rangle)$ cells leads to the firing of Type-2 inhibitory interneurons associated with bed circuits of the form $bed(\langle$ *owner=Tom* $\rangle|*)$ and $bed(\langle$ *o-obj=Fido* $\rangle|*)$ (here * refers to any memorized event in which these bindings occur). The firing of these interneurons blocks the firing of bed cells of the form $bed(\langle$ *owner=Tom* $\rangle|*)$ and $bed(\langle$ *o-obj=Fido* $\rangle|*)$. Thus these bed cells fire for a couple of cycles, and then shut off.
4. Other bed cells of the form $bed(\langle$ *owner=fx* $\rangle|*)$ and $bed(\langle$ *o-obj=fy* $\rangle|*)$, however, continue to fire (here fx corresponds to any entity other than Tom, and fy corresponds to any entity other than Fido).
5. The firing of bed cells as described above leads to the sustained firing of all $bei(E')$ cells, where E' is any buy event in which Tom is not the buyer and/or Fido is not the bought object.[28]
6. The firing of $bei(E')$ cells blocks the firing of all $remind(E')$ cells, where E' is any buy event that mismatches the event "Tom bought Fido at the pet store in January."
7. However, $remind$ cells associated with buy events that have binding-errors for neither the binding $\langle$ *owner=Tom* $\rangle$, nor for the binding $\langle$ *o-obj = Fido* $\rangle$, continue to fire. This includes $remind(E)$ circuits, where E is the event "Tom bought Fido at the pet store in January."
8. The sustained firing of $remind(E)$ cells leads to the sustained firing of *+:buy* linking cells in ECcr.
9. The firing of $remind(E)$ cells together with the firing of the linking cells for *owner* and *o-obj* roles in ECro leads to the activation of $reinstate(\langle owner = Tom \rangle|E)$ and $reinstate(\langle$ *o-obj = Fido* $\rangle|E)$ cells.
10. The firing of the above $reinstate$ cells in turn activates the linking cells for *+:Tom* and *+:Fido* in ECce.
11. The firing of *+:buy* linking cells in ECcr activates the ensemble *+:buy* in the cortical focal-cluster of BUY, and the firing of linking cells for *+:Tom* and *+:Fido* in ECce activates the ensembles *+:Tom* and *+:Fido* in the cortical focal-clusters *Tom* and *Fido*.

The example presented above is indicative of the sorts of inferences that can be drawn as a result of the interaction between semantic, causal and episodic knowledge. A better understanding of the full scope of the inferential power of such an integrated system can be obtained by referring to descriptions of SHRUTI (Shastri & Ajjanagadde 1993; Shastri 1999a, Shastri & Wendelken 2000) and process-based representations of actions schemas (Arbib 1994; Bailey et al. 1998; Narayanan 1997; Shastri et al. 2001). The HS-based episodic memory system can support a detailed simulation of a memorized event by activating sensorimotor programs and schemas and binding their roles and parameters to

[28] Note that $bei(E)$ cells, where E refers to the event "Tom bought Fido at the pet store in January," fire for a couple of cycles due to the transient firing of $bed(\langle$ *owner=Tom* $\rangle|E)$ and $bed(\langle$ *o-obj=Fido* $\rangle|E)$ cells. Other bei cells associated with events matching the cue also produce a similar transient activity.

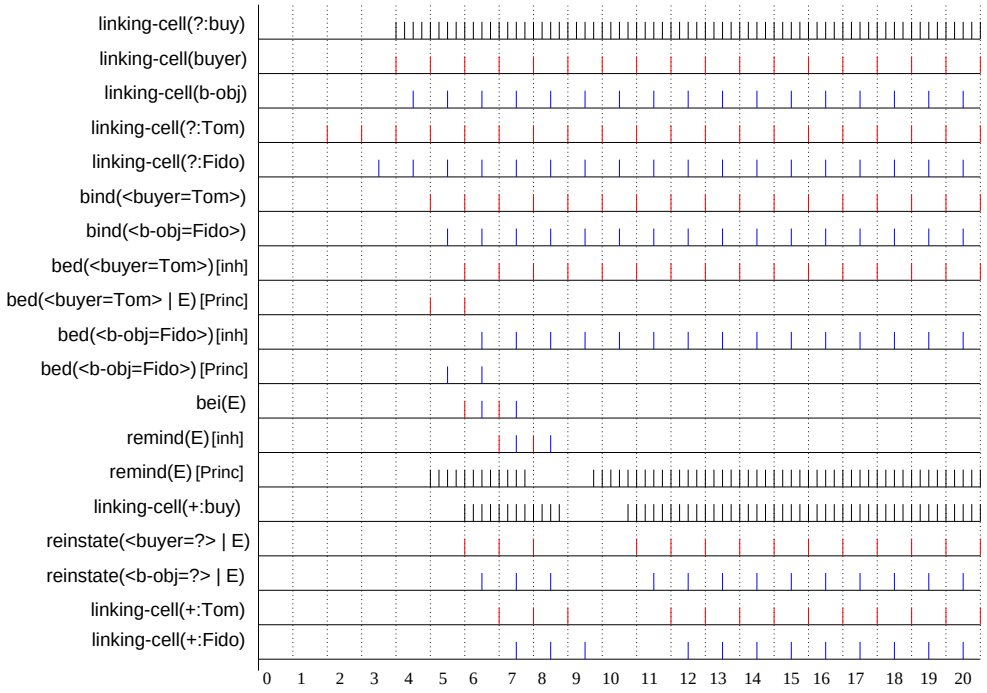

Figure 24: A schematic of the pattern of activity in the HS-based memory trace of the event "Tom bought Fido at the pet store in January" subsequent to the posing of the query "Did Tom buy Fido?". See text for details.

the appropriate entities and values retrieved from the event's episodic memory trace. In the case of more complex and extended events, several sets of bindings — one for each subevent comprising the event — may be required to carry out a simulated recreation of the event. The resulting evocation may be viewed as an interpolation of memorized snap-shots of the complex event.

10 Transfer of information from the HS to semantic structures

We have seen several examples of semantic and conceptual structures in previous sections. Let us consider another kind of semantic representation before discussing the transfer of information from the HS-based episodic memory system to cortically expressed semantic and conceptual representations.

10.1 Taxon-facts

While an episodic memory encodes a specific event, a taxon-fact[29] encodes a distillation or statistical summary of multiple events pertaining to a relational schema. In its simplest form, a taxon-fact encodes a measure of the probability that certain *types* of entities participate in a certain kind of event. Examples of taxon-facts (rather a verbalization of their representational import) are: "I often eat cereal for breakfast," "It usually rains during winter in Northern California," and "Soccer moms are likely to own minivans."

Figure 25(a) depicts a schematic of the encoding of a taxon-fact. Note that a taxon-fact, like an episodic memory, provides closure between the enabler (?) and collector (+) ensembles of a relational schema's focal-cluster. In the absence of relevant episodic memories, taxon-facts provide plausible and likely answers to queries posed about the relation.[30]

[29] The use of the term "taxon-fact" was inspired by the taxon-locale distinction proposed by O'Keefe and Nadel (1978).

[30] Whereas answers provided by episodic memories have a strong subjective feel of "I remember", those provided by taxon-facts do not.

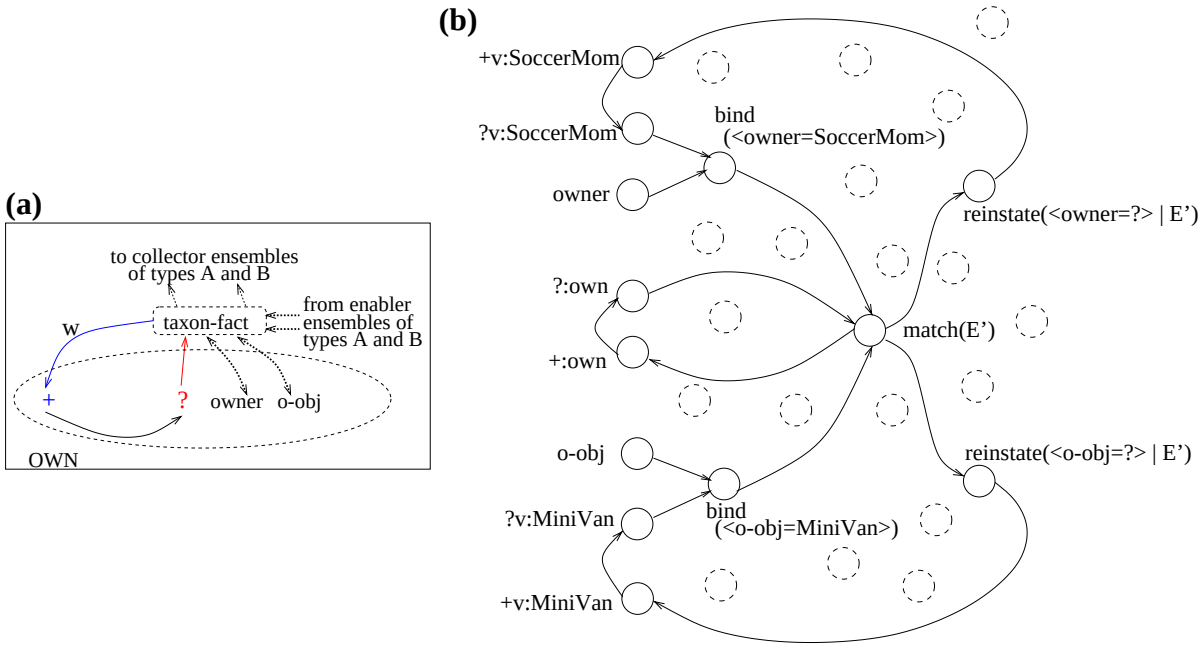

Figure 25: (**a**) A schematic of a cortically encoded taxon-fact representing that an owner of type A is likely to own an object of type B. Here A and B are types (categories) and w is a measure of the likelihood associated with this taxon-fact. Types A and B could refer to parents of young children and minivans, respectively. Note that each label in the figure denotes an ensemble of cells. (**b**) A detailed depiction of a neural circuit underlying the taxon-fact "Soccer moms are likely to own minivans." A single circuit is depicted, but the encoding will comprise of several copies of such a circuit.

10.2 Memory transfer from the HS to neo-cortical circuits

It is widely believed that over time episodic memories migrate from the HS to cortical circuits external to the HS (Squire 1992; Bontempi, et al. 1999). SMRITI, however, suggests a different sort of information migration from the HS to cortical circuits.

Events in episodic memory can affect cortically expressed semantic representations in an incremental manner. The HS-based memories of events are activated while reminiscing, reflecting, problem solving and sleeping. Once activated, these events trigger *reflexive* inferences (or "mental simulations") in cortical representations that correspond to the mind/brain seeking explanations and making predictions (Shastri 1999a; Shastri & Wendelken 2000). The resulting flurry of activity can lead to the fine-tuning of synaptic strengths, and in effect, to the modification of prior and conditional probabilities encoded in cortically expressed causal and semantic structures. For example, hearing about the event "John was mugged in Central Park last night" may increase the strength of the taxon-fact "Central Park is a dangerous place after dark".[31]

As argued in Section 11.3, the error-sensitive encoding of a memorable event (that is, the event's episodic memory trace) must remain in the HS. But *simpler* similarity-based versions of event memories could be encoded within cortical circuits over time (cf. Cermak 1984) using representational machinery analogous to that of taxon-facts. An example of such an encoding is shown in Figure 25(b). The encoding consists of a *binding-detector* cell for each binding, a *match* (or chunking) node for forming a soft-conjunct of the bindings, and a *binding-reinstator* node for each binding. A comparison of Figures 25(b) and 7 — both of which depict the encoding of a relational item involving two bindings — shows the relative simplicity of the cortical representation.

Such a cortical representation of a "taxon-event" would have several properties that would distinguish it from an episodic memory trace: First, such an encoding would be insensitive to binding-errors and respond to a cue based on overall similarity. Second, since it is unlikely that any single cortical area outside of the HS receives converging inputs

[31] The synaptic weights associated with links shown in Figure 22 have evidential and probabilistic interpretations. For a detailed description of how causal knowledge is encoded and how synaptic weights map to evidential strengths and probabilities see (Shastri 1999a; Wendelken & Shastri 2000; Shastri & Wendelken 2000).

from focal-clusters of the *full spectrum* of relations and entities, it is unlikely that such cortical representations can encode bindings between arbitrary roles and entities. Hence, these reduced representations are likely to encode only a restricted set of bindings, and reflect a preference for generic role-fillers such as categories rather than specific role-fillers such as entities. One casualty of this reduced and simpler cortical encoding appears to be *source information* indicating where and how some information was obtained.

An intriguing exception to the above observations about the limitations of cortical representations vis-a-vis episodic representations is the seeming ability of the cortex to encode episodic-memory-like structures in the first 10-15 years of life. While patients with bilateral hippocampal lesions exhibit retrograde amnesia extending several decades prior to the insult, they still seem to retain episodic memories of their childhood (for example, Rempel-Clower et al. 1996). This is perhaps, another example of the rich connectivity and exceptional plasticity of the pre-adolescence brain.

Other researchers have suggested that the HS-based episodic memory traces of events contribute to the fine tuning of cortically based semantic representations (for example, McClelland, McNaughton & O'Reilly 1995; O'Reilly & Rudy 2000). But these models posit that eventually, episodic memory traces also get transferred to cortical circuits. In contrast, the information transfer from the HS to the cortex proposed here is one where the error-sensitive episodic memory trace of a *significant* and *memorable* event continues to persist in the HS — even after the event's episodic memory trace has contributed to the fine-tuning of cortically based causal and semantic structures, and even after a similarity based reduced description of the event may have been formed within cortical circuits.

11 Predictions

SMRITI makes a number of predictions. These include predictions regarding the functional roles of different components of the HS, the properties of relational schemas underlying episodic memories, the sorts of memories that must persist in the HS for the long-term, the nature of memory consolidation, and memory deficits resulting from cell loss in the HS and cortical circuits encoding focal-clusters of relational schemas and entities. Among other things, these predictions help explain differences in the temporal gradient of retrograde amnesia observed in hippocampal and semantic dementia patients.

11.1 Functional role of components of the HS and cortical regions

SMRITI predicts a specific mapping between components of the HS and functional components of an episodic memory trace. This mapping has been described in Section 6.1 and illustrated in Figure 7. Furthermore, SMRITI predicts that local inhibitory circuits serve as critical elements of functional units recruited for encoding episodic memory traces.

SMRITI also predicts that cell ensembles making up focal-clusters of entities and relational schemas are likely to be located in the perirhinal cortex, parahippocampal cortex, and other cortical areas that project directly to EC (refer to Section 2). Some focal-clusters may also be located in EC.

11.2 Arity of frames and schemas subserving episodic memory

The qualitative analysis of SMRITI suggests that a robust memory trace of an event can be formed with practical certainty as long as the arity of the event – that is, the number of bindings required to encode the event — does not exceed six. Beyond this, the probability that an event's memory trace may contain too many ill-formed functional units becomes high.

Restricting the number of bindings in typical event representations to ca. six is not as limiting as it might appear. In general, an arity of five suffices to encode *who* did *what* to *whom*, *where*, *when* and *how* (the "what" selects the appropriate relational frame or schema and does not count toward the number of bindings in the event). Similarly, six bindings suffice to encode an agent, a patient, an instrument, a benefactor, a spatial location, and a temporal location. Moreover, the restriction on the number of bindings in an event resonates with a comparable constraint on the number of role-entity bindings in a dynamic (activity based) encoding of relational information in the mind/brain (Shastri & Ajjanagadde 1993; also see Lisman & Idiart 1995; Jensen & Lisman 1996; Luck & Vogel 1997).

The restriction on the number of bindings in an event suggests that the relational schemas underlying the construal of experiences must be highly differentiated and must make use of pre-programmed parameter values. Thus the schemas (or frames) used in the construal of events are more likely to correspond to specific actions such as *pull*, *push*, and *shove*, and *walk*, *run*, and *crawl*, than to generic actions such as *apply-force* and *change-location*. Note that

expressing an event using a generic schema requires more bindings than expressing the event using a specific schema. For example, encoding a push event using the *apply-force* schema would require bindings to specify, among other things, the direction of force application and the intensity of the applied force. Encoding the same event as an instance of a *push* schema, however, will not require any bindings to specify the direction of force application (away from the body) and the intensity of applied force (moderate) since these values are integral to the bodily grounded "meaning" of push, and hence, would be pre-programmed into the *push* action schema (cf. *push*, *pull*, and *shove*).

11.3 Persistence of specific event memories in the HS

SMRITI suggests that the HS may be essential not only for the acquisition of certain types of memories, but also for their long-term maintenance and retrieval (see Nadel & Moscovitch, (1997) and Wickelgren (1977) for supporting views). This prediction is based on the observation that only the HS appears to have the appropriate convergence of high-level multimodal inputs *and* the appropriate architecture and circuitry to support the formation of *binding-error-detector*, *binding-error integrator*, and *relational-match-indicator* circuits that are essential for the formation of error-sensitive memory traces of relational items. Key representational properties of an "item" that might *jointly* necessitate the participation of the HS in its memorization and long-term maintenance are as follows:

1. The item is best viewed as an *instance* of a generic *relational* schema (or a frame). Such a *relational instance* corresponds to a collection of bindings between roles (or parameters) of a generic relational schema and entities that fill these roles in the given instance.

2. The role-fillers (or parameter values) of the relational instance are not restricted to any particular domain, and almost any entity in the organism's conceptual structure may be chosen as a role-filler.

3. Fillers of multiple roles may belong to the same domain. This entails that the identity of an entity (or its type) alone is not sufficient to determine the role filled by the entity in a relational instance.

4. The item is to be memorized as a specific item *per se*, and distinguished from other items — even similar ones. It is expected that the item's memory trace will not respond to a cue that specifies an incompatible binding — even if the cue is otherwise highly similar to the encoded item.

5. The item's memory trace is expected to support the selective extraction of component role-fillers of an item.

Since attribute-values associated with an exemplar (for example, a person) or a category (for example, dogs) can be viewed as a collection of bindings between attributes and values, is it being suggested above that exemplars and categories are also encoded in the HS? The answer, clearly, is *no*. Note that attribute-value bindings required to encode exemplars and categories typically do not have properties (2) and (3), since the values of a given attribute typically belong to a specific domain, and the values of different attributes typically belong to non-overlapping domains. For example, the values of attributes shape, color, and texture are restricted to the domains of shapes, colors, and textures, respectively, and these domains are non-overlapping. Finally, the representations of categories are typically more sensitive to similarities than to fine distinctions, and hence, their representation does not require property (4). Consequently, the HS is not required for the representation of exemplars and categories.

Although error-sensitive episodic memory traces of memorable events must remain in the HS, as explained in Section 10.2, several types of information *is* transferred from the HS to similarity-based semantic and causal representations in cortical circuits.

11.4 Memory deficits resulting from damage to regions of the HS

SMRITI suggests a number of predictions about the nature of memory deficits that would result from focal insults to various components of the HS. These predictions concern the effect on memories acquired prior to the insult (retrograde effects) as well as on the acquisition of new memories after the insult (anterograde effects). Before discussing specific predictions related to cell loss, let us reiterate that the physically dispersed and redundant nature of functional units makes an episodic memory trace robust against low to moderate levels of cell loss (cf. Section 8.7). Only significant amounts of cell loss lead to memory deficits discussed below.

11.4.1 Retrograde effects

The model predicts that focal damage to different regions of the HS will affect prior memories differently depending on the functional role of the affected region. These predictions are summarized in Table 6. Some additional observations are as follows:

The effect of damage to EC (misses or false-alarm) will depend on which subregions of EC are damaged. But significant cell loss in EC will lead to a catastrophic failure in the recognition and recall of existing memories. EC is one of the first areas to be affected by Alzheimer's disease and the predicted behavior is consistent with the nature of memory deficits observed during the progression of the disease.

By recognition we mean recognizing full-blown events, and not just recognizing whether or not an object had been observed earlier in a certain context. It is important to distinguish between recognition memory for an object and recognition memory for an event (or situation) involving multiple objects playing specific roles in the event (or situation) (Wan, Aggleton & Brown 1999; Gaffan & Parker 1996). While it may be possible for the mind/brain to determine whether or not an object had appeared *recently* in a particular visual context by leveraging purely cortical representations (for example, via short-term synaptic changes within circuits in the inferior temporal cortex), it would not be possible for it to recognize an event and distinguish it from similar events without an intact HF.

In spite of damage to DG, the HS will continue to produce a correct affirmative response to queries such as "Did someone give something to someone?" if *any give* events have been memorized. In general, a memorized event will be recognized correctly in spite of damage to DG, if the query does not specify bindings for *any* of the roles that are bound in the memorized event.[32] Damage to the perforant path fibers between EC and DG, or to mossy fibers between DG and CA3 will have the same affect on the retrieval of prior memories as will damage to principal cells in DG.

If there is significant damage to CA3, the HS will produce erroneous responses to wh-questions and *reinstate role-fillers from other memorized events pertaining to the* **same** *relational schema as the one specified in the query*. Significant damage to perforant path fibers from EC to CA3 will also have a similar effect.

CA1 encodes $remind$ circuits that sit at the apex of an event's memory trace. Hence, a catastrophic loss of cells in CA1 will lead to a catastrophic failure in the functioning of the HS-based episodic memory system.[33] This prediction is consistent with the empirical findings reported by (Rempel-Clower et al. 1996).[34]

Focal damage to the subiculum will render the system unable to respond to wh-queries (Who gave Mary a book?), but its ability to respond to yes-no queries (Did John give Mary a book?) will remain intact.

11.4.2 Anterograde effects

The model predicts that focal damage to different regions of the HS will affect the formation of new memories differently. These predictions are summarized in Table 7. Some additional observations are as follows.

Damage to EC will make it extremely difficult to memorize events involving entities, roles, and relational schemas that are novel, or whose linking cells have been destroyed. Events involving entities, roles, and relational schemas whose linking cells are intact will continue to be memorized normally. Significant damage to EC, however, will render the HS incapable of forming new memories.

The blocking of LTP in DG will also have the same impact as the loss of cells in DG.

Significant damage to CA3 principal cells will render the HS incapable of forming new memories. In contrast, the selective destruction of Type-2 interneurons, or equivalently, the disabling of inhibitory activity, will produce memory traces that will produce false-negative responses to retrieval cues.

The effect of significant damage to CA2 principal cells will be analogous to that of significant damage to CA3 principal cells, but the effect of damage to Type-3 interneurons in CA2 will be opposite to that of damage to Type-2 interneurons in CA3.

Significant damage to CA1 principal cells - even in the absence of damage to any other component of the HS - will be sufficient to render the HS incapable of acquiring new memories. This prediction is validated by empirical findings reported by (Zola-Morgan, Squire & Amaral 1986; Rempel-Clower et al. 1996).

[32] Diffuse activation of subicular cells may occur due to activity arriving from EC. This in turn may activate some EC cells and produce low-confidence false-positive responses in some cases. Nevertheless, the predominant effect of significant damage to DG will be the production of erroneous "don't know" responses.

[33] Diffuse activation of subicular cells may occur via EC and lead to low-confidence false-positive responses in some cases. But the predominant effect of significant damage to CA1 will be the production of erroneous "don't know" responses.

[34] Patients with damage limited to CA1 exhibit profound anterograde amnesia (see Section 11.4.2) and retrograde amnesia for events over an interval of several years prior to the insult.

Region	Effect
EC: significant cell loss	Catastrophic failure of memory.
EC: only linking cells of relational schemas and entities damaged	Erroneous "don't know" (that is, miss) responses. Behaviorally equivalent to forgetting.
EC: only linking cells of roles damaged	Excessive "false-positive" (that is, false-alarm) responses. Behaviorally equivalent to the existence of spurious or illusionary memories.
DG principal cells damaged	Erroneous "don't know" responses. Behaviorally equivalent to forgetting. (CA3 principal cell response becomes promiscuous)
CA3 and/or CA2: principal cells damaged	Excessive false-positive responses. Behaviorally equivalent to the existence of spurious or illusionary memories.
CA3 Type-2 interneurons damaged	Erroneous "don't know" responses. Behaviorally equivalent to forgetting.
CA1 principal cells damaged	Catastrophic memory failure. Erroneous "don't know" responses. Behaviorally equivalent to forgetting.
CA1 Type-2 interneurons damaged	Catastrophic memory failure. Excessive false-positive responses. Behaviorally equivalent to the existence of spurious or illusionary memories.
Sub: Principal cells damaged	Ability to respond to yes-no queries intact, but failure to answer wh-questions.

Table 6: Affect of significant cell loss within specific regions of the HS on memories acquired prior to the insult. See text for details.

11.5 Episodic memory deficits resulting from damage to high-level cortical circuits

Cell loss in cortical regions where focal-clusters of entities and relational schemas are located will destroy cells in the enabler and collector ensembles of relational schemas and entities. Since relational schemas and entities are components of semantic knowledge, the progressive loss of cells in these cortical regions would lead to a condition known as *semantic dementia*[35] (Hodges et al. 1992; Graham, Simons, Pratt, Patterson & Hodges 2000; Graham & Hodges 1997). Since episodic memory traces are grounded in cortically expressed focal-clusters of relational schemas and entities, changes in these cortical representations will also have a distinct impact on the functioning of the HS-based episodic memory system.

In analyzing the effect of changes in cortical circuitry on episodic memory, we will distinguish between two types of items: items that are *in-use* and items that are *out-of-use*. In-use items are relational schemas and entities that have been activated often in the recent past as a result of one's experiences, thoughts, activities, and interactions. Out-of-use items are those relational schemas and entities that have not been activated for a long period of time. Being in-use and out-of-use is a matter of degree, but we will treat this distinction as a categorical one to simplify the following discussion. The generalization to the graded case is straightforward.

11.5.1 Effect of cell loss on in-use items

The process of cell loss will gradually destroy cells in the focal-cluster of an in-use item. However, this loss will be offset by the *incorporation of new cells into the focal-cluster* as a consequence of frequent and recent activation of the

[35] Semantic dementia results in a progressive deterioration of both verbal and non-verbal aspects of semantic knowledge about conceptual entities. Semantic dementia patients have difficulty performing tasks such as picture naming, generating category exemplars, and word-picture matching. Their phonological and syntactic knowledge, episodic memory, and perceptual, visuo-spatial, and non-verbal problem solving abilities, however, remain largely intact. Semantic dementia patients typically exhibit atrophy of one or both temporal lobes, especially, the temporal pole, middle temporal gyrus, and inferior temporal gyrus.

Region	Effect
EC	Inability to form new memories.
DG	Inability to form new memories.
CA3 or CA2 Principal cells	Inability to form new memories
CA3 Type-2 inhibitory interneurons	Memory trace prone to producing *false-negative* responses
CA2 Type-3 inhibitory interneurons	Memory trace prone to producing *false-positive* responses
CA1 principal cells	Catastrophic loss of the ability to form new memories
CA1 Type-2 inhibitory interneurons	Memory trace prone to producing *false-positive* responses
Sub: Principal cells damaged	Acquired memory traces can respond to yes-no queries, but do not encode functional units required to answer wh-questions.

Table 7: Affect of significant cell loss within specific regions of the HS on the formation of new memories. See text for details.

item. The degree to which the loss will be offset in this manner will depend on the rate and duration of cell loss, the item's significance, and how frequently the item is activated.

For moderate levels of cell loss, the incorporation of new cells in the enabler and collector ensembles of an in-use item will be able to maintain the number of cells in these ensembles at levels close to the normal. Hence, a process of cell loss in cortical areas where focal-cluster of relational schemass and entities are located will *not* lead to anterograde amnesia for memories involving in-use items (at least for low to moderate levels of cell loss).

But the incorporation of new cells in the enabler and collector ensembles of an in-use item, together with the loss of existing cells in these ensembles, means that these cortical functional units will not be a stable collection of cells. Instead, the collection of cells forming these ensembles will *progressively change* over time. This progressive morphing of enabler and collector ensembles of in-use items in the cortex has the following interesting consequence: an episodic memory trace formed in the HS will gradually lose its cortical grounding and become inaccessible because the cortical collector and enabler cells at the source of this memory trace and the cortical collector cells at the destination of this memory trace (see Figure 7) will gradually cease to exist. Behaviorally, this will amount to a progressive forgetting (memory decay) process.

In view of the above, it follows that any process that leads to a gradual loss of principal cells in critical cortical areas, but which leaves the HS spared, will lead to a form of amnesia wherein recent memories involving in-use items are remembered well, but older memories are gradually forgotten. In other words, the temporal gradient of this retrograde amnesia will have a positive slope (see Figure 26). Note that the initial loss of ca. 5% in effectiveness of a memory trace may not have any overt behavioral manifestation. Hence an episodic memory involving in-use-items may not exhibit any sign of degradation for some time subsequent to its acquisition.

Thus SMRITI explains why damage to cortical circuits encoding focal-clusters of entities and relational schemas in semantic dementia patients can lead to the type of amnesia wherein new memories can be acquired, recent memories are remembered well, but older memories are forgotten. This is in contrast to the type of retrograde amnesia reported in hippocampal patients who cannot remember recent events, but who seem to exhibit some recall of events from the distant past.

As should be expected, as the process of cell loss in cortically expressed focal-clusters will continue, the rate at which new cells are incorporated into enabler and collector ensembles will keep declining. Eventually, massive cell loss will reduce the number of cells in enabler and collector ensembles to below a critical level and lead to a catastrophic failure in the formation of new memories as well as in the retrieval of existing memories.

11.5.2 Effect of cell loss on out-of-use items

A process of gradual cell loss will lead to a gradual decrease in the (neural) mass of an out-of-use item's collector ensemble. This will lead to a steadily worsening anterograde and retrograde amnesia for memories involving out-of-use items, even though older episodic memory traces involving out-of-use items may continue to exist within the HS.

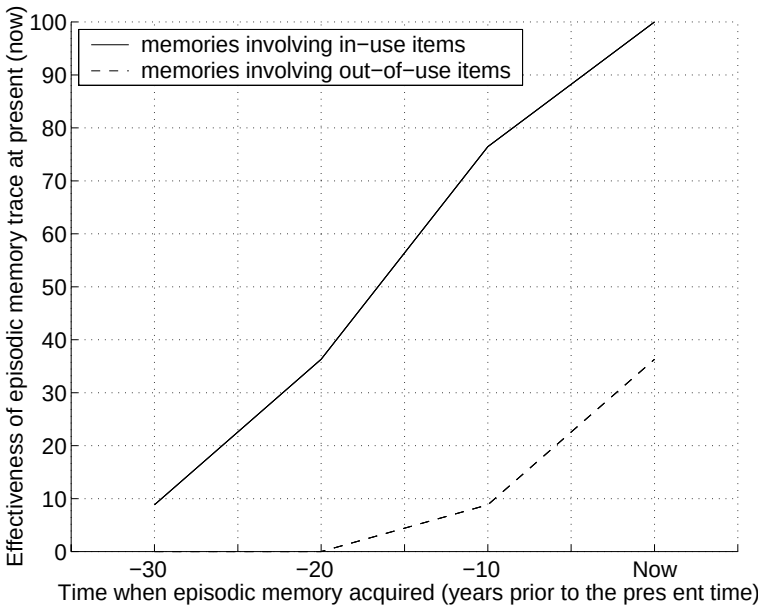

Figure 26: Effect of semantic dementia on two types of episodic memories: those involving in-use and those involving out-of-use items. The acquisition of episodic memories is normal (no anterograde amnesia) for memories involving in-use items, but significantly impaired for memories involving out-of-use items. In both cases, retrograde amnesia exhibits a positively sloped temporal gradient. The analysis assumes that (i) the rate of cortical cell loss is 0.5% per year, (ii) cell loss in focal-clusters of in-use items is compensated by the recruitment of new cells, (iii) such compensation does not occur for out-of-use items, and (iv) for out-of-use items, the accumulated cortical cell loss at the present time is 10%. The results are extrapolated from the quantitative analysis of the model discussed in Section 8. Note that a limited loss in the effectiveness of a memory trace may not have an overt behavioral manifestation. Hence an episodic memory involving in-use-items may not exhibit any sign of degradation for some time after its acquisition.

11.5.3 Episodic memory deficits resulting from damage to inhibitory cortical circuits

A degradation of inhibitory circuits in cortical networks (for example, due to loss of inhibitory interneurons) will lead to increased interference among focal-clusters of similar entities. This will adversely impact the quality of episodic memory traces recruited after the breakdown. Note that as a result of spurious activity in focal-clusters of similar entities, the functional units recruited for a memory trace will not only be driven by cortical cells in the focal-cluster of the correct entities, but also by cortical cells in the focal-clusters of similar entities. Consequently, the memory trace will produce cross-talk during retrieval and activate not only the correct entities, but also entities similar to the correct entities. This will make episodic memory traces highly susceptible to semantically related lures and cause them to produce false-alarm responses to such cues.

11.6 Hippocampal system damage and the acquisition of semantic memory

SMRITI predicts that the availability of episodic memory traces in the HS would greatly facilitate the acquisition and updating of semantic memory. As discussed earlier, the activation of an event's memory trace can reinstate an elaborate and activity-based representation of the event within cortical circuits. Such an activity can trigger synaptic changes within cortical circuits that modify causal and semantic structures to reflect the information contained in the event. Note that as long as an episodic memory trace is not forgotten, it can be reinstated multiple times (for example, during sleep), and this can enable a thorough incorporation of the event's information content into causal and semantic structures.

While the existence of episodic memory traces greatly facilitates the acquisition and updating of semantic knowledge, it is not the case that the HS-based episodic memory system is *essential* for acquiring and updating semantic knowledge. In the cortico-hippocampal interactions envisaged by SMRITI, the construal of an experience as an event is initially expressed as a pattern of activity over focal-clusters and other distributed neural circuits. Though this pattern

of activity is transient, and occurs only once, it can trigger incremental changes in synaptic strengths within cortical circuits. Over time, and with sufficient repetitions, these changes could accumulate and result in qualitative changes in causal and semantic structures.

Thus, while a dysfunctional HS should make it extremely difficult to acquire causal and semantic knowledge, it should still be possible to acquire such knowledge without a functioning HS. This may be especially true, if the insult to the HS occurs at an early age (see Vargha-Khadem et al. 1997) since it may be easier during ontogeny for the brain to develop alternate strategies to compensate for the absence of the HS. One possible compensatory strategy would be to leverage working-memory-like representational mechanisms in the prefrontal cortex to prolong the duration of activity based representations of significant events. Doing so might give each experience more time to affect the representation of causal and semantic knowledge.[36]

12 Discussion

SMRITI demonstrates how a transient pattern of activity representing an event can be transformed rapidly into a persistent and robust memory trace as a result of LTP (and LTD) within structures whose architecture and circuitry resemble those of the HS.

Episodic memory traces formed by SMRITI respond to highly partial cues, and reject similar but erroneous cues. An episodic memory trace acting in concert with cortical circuits encoding semantic, causal, and procedural knowledge can recreate an activation based representation of the memorized event in high-level cortical circuits. The retrieval of memories in SMRITI is rapid and corresponds to a parallel search wherein the query (cue) is matched simultaneously against memory traces encoded in the HS.

At a macro-level of description, SMRITI is similar to other models of hippocampal function that view the HS as a structure for binding together items represented in cortical circuits (for example, see Marr 1971; Halgren 1984; Teyler & DiScenna 1986; Damasio 1989; Squire & Zola-Morgan 1991; Alvarez & Squire 1994; O'Reilly & McClelland 1994; Rudy & Sutherland 1994; Treves & Rolls 1994; Cohen & Eichenbaum 1995; Hasselmo & Stern 1995; McClelland & Goddard 1996; Murre 1996; and Nadel & Moscovitch 1997). The proposed model however, is quite distinct from previous models both in its representational power, and in the functional role it attributes to various components of the HS. In particular, SMRITI contributes to our understanding of episodic memory by:

1. explicating the representational requirements of encoding events and situations,
2. proposing a detailed neural circuit that satisfies these representational requirements, and
3. demonstrating that the propagation of a rhythmic pattern of activity encoding an event within structures whose architecture and local circuitry resemble those of the HS can lead to the rapid and automatic recruitment of the requisite neural circuit within these structures.

The neural circuit required for a representationally adequate encoding of an episodic memory trace is fairly complex. But the complexity of this circuit is well matched by the complexity of the architecture and local circuitry of the HS. Thus SMRITI provides a rationale for various components of the HS and their interactions, and suggests that the idiosyncratic architecture of the HS is tailored to the representational problems it must solve in order to support the episodic memory function.

A number of modelers have proposed that inhibitory interneurons may serve to limit the activity of cells in the HS and have carried out simulations and mathematical analysis of such interactions (for example, Marr 1971; McNaughton & Morris 1987; O'Reilly & McClelland 1994). Other modelers have suggested that inhibitory interneurons may subserve long-distance synchronization of hippocampal activity (Sik, Ylinen, Penttonen & Gyorgy Buzsaki 1994; Traub, Whittington, Stanford & Jefferys 1996). But SMRITI suggests that in addition to limiting the level of excitatory activity and subserving long-range synchronization, local inhibitory circuits serve as critical elements of functional units recruited for encoding episodic memory traces. SMRITI also offers an alternative interpretation of the functional role of hippocampal region CA3; while most models view CA3 as an associative memory (for example, Marr 1971; Treves & Rolls 1994; O'Reilly & McClelland 1994; Hasselmo & Stern 1995; Levy 1996; McClelland & Goddard 1996; Lisman 1999), SMRITI suggests that a key representational role of CA3 in humans might be the detection of binding *errors*.

[36] For an interesting discussion of issues relating to semantic memory in early onset hippocampal patients refer to (Tulving & Markowitsch 1998; Squire & Zola 1998; Mishkin, Vargha-Khadem & Gadian 1998).

Learning in SMRITI is based on LTP and LTD that have emerged as the likely biological mechanisms underlying activity dependent learning in the HS and the cortex. Hippocampal pyramidal cells are known to produce simple spikes as well as spike-bursts. The proper functioning of SMRITI also requires two modes of firing; one analogous to simple spikes, and the other to spike-bursts.

SMRITI predicts that relational schemas underlying the construal of our experience in terms of events have a low arity. This entails that the relational schemas underlying the construal of everyday events are highly differentiated.

SMRITI helps delineate the distinction between semantic and episodic memory, and identifies the sorts of memories that must continue to be encoded in the HS and are not "transferred" to the cortex via a process of consolidation. In particular, it suggests that the HS may be essential, not only for acquiring memories of specific events and situations, but also for the long-term maintenance and accurate retrieval of such memories.

Finally, SMRITI also predicts the nature of encoding and retrieval deficits that would result from significant damage to specific components and pathways of the HS, and from cell loss in cortical circuits encoding semantic knowledge.

12.1 A link between episodic memory and spatial maps

Studies of animals including the monkey (for example, Beason-Held, Rosene, Killiany & Moss 1999; Zola, Squire, Teng, Stefanacci, Buffalo & Clark 2000; Murray & Mishkin 1998) and the rat (for example, O'Keefe & Nadel 1978; McNaughton, Barnes & O'Keefe 1983; Otto & Eichenbaum 1992; Busney & Eichenbaum 1996; O'Keefe & Burgess 1996; Wood, Dudchenko & Eichenbaum 1999; Redish 1999) provide additional evidence for the putative role of the HS. The rat HS has been shown to participate in spatial as well as nonspatial memory tasks. But a key finding is that the rat hippocampus encodes spatial maps of its environs and certain cells in the hippocampus behave as "place cells" that respond maximally when the animal is in a relatively circumscribed spatial region. It is also known that certain cells in the monkey hippocampus respond when the monkey looks at a certain part of space (Rolls, Miyashita, Cahusac, Kesner, Niki, Feigenbaum & Bach 1989) and may be thought of as "spatial view cells" (Rolls 2000). Birds remember where they have stored food and also what they have stored and when. The neural substrate of this avian memory shares embryological and anatomical features with the mammalian hippocampus (Kamil & Balda 1985; Clayton & Dickinson 1998). The HS also plays a role in humans in the navigation of large-scale and well-learned spatial environments (Maguire, Frackowiak & Firth 1997).

As discussed in Section 4.2, an event or a situation can be viewed as an instance of a relational schema. It is also possible to view a spatial map of a specific environment (or context) as a special sort of relational instance by viewing any given environment (or context) as a relation, spatial locations in the given environment as roles, and the occurrence of an object (or landmark) at the location as a binding between the appropriate location and object. Thus a spatial map specifies location-object bindings in a given environment just as an event specifies role-entity bindings in a conceptual schema or frame.[37]

Furthermore, an event can be viewed as a natural generalization of a spatial map. An event specifies bindings not only for spatial locations, but also for conceptual roles. Thus while a spatial map locates objects in a 3-dimensional map and specifies "what is where," an event memory locates an event in a high dimensional conceptual space and specifies "who did what to whom where and when" (see O'Keefe & Nadel 1978; Tolman 1948).

The possibility that episodic memory faculty in humans may be a natural progression of spatial memory faculty in lower mammals is supported by the finding of "place cells" and "spatial view" cells. There seems to be an evolutionary progression in the representational import of hippocampal cells from the egocentric "I am here" in the rat, to the more general allocentric "What is where?" in the monkey, to eventually "Who did what to whom where and when?" in the human.

12.2 Ongoing work and future directions

The current work sheds light on several aspects of episodic memory, and its realization in the mind/brain. But it takes only a small step toward a complete understanding of episodic memory function and its neural basis. A large number of issues remain to be addressed. These include, a detailed understanding of (i) memory consolidation, (ii) forgetting, (iii) the role of sleep in learning and memory, (iv) the role of neurogenesis in learning and memory (Eriksson, Perfilieva, Bjork-Eriksson, Alborn, Nordborg, Peterson & Gage 1998; Gould, Beylin, Tanapat, Reeves & Shors 1999; Shors, Miesegaes, Beylin, Zhao, Rydel & Gould 2001) (v) how memories are organized into rich clusters and sequences, (vi)

[37] A sequence can also be viewed as a special case of a relational instance by viewing each position in the sequence as a role and the occurrence of an item in that position as a binding between the appropriate role and the item.

how working memory and attentional processes influence episodic memory encoding and retrieval, (vii) how aging and pathological processes affect memory function, and (viii) how representational differences (for example, spatial versus verbal) between the left and right hippocampi (for example, Kroll et al. 1996) and along the longitudinal axis of the hippocampus (for example, see Hampson, Simeral & Deadwyler 1999) contribute to hippocampal function.

Another area of research that merits further effort is the establishing of links between specific experimental findings about memory and a detailed circuit-level model of episodic memory such as SMRITI. An example of such a linkage is the fan effect (see Section 8.8). Other examples of experimental findings that seem appropriate candidates for investigating such a linkage are: episodic memory is susceptible to various types of distortions and illusions (Roediger 1996) and hippocampal patients are less prone to produce false-positive responses to semantically similar items than normal subjects (Schacter 1996a).

The ultimate goal of the research reported here is to extend our understanding of what we remember and how we remember. Doing so will help us better understand ourselves and our kind. But it is also hoped that if the proposed research bears fruit, and if technology advances at its current pace, this work might also contribute to the development of a prosthesis for patients who have suffered massive and irreversible damage to their hippocampal region.

Acknowledgment

The research reported in this article was funded by NSF grants 9720398 and 9970890 and ONR grant N00014-93-1-1149. Thanks to Collin Baker, Jerry Feldman, Bob Knight, Bruce Miller, Srini Narayanan, Sadhana Shastri, and Carter Wendelken for stimulating discussions. Certain aspects of this work were presented at the Nineteenth Annual Conference of the Cognitive Science Society and the Ninth Annual Computational Neuroscience Meeting.

References

Abe, K., Inokawa, M., Kashiwagi, A. & Yanagihara, T. (1998) Amnesia after a discrete basal forebrain lesion. *Journal of Neurology, Neurosurgery, and Psychiatry*, 65: 126–130.

Acsady, L., Kamondi, A., Sik, A., Freund, T. & Buzsaki, G. (1998) GABAergic Cells Are the Major Postsynaptic Targets of Mossy Fibers in the Rat Hippocampus. *Journal of Neuroscience* 18: 3386-3403.

Aggleton, J.P. & Brown, M.W. (1999) Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences* 22: 425–489.

Ajjanagadde, V. & Shastri, L. (1991) Rules and variables in neural nets. *Neural Computation* 3: 121–134.

Alvarez, P. & Squire, L.R. (1994) Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences, USA* 91: 7041–7045.

Amaral, D.G. & Cowan, M. (1980) Subcortical afferents to the hippocampal formation in the monkey. *Journal of Comparative Neurology* 189: 573–591..

Amaral, D.G. & Insausti, R. (1990) Hippocampal formation. In *The Human Nervous System* ed. G. Paxinos, Academic Press.

Amaral, D.G., Ishizuka, N. & Claiborne, B. (1990) Neurons, numbers and the hippocampal network. *Progress in Brain Research* 83: 1–11.

Amaral, D.G. & Witter, M.P . (1995) Hippocampal Formation. In *The Rat Nervous System* Second Edition, ed. G. Paxinos, Academic Press.

Anderson, J. R.(1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology* 6: 451-474.

Anderson, J. R. & Reder, L. M. (1999). The Fan Effect: New Results and New Theories. *Journal of Experimental Psychology: General* 128: 186-197.

Arbib, M.A. (1994). *The Metaphorical Brain: 2* Wiley Interscience.

Bailey, D., Chang, N., Feldman, J.A. & Narayanan, S. (1998) Extending Embodied Lexical Development. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, Madison, WI. August, 1998.

Barsalou, L. (1999) Perceptual Symbol Systems. *Behavioral and Brain Sciences* 22: 577-609.

Bartlett, F.C. (1932) *Remembering*, Cambridge University Press.

Beason-Held, L.L., Rosene, D.L., Killiany, R.J. & Moss, M.B. (1999) Hippocampal formation lesions produce memory impairment in the rhesus monkey. *Hippocampus* 9: 562-574.

Bliss. T.V.P. & Lomo, T. (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology* 232: 331–356.

Bontempi, B., Laurent-Demir, C., Destrade, C. & Jaffard, R. (1999) Time-dependent reorganization of brain circuitry underlying long-term memory storage. *Nature* 400: 671 - 675.

Braak, H. (1974) On the Structure of the Human Archicortex. I. The Cornu ammonis. A Golgi and Pigmentarchitectonic Study. *Cell and Tissue Research* 152: 349–383.

Brown, T.H, Kairiss, E.W. & Keenan, C.L. (1990) Hebbian Synapses: Biophysical Mechanisms and Algorithms. *Annual Review of Neuroscience* 13: 475-511.

Buckley, M.J., Gaffan, D. & Murray, E.A. (1997) Functional double dissociation between two inferior temporal cortical areas: perirhinal cortex versus middle temporal gyrus. *Journal of Neurophysiology* 77: 587–598.

Buckmaster, P.S. & Schwartzkroin, P.A. (1995) Interneurons and inhibition in the Dentate Gyrus of the Rat *in vivo*. *Journal of Neuroscience* 15:774–789.

Burnard, Lou. (ed.) 1995. Users' reference guide to the British National Corpus. Oxford: Oxford University Computing Services.

Busney, M. & Eichenbaum, H. (1996) Conservation of hippocampal memory function in rats and humans. *Nature* 379: 255–257.

Buzsaki, G. & Chrobak, J.J. (1995) Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks. *Current Opinion in Biology* 5: 504–510.

Cahill, L., Haier, R.J., Fallon, J., Alkire, M.T., Tang, C. Keator, D., Wu, J. & McGaugh, J.L. (1998) Amygdala activity at encoding correlates with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences, USA* 93: 8016–8021.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J.D.E. & Cahill, L. (2000) Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience*, 20:RC99: 1–5.

Carpenter, G.A. & Grossberg, S. (1993) Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in Neuroscience* 16: 131–136.

Cermak, E. (1984) The episodic/semantic distinction in amnesia. In *The Neuropsychology of Memory*, eds. L.R. Squire and N. Butters, Guilford Press.

Chrobak, J.J. & Buzsaki, G. (1998) Gamma Oscillations in the Entorhinal Cortex of the Freely Behaving Rat. *Journal of Neuroscience* 18: 388-398.

Clayton, N.S. & Dickinson, A. (1998) Episodic-like memory during cache recovery by scrub jays. *Nature* 395: 272–274.

Cohen, N.J. & Eichenbaum, H. (1995) *Memory, Amnesia, and the Hippocampal System*, MIT Press.

Corkin, S. (1968) Acquisition of motor skill after bilateral medial temporal lobe excision. *Neuropsychologia* 6: 255–265.

Corkin, S., Amaral, D.G., Gonzales, R.G., Johnson, K.A. & Hyman, B.T. (1997) H.M.'s medial temporal lobe lesion: findings from magnetic resonance imaging. *Journal of Neuroscience* 17: 3964-3979.

Damasio, A. R. (1989) Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33: 25-62.

Derrick, B.E. & Martinez, J.L. (1996) Associative, bidirectional modifications at the mossy-fibre-CA3 synapse. *Nature*, 381: 429-434.

Diederich, J. & Hogan J.M. (1997) Constructive learning in connectionist semantic networks. In *Neural Networks and a New Artificial Intelligence*, ed. G. Dorffner.

Dolan, R.J. & Fletcher, P.F. (1999) Encoding and retrieval in human medial temporal lobes: an empirical investigation using functional magnetic resonance imaging (fMRI). *Hippocampus* 9: 25–34.

Dudek, S.M. & Bear, M.F. (1992) Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proceedings National Academy of Sciences, USA*89: 4363–4367.

Duvernay, H.M. (1988) *The Human Hippocampus*, J.F. Bergmann.

Eichenbaum, H., Otto, T. & Cohen, N.J. (1994). Two component function of the hippocampal memory system. *Behavioral and Brain Sciences* 17: 449–518.

Eriksson, P.S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A.-M., Nordborg, C., Peterson, D.A. & Gage, F.H. (1998) Neurogenesis in the adult hippocampus. *Nature Medicine* 4: 1313–1317.

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

Fernandez, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dumpelmann, M., Roost, D.V., Elger, C.E. (1999) Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science* 285: 1582–1585.

Fillmore, C.J. (1968) The case for case. In *Universals in Linguistic Theory* eds. E. Bach & R.J. Harms, Holt, Rinehart & Winston.

Freund, T.F. & Buzsaki, G. (1996) Interneurons of the hippocampus. *Hippocampus* 6: 347-470.

Fuster, J.M. (1995) *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*, MIT Press.

Gaffan, D. & Parker, A. (1996) Interaction of Perirhinal Cortex with the Fornix-Fimbria: Memory for Objects and "Object-in-Place" Memory. *Journal of Neuroscience* 16: 5864-5869.

Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996) Action recognition in the premotor cortex. *Brain* 119: 593-609.

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

Gomez-Isla, T., Price, J.L., McKeel, D.W., Morris, J.C., Growdon, J.H. & Hyman, B.T. (1996) Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *Journal of Neuroscience* 16: 4491-4500.

Gould, E., Beylin, A., Tanapat, P., Reeves, A. & Shors, T.J. (1999) Learning enhances adult neurogenesis in the hippocampal formation. *Nature Neuroscience* 2: 260–265.

Graf, P. & Schacter, D.L. (1985) Implicit and explicit memory for new associations in normals and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition* 11: 501–518.

Graham, K.S. & Hodges, J.R. (1997) Differentiating the Roles of the Hippocampal Complex and the Neocortex in Long-Term Memory Storage: Evidence From the Study of Semantic Dementia and Alzheimer's Disease. *Neuropsychology* 11: 77–89.

Graham, K.S., Simons, J.S., Pratt, K.H., Patterson, K. & Hodges, J.R. (2000) Insights from semantic dementia on the relationship between episodic and semantic memory. *Neuropsychologia* 38: 313–324.

Granger, R., Wiebe, S.P., Taketani, M. & Lynch, G. (1996) Distinct memory circuits composing the hippocampal region. *Hippocampus* 6: 567–578.

Halgren, E. (1984) Human hippocampal and amygdala recording and stimulation: evidence for a neural model of recent memory. In *The Neuropsychology of Memory* eds. L.R. Squire & N. Butters, Guilford Press.

Hampson, R.E., Simeral, J.D. & Deadwyler, S.A. (1999) Distribution of spatial and nonspatial information in dorsal hippocampus. *Nature* 402: 610–614.

Harnad, S. (1996) On the virtues of theft over honest toil: Grounding language and thought in sensorimotor categories. Presented at Hang Seng Center Conference on Language and Thought, Sheffield University, June 1996. (http://www/cogsci.soton.ac.uk/ harnad/Papers/Harnad/ harnad96.language.theft.html)

Hasselmo, M.E. & Stern, C.E. (1995) Linking LTP to network function: A simulation of episodic memory in the hippocampal formation. In *Long-Term Potentiation, Volume 3* eds. M. Baudry & J. Davis, MIT Press.

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

Hodges, J.R., Patterson, K., Oxbury, S. & Funnel, E. (1992) Semantic Dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain* 115: 1783–1806.

Horel, J.A. (1988) Limbic neocortical interrelations. In *Comparative Primate Biology Vo. 4. Neurosciences.* ed. H.D. Steklis and J. Erwin, Alan R. Liss.

Huerta, P. & Lisman, J. (1995) Bidirectional Synaptic Plasticity Induced by a Single Burst during Cholinergic Theta Oscillation in CA1 In Vitro. *Neuron* 15, 1053-1063.

Hummel, J.E. & Holyoak, K.J. (1997) Distributed representations of structure: a theory of analogical access and mapping. *Psychological Review* 104: 427-466.

Hyman, B.T. & Van Hoesen, G.W. (1989) Hippocampal and Entorhinal Cortex Cellular Pathology in Alzheimer's Disease. In *The Hippocampus–New Vistas*, eds. V. Chan-Palay & C. Kohler, Alan R. Liss.

Insausti, R., Amaral, D.G. & Cowan, W.M. (1987) The entorhinal cortex of the monkey: II Cortical afferents. *Journal of Comparative Neurology* 264: 356–395.

Insausti, R., Tunon, T, Sobreviela, T., Insausti, A.M. & Gonzalo, L.M. (1995) The human entorhinal cortex: a cytoarchitectonic analysis. *Journal of Comparative Neurology* 355: 171–98.

Ivanco, T.L. & Racine, R.J. (2000) Long-term potentiation in the reciprocal corticohippocampal and corticocortical pathways on the chronically implanted, freely moving rat. *Hippocampus* 10: 143–152.

Jackendoff, R. (1990) *Semantic Structures*, MIT Press.

Jackson, M.B. & Scharfman, H.E. (1996) Positive feedback from hilar mossy cells to granule cells in the dentate gyrus revealed by voltage-sensitive dye and microelectrode recordings. *Journal of Neurophysiology* 76: 601–616.

Jensen, O. & Lisman, J. E. (1996) Novel lists of 7 ± 2 known items can be reliably stored in an oscillatory short-term memory network: Interaction with long term memory. *Learning and Memory* 3: 257-263.

Johnson, M.K. & Chalfonte, B.L. (1994) Binding complex memories: The role of reactivation and the hippocampus. In *Memory Systems 1994*, eds. D.L. Schacter & E. Tulving, MIT Press

Johnston, D. & Amaral, D. (1998) Hippocampus. In *The Synaptic organization of the brain*. ed. G.M. Shephard. Oxford University Press.

Kamil, A.C. & Balda, R.P. (1985) Cache recovery and spatial memory in Clark's nutcrackers (*Nucifraga columbiana*). *Journal of Experimental Psychology Animal Behavior and Processes* 11: 95-111.

Kartsounis, L.D., Rudge, P. & Stevens, J.M. (1995) Bilateral lesions of CA1 and CA2 fields of the hippocampus are sufficient to cause a severe amnesic syndrome in humans. *Journal of Neurology, Neurosurgery, and Psychiatry* 59: 95-98.

Klimesch, W. (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews* 29: 169-195.

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

Knowlton, B.J. & Squire, L.R. (1993) The learning of categories: parallel brain systems for item memory and category knowledge. *Science* 262: 1747–9.

Kroll, N.E.A., Knight, R.T., Metcalfe, J., Wolf, E.S. & Tulving, E. (1996) Cohesion failure as a source of memory illusions. *Journal of Memory and Language* 35: 176–196.

Lacaille, J.C., Kunkel, D.D. & Schwartzkroin, P.A. (1989) Electrophysiological and morphological characterization of hippocampal interneurons. In *The Hippocampus–New Vistas*, eds. V. Chan-Palay & C. Kohler, Alan R. Liss.

Lakoff, G. & Johnson, M. (1999) *Philosophy in the Flesh: The embodied mind and its challenge to Western thought*, Basic Books.

Langacker, R. (1986) *Foundations of Cognitive Grammar* Volume 1, Stanford University Press.

Laroche, S., Davis, S. & Jay, T.M. (2000) Plasticity at hippocampal to prefrontal cortex synapses: Dual roles in working memory and consolidation. *Hippocampus* 10: 438–446.

Larson, J., Wong, D. & Lynch, G. (1986) Patterned stimulation at the theta-frequency is optimal for the induction of hippocampal long-term potentiation. *Brain Research* 368, 347-350.

Lavenex, P. & Amaral, D.G. (2000) Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus* 10: 420–430.

Lepage, M., Habib, R. & Tulving, E. (1998) Hippocampal PET activation of memory encoding and retrieval: the HIPER model. *Hippocampus* 8: 313–322.

Levy, W.B. (1996) A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus* 6: 579–590.

Levy, W.B. & Steward, O. (1979) Synapses as associative memory elements in the hippocampal formation. *Brain Research* 175: 233-245.

Linden, D.J. (1994). Long-term synaptic depression in the mamallian brain. *Neuron 12*: 457–472.

Lisman, J. (1999) Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron* 22: 233-242.

Lisman, J. & Idiart, M.A.P. (1995) Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science* 267: 1512–1515.

Louie, K. & Wilson, M.A. (2001) Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* 29: 145–156.

Luck, S. J. & Vogel, E. K., (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390: 279-281.

Lynch, G., Dunwiddie, T. & Gribkoff, V. (1977) Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature* 266: 737-739.

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

MacWhinney, B. (1999) The emergence of language from embodiment. In *The emergence of language*, ed. B. MacWhinney, Lawrence Erlbaum.

Maddock, R.J. (1999) The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends in Neuroscience* 22: 310–316.

Maguire, E.A., Frackowiak, R.S.J. & Firth, C.D.(1997) Recalling routes around London: Activation of the right hippocampus in taxi driver. *Journal of Neuroscience* 17: 7103–7110.

Malenka, R.C. & Nicoll, R.A. (1999) Long-term Potentiation - A Decade of Progress? *Nature* 285: 1870-1874.

Maquet, P., Pters, J-M., Aerts, J., Delfiore, G., Degueldre, C., Luxen, A. & Franck, G. (1996) Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature* 383: 163-166.

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

McClelland, J.L & Goddard, N.H. (1996) Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus* 6: 654–665.

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: 419-57.

McGaugh, J.L. (2000) Memory – a century of consolidation. *Science* 287: 248–251.

McNaughton, B.L., Barnes, C.A. & O'Keefe, J. (1983). The contributions of position, direction, and velocity to single unit activity in the hippocampus. *Experimental Brain Research* 52: 41–49.

McNaughton, B.L., Douglas, R.M. & Goddard, G.V. (1978). Synaptic enhancement in fascia dentata: Cooperativity among coactive afferents. *Brain Research* 157: 277-293.

McNaughton, B.L. & Morris, R.G.M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neuroscience* 10: 408-415.

Menschik, E.D. & Finkel, L.H. (1998) Neuromodulatory control of hippocampal function: Toward a model of Alzheimer's disease. *Artificial Intelligence in Medicine* 13: 99-121.

Metcalfe, J., (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: implications for Korsakoff amnesia. *Psychological Review* 100: 3-22.

Meunier, M., Bachevalier, J., Mishkin, M. & Murray, E.A. (1993) Effects of visual recognition of combined and separate abalations of the entorhinal and perirhinal cortex in rhesus monkey. *Journal of Neuroscience* 13: 5418-5432.

Mishkin, M. & Petri, H.L. (1984) Memories and habits: some implications for the analysis of learning and retention. In *The Neuropsychology of Memory* eds. L.R. Squire & N. Butters, Guilford Press.

Mishkin, M., Vargha-Khadem, F. & Gadian, D.G. (1998) Amnesia and the organization of the hippocampal system. *Hippocampus* 8: 212–216.

Moll, M. & Miikkulainen, R. (1997) Convergence-Zone Episodic Memory: Analysis and Simulations. *Neural Networks* 10: 1017-1036.

Morris, R.G.M. & Frey, U. (1997) Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience? *Philosophical Transactions of the Royal Society of London* 352: 1489–1503.

Moscovitch, M. (1994) Memory and working-with-memory: evaluations of a component process model and comparisons with other models. In *Memory Systems 1994*, eds. D.L. Schacter & E. Tulving, MIT Press.

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

Murray, E.A. & Bussey, T.J. (1999) Perceptual-mnemonic functions of the perirhinal cortex. *Trends in Cognitive Science* 3: 142–151.

Murray, E.A. & Bussey, T.J. (2001) Consolidation and the medial temporal lobe revisited: Methodological considerations. *Hippocampus* 11: 1-7.

Murray, E.A. & Mishkin, M. (1998) Object Recognition and Location Memory in Monkeys with Excitotoxic Lesions of the Amygdala and Hippocampus. *Journal of Neuroscience* 18: 6568-6582.

Nadel, L. & Moscovitch, M. (1997) Memory Consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology* 7: 217–227.

Nadel, L., Samsonovich, A., Ryan, L. & Moscovitch, M. (2000) Multiple trace theory of human memory: Computational, neuroimaging, and neuropsychological results. *Hippocampus* 10: 352–368.

Narayanan, S. (1997) Talking The Talk Is Like Walking the Walk: A Computational Model of Verbal Aspect. *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society*, Stanford, CA.

Neisser, U. (1967) *Cognitive Psychology*, Appleton-Century-Crofts.

Nicoll, R.A. & Malenka, R.C. (1995) Contrasting properties of two forms of long-term potentiation in the hippocampus. *Nature* 377: 115–118.

Nyberg, L., Habib, R., McIntosh, A.R. & Tulving, E. (2000) Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences, USA* 97: 11120–11124.

O'Keefe, J. & Burgess, N. (1996) Geometric determinants of the place fields of hippocampal neurons. *Nature* 381: 425–428.

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

Olbrich, H.G. & Braak, H. (1985) Ratio of pyramidal cells versus non-pyramidal cells in sector CA1 of the human Ammon's horn. *Anatomy and Embryology* 173: 105-110.

Olton, D.S., Becker, J.T. & Handelmann, G.E. (1979) Hippocampus, space, and memory. *Behavioral and Brain Sciences* 2: 313–365.

O'Reilly, R.C. & McClelland, J.L. (1994) Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off, *Hippocampus* 4: 661-682.

O'Reilly, R.C. & Rudy, J.. (2000) Computational principles of learning in the neocortex and hippocampus. *Hippocampus* 10: 389–397.

Otto, T. & Eichenbaum, H. (1992) Dissociable roles of the hippocampus and orbitofrontal cortex in an odour-guided delayed nonmatch to sample task. *Behavioral Neuroscience* 105: 111–119.

Page, M. (2000) Connectionist Modelling in Psychology: A Localist Manifesto *Behavioral and Brain Sciences* 23: 443–467.

Pavlides, C., Greenstein, Y.J., Grudman, M. & Winson, J. (1988) Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of θ-rhythm. *Brain Research* 439, 383—387.

Penttonen, M., Kamondi, A., Sik, A., Acsady, L. & Buzsaki, G. (1997) Feed-forward and feedback activation of the dentate gyrus in vivo during dentate spikes and sharp wave bursts. *Hippocampus* 7: 437–450.

Pinker, S. (1989) *Learnability and Cognition: The Acquisition of Argument Structure*, The MIT Press.

Qin, Y.L., McNaughton, B.L., Skaggs, W.E. & Barnes, B.A. (1997) Memory reprocessing in corticocortical and hippocampocortical neural ensembles. *Philosophical Transactions of the Royal Society of London, Series B,* 352: 1525–1553.

Radvansky, G.A. (1999) The fan effect: A tale of two theories. *Journal of Experimental Psychology: General* 128: 198–206..

Radvansky, G.A., Spieler, D.H. & Zacks, R.T. (1993) Mental model organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19: 95–114.

Rawlins, J.N.P. (1985) Associations across time: The hippocampus as a temporary memory store. *Behavioral and Brain Sciences* 8: 479-496.

Redish, A.D. (1999) *Beyond the cognitive map: From place cells to episodic memory*, MIT Press.

Rempel-Clower, N.L., Zola, S.M., Squire, L.R. & Amaral, D.G. (1996) Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. *Journal of Neuroscience* 16: 5233–5255.

Rioult-Pedotti, M.-S., Friedman, D. & Donoghue, J.P. (2000) Learning-induced LTP in Neocortex. *Science* 290: 533-536.

Rizzolatti, G. & Arbib, M.A. (1998) Language within our grasp. *Trends in Neurosciences* 21: 188–194.

Roediger, H.L. (1996) Memory illusions. *Journal of Memory and Language* 35: 76-100.

Rolls E.T. (2000) Hippocampo-cortical and cortico-cortical backprojections. *Hippocampus* 10: 380-388.

Rolls, E.T., Miyashita, Y., Cahusac, P.M.B., Kesner, R.P., Niki, H., Feigenbaum, J. & Bach, L. (1989) Hippocampal neurons in the monkey with activity related to the place in which a stimulus is. *Journal of Neuroscience* 9: 1835-1845.

Rudy, J.W. & Sutherland, R.J. (1994) The memory coherence problem, configural association, and the hippocampal system. In *Memory Systems 1994*, eds. D.L. Schacter & E. Tulving, MIT Press

Schacter, D.L. (1996a) Illusory memories: A cognitive neuroscience analysis, *Proceedings of the National Academy of Sciences, USA* 93: 13527–13533.

Schacter, D.L. (1996b) *Searching for memory: The Brain, the Mind, and the Past*, Basic Books.

Schacter, D.L. & Wagner, A.D. (1999) Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus* 9: 7–24.

Scharfman, H.E. (1994) EPSPs of dentate gyrus granule cells during epileptiform bursts of dentate hilar "mossy" cells and area CA3 pyramidal cells in disinhibited rat hippocampal slices. *Journal of Neuroscience* 14: 6041–6057.

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

Schwartzkroin, P.A., Scharfman, H.E. & Sloviter, R.S. (1990) Similarities in circuitry between Ammon's horn and dentate gyrus: local interactions and parallel processing. *Progress in Brain Research* 83: 269–286.

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

Shastri, L. (1988) A connectionist approach to knowledge representation and limited inference. *Cognitive Science* 12: 331–392.

Shastri, L. (1997) A Model of Rapid Memory Formation in the Hippocampal System, In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, 680-685, Stanford University, CA. August 1997.

Shastri, L. (1999a) Advances in Shruti - A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence* 11: 79–108.

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

Shastri, L. (2000) Types and Quantifiers in Shruti – a connectionist model of rapid reasoning and relational processing, *Hybrid Neural Symbolic Integration*, eds. S. Wermter & R. Sun, Springer-Verlag.

Shastri, L. (2001a) Biological Grounding of Recruitment Learning and Vicinal Algorithms in Long-term Potentiation. *Emergent neural computational architectures based on neuroscience*, eds. J. Austin, S. Wermter & D. Wilshaw, Springer-Verlag. (In Press).

Shastri, L. (2001b) Episodic memory trace formation in the hippocampal system: a model of cortico-hippocampal interaction. Technical Report 01-004, International Computer Science Institute, Berkeley, CA.

Shastri, L. & Ajjanagadde V. (1993) From simple associations to systematic reasoning: A connectionist encoding of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences* 16: 417-494.

Shastri, L., Grannes, D., Narayanan, S., and Feldman, J. (2001) A Connectionist Encoding of Parameterized Schemas and Reactive Plans. In *Hybrid Information Processing in Adaptive Autonomous vehicles*, eds. G.K. Kraetzschmar & G. Palm, Springer-Verlag. In Press.

Shastri, L. and Wendelken, C. (2000) Seeking coherent explanations – a fusion of structured connectionism, temporal synchrony, and evidential reasoning. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society,* August, 2000, Philadelphia, PA., pp 453-458.

Shen, B. & McNaughton, B.L. (1996) Modeling the spontaneous reactivation of experience-specific hippocampal cell assembles during sleep. *Hippocampus* 6: 685–692.

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. *Quarterly Journal of Experimental Psychology* 38A: 619–644.

Shimizu, E., Tang, Y-P., Rampon, C. & Tsien, J.Z. (2000) NMDA Receptor-Dependent Synaptic Reinforcement as a Crucial Process of Memory Consolidation. *Science* 290: 1170–1174.

Shors, T.J. & Matzel, L.D. (1997) Long-term Potentiation: What's Learning Got to Do with It? *Behavioral and Brain Sciences* 20: 597–614.

Shors, T.J., Miesegaes, G., Beylin, A., Zhao, M., Rydel, T. & Gould, E. (2001) Neurogenesis in the adult is involved in the formation of trace memories. *Nature* 410: 372–376.

Sik, A., Ylinen, A., Penttonen, M. & Buzsaki (1994) Inhibitory CA1-CA3-Hilar Region Feedback in the Hippocampus. *Science* 265: 1722–1724.

Simons, J.S., Graham K.S & Hodges, J.R. (1999) What does semantic dementia reveal about the functional role of the perirhinal cortex? (with response by Bussey, T.J. & Murray, E.A.) *Trends in Cognitive Science* 3: 248–250.

Singer, W. (1993) Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology* 55: 349-74.

Skaggs, W.E. & B. L. McNaughton (1996) Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271: 1870–1873.

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.

Squire, L.R. & Zola-Morgan, S. (1998) Episodic memory, semantic memory and amnesia. *Hippocampus* 8: 205–211.

Staubli, U. & Lynch, G. (1987) Stable hippocampal long-term potentiation elicited by 'theta' pattern stimulation. *Brain Research* 435: 227–234.

Stefanacci, L., Buffalo, E.A., Schmolck, H. & Squire, L.R. (2000) Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neurophysiological profile of patient E.P. *Journal of Neuroscience* 20: 7024–7036.

Stickgold, R. (1998) Sleep: off-line memory reprocessing. *Trends in Cognitive Science* 2: 484-492.

Suzuki, W.A. & Amaral, D.G. (1994) Perirhinal and parahippocampal cortices of the macaque monkey: Cortical Afferents. *The Journal of Comparative Neurology* 350: 497–533.

Tang, Y., Shimizu, E., Dube, G.R., Rampon, C., Kerchner, G.A., Zhuo, M., Liu, G. & Tsien, J.Z. (1999) Genetic enhancement of learning and memory in mice. *Nature* 401: 63–69.

Teng, E. & Squire, L.R. (1999) Memory for places learned long ago is intact after hippocampal damage. *Nature* 400: 675–677.

Tesche, C.D. & Karhu, J. (2000) Theta oscillations index human hippocampal activation during a working memory task. *Proceedings of the National Academy of Sciences, USA* 97: 919–924.

Teyler, T.J. & DiScenna P. (1986) The Hippocampal Memory Indexing Theory. *Behavioral Neuroscience* 100: 147–154.

Tolman, E.C. (1948) Cognitive maps in rats and men. *Psychological Review*: LV: 189–208.

Traub, R.D., Whittington, M.A., Stanford, I.M. & Jefferys, J.G.R. (1996) A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature* 382: 621–624.

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

Treves, A., Skaggs, W.E. & Barnes, C.A. (1996) How much of the hippocampus can be explained by functional constraints? *Hippocampus* 6: 666–674.

Tulving, E. (1983) *Elements of Episodic Memory*, Oxford University Press.

Tulving, E. (1995) Organization of Memory: Quo Vadis? In *The Cognitive Neuroscience*, ed. M.S. Gazzaniga, MIT Press.

Tulving, E. & Markowitsch, H.J. (1998) Episodic and declarative memory: role of the hippocampus. *Hippocampus* 8: 198–204.

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

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

Van Hoesen, G.W. & Hyman, B.T. (1990) Hippocampal formation: anatomy and patterns of pathology in Alzheimer's disease. *Progress in Brain Research* 83: 445–457.

Vargha-Khadem, F., Gadian, 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: 376-380.

Vertes, R.P. & Eastman, K.E. (2000) The case against memory consolidation in REM sleep. *Behavioral and Brain Sciences* 23. In Press.

Vertes, R.P. & Kocsis, B. (1997) Brainstem-diencephalo-septohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience* 81: 893-926.

von der Malsburg, C. (1986) Am I thinking assemblies? In *Brain Theory*, eds. G. Palm & A. Aertsen. Springer-Verlag.

Wan, H., Aggleton, J.P. & Brown, M.W. (1999) Different contributions of the hippocampus and perirhinal cortex to recognition memory. *Journal of Neuroscience* 19: 1142-1148.

Wendelken, C. & Shastri, L. (2000) Probabilistic learning and inference in a connectionist causal network. In *Proceedings of Neural Computation 2000*, Berlin, May 2000.

West, M.J. (1990) Stereological studies of the hippocampus: a comparison of the hippocampal subdivisions of diverse species including hedgehogs, laboratory rodents, wild mice and men. *Progress in Brain Research* 83: 13-36.

West, M.J. & Slomianka, L. (1998a) Total numbers of neurons in the layers of the human entorhinal cortex. *Hippocampus* 8: 69–82.

West, M.J. & Slomianka, L. (1998b) Corrigendum West MJ, Slomianka L. Total numbers of neurons in the layers of the human entorhinal cortex. Hippocampus 8:69–82. *Hippocampus* 8: 426.

Wheeler, M.E., Petersen, S.E. & Buckner, R.L., (2000) Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, USA* 97: 11125–11139.

Wickelgren, W.A. (1977) *Learning and Memory*, Prentice-Hall.

Wickelgren, W.A. (1979) Chunking and consolidation: A theoretical synthesis of semantic networks configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review* 86: 44-60.

Wilson, M.A. & McNaughton, B.L. (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* 265: 676–679.

Winson, J.(1985) *Brain and Psyche: The biology of the unconscious*, Doubleday/Anchor Press.

Witter, M.P., Naber, P.A., van Haeften, T., Machielsen, W.C.M., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P. & Lopes da Silva, F.H. (2000) Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus* 10: 398–410.

Wood, E.R., Dudchenko, P.A. & Eichenbaum, H. (1999) The global record of memory in hippocampal neuronal activity. *Nature* 397: 613 - 616.

Yonelinas, A. P. (1997) Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition* 25: 747-763.

Zola-Morgan, S., Squire, L.R. & Amaral, D.G. (1986) Human Amnesia and the Medial Temporal Region: Enduring Memory Impairment Following Bilateral Lesion Limited to Field CA1 of the Hippocampus. *Journal of Neuroscience* 6: 2950–2967.

Zola, S.M., Squire, L.R., Teng, E., Stefanacci, L., Buffalo, E.A. & Clark, R.E. (2000) Impaired Recognition Memory in Monkeys after Damage Limited to the Hippocampal Region. *Journal of Neuroscience* 20: 451-463.