

The Temporal Context Model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains

Marc W. Howard, Mrigankka S. Fotedar,
Aditya V. Datey
Department of Psychology
Syracuse University

Michael E. Hasselmo
Department of Psychology, Program in Neuroscience, and
Center for Memory and the Brain
Boston University

Accepted March 23, 2004

The medial temporal lobe (MTL) has been studied extensively at all levels of analysis, yet its function remains unclear. Theory regarding the cognitive function of the MTL has centered along 3 themes. Different authors have emphasized the role of the MTL in episodic recall, spatial navigation, or relational memory. Starting with the temporal context model (M. W. Howard and M. J. Kahana, 2002), a distributed memory model that has been applied to benchmark data from episodic recall tasks, the authors propose that the entorhinal cortex supports a gradually changing representation of temporal context and the hippocampus proper enables retrieval of these contextual states. Simulation studies show this hypothesis explains the firing of place cells in the entorhinal cortex and the behavioral effects of hippocampal lesion in relational memory tasks. These results constitute a first step towards a unified computational theory of MTL function that integrates neurophysiological, neuropsychological and cognitive findings.

The medial temporal lobe (MTL) is a region that includes the hippocampus proper, the subicular complex and parahippocampal cortical regions, including entorhinal, perirhinal, and parahippocampal/postrhinal cortices. A great deal of data from neuropsychology (e.g. Eichenbaum & Cohen, 2001; Scoville & Milner, 1957; Squire, 1992) and functional imaging (e.g. Fernandez, Effern, Grunwald, et al., 1999; Stern, Corkin, Gonzalez, et al., 1996; Wagner et al., 1998) has converged on the idea that the MTL is important in learning and memory. In order to bridge the gap between cognition and cellular-level physiology, we need a mechanistic, mesoscopic description of MTL computational function. We already have several successful verbally-formulated theories

of the cognitive function of the MTL. These are described in turn in the following subsections. This paper will attempt to draw these multiple verbal theories together into a single computational framework that is consistent with known neurophysiological and neuroanatomical data.

Episodic recall

All of the diverse skills and facts that differentiate an adult from an infant must be some form of memory—we say that one learns to ride a bike, or remembers the alphabet. In the early part of this century, memory theory strove to describe general laws that would presumably apply to all these different types of learning (e.g. Estes, 1950; Osgood, 1949). Recent decades have seen this unitary approach to memory fragment into the categorization of multiple types of memory, typically with separable neural substrates for each (e.g. Eichenbaum & Cohen, 2001; Nadel & Moscovitch, 1997; Tulving & Schacter, 1991). One of the most fruitful of these distinctions has been that of episodic memory.

Episodic memory refers to the ability to remember specific events from one's personal experience (Tulving, 1983, 2002). For instance, one might have an episodic memory of having eaten a banana at breakfast. The memory for this episode, perhaps with details about the other objects and people present at breakfast, with the taste of the banana, the sounds and smells that were present in the room, is in principle quite distinct from other types of memory one might have for bananas. For instance, one could remember many things about bananas—that they are yellow, that they are good to eat, that people like to eat them at breakfast—without memory for any specific experience with a banana. Recent

Michael Kahana developed the first versions of TCM and has collaborated closely on its development as a model of episodic recall. He also made numerous helpful comments on an earlier draft of the ms. Thanks are due to Loren Frank for generously sharing and patiently explaining positional and derived data from the W-maze, and to Neil Burgess and Colin Lever for sharing positional data from the open field. The early development of the ideas about the place code benefited from discussions with Larry Abbott and John Lisman. Supported by Conte Center grant NIMH P50 MH60450 (principal investigator Joseph Coyle), R01 MH61492 and MH60013 to MEH, 2-R01MH55687 (principal investigator Michael Kahana), F32 MH65841 to MWH, and the College of Arts and Sciences of Syracuse University. Correspondence concerning this article should be addressed to marc@memory.syr.edu. Marc Howard, Syracuse University, Department of Psychology, 430 Huntington Hall, Syracuse, NY 13244-2340.

work has argued that episodic memory relies on the MTL, in particular the hippocampus (Nadel & Moscovitch, 1997; O'Keefe & Nadel, 1978; Tulving & Markowitsch, 1998).

A number of behavioral tasks test episodic memory. For example, in the free recall task, the subject is presented with a list of stimuli, typically words. The task is to recall as many words as possible from the list, with the subject free to determine the order of recall. Free recall is an episodic task in that performance requires that the subject recall the words presented in a particular episodic setting. Free recall is sufficiently sensitive to MTL damage that it can be used as a diagnostic tool for MTL damage in clinical settings (Graf, Squire, & Mandler, 1984).

Spatial navigation

O'Keefe and Nadel (1978) proposed that the primary function of the hippocampus is to construct and read out "cognitive maps." In the following years, however, this theoretical approach has focused on the role of the hippocampus and related structures in learning and navigating through spatial environments. The most remarkable piece of evidence supporting this view is the existence of place cells (O'Keefe & Dostrovsky, 1971). Pyramidal cells within the hippocampus, recorded from rats moving throughout an environment, fire selectively when the animal is in one particular region of the environment. In open environments, this doesn't depend on the direction the animal is facing (Muller, Bostock, Taube, & Kubie, 1994), and firing persists in the dark (Quirk, Muller, & Kubie, 1990), ruling out an explanation based on simple visual stimuli correlated with place.

There is an extensive literature describing characteristics of place cells in dorsal CA1 (e.g. Muller & Kubie, 1987; O'Keefe & Burgess, 1996; O'Keefe & Dostrovsky, 1971; Wilson & McNaughton, 1993). Less is known about the place code in other MTL structures. It is known that there are place cells in the entorhinal cortex (EC, Barnes, McNaughton, Mizumori, Leonard, & Lin, 1990; Frank, Brown, & Wilson, 2000; Quirk, Muller, Kubie, & Ranck, 1992), a region of cortex that provides input to the hippocampus proper. The place response in EC differs in some respects from the place code observed in CA1, indicating that the hippocampus performs significant computations on the incoming place representation. Nonetheless, it is clear that we can't have a meaningful understanding of the function of the hippocampus until we have a correct understanding of the nature of the entorhinal place code.

Transitive associations and relational memory

Data from olfactory learning in the rat (Bunsey & Eichenbaum, 1996; Dusek & Eichenbaum, 1997) has been used to argue that the hippocampus, the central structure of the MTL, enables transitive associations, a function believed to be important in relational memory. In these experiments, rats learned associations or relationships between arbitrary stimuli. For instance, in the study of Bunsey and Eichenbaum (1996), rats with hippocampal lesions were able to learn associations between odors *A* and *B*, and between *B* and *C*.

Unlike normal rats, however, lesioned rats did not show a transitive generalization for the association $A \rightarrow C$. Although the lesioned animals were able to learn simple associations between the stimuli, Bunsey and Eichenbaum (1996) argued that they did not learn the relationships among stimuli that weren't presented together (see also Dusek & Eichenbaum, 1997).

The mnemonic deficit exhibited by hippocampal-lesioned animals cannot apparently be described as a deficit in the development of simple stimulus-response associations. However, when complex relationships between stimuli must be learned, the MTL, and the hippocampus in particular, appear to be critically involved. This emphasis on relational memory is not at all contradictory to a role for the MTL in episodic memory. After all, memory for an episode involves drawing together the many different stimuli present within the episode, in a unique configuration.

Toward a unified framework

These three theoretical approaches to MTL function, episodic recall, spatial navigation and relational memory, are not mutually contradictory. As mentioned previously, memory for an episode should include memory for the configuration of stimuli present in that episode. Similarly, O'Keefe and Nadel (1978) pointed out that a cognitive map could be used to encode the relationships between non-spatial sets of stimuli, resulting in binding items to a temporal-spatial context, supporting episodic memory (ch 14 O'Keefe & Nadel, 1978). Because the neurobiology of the MTL is such an intensely studied subject, there is a tremendous incentive to construct a model that can address questions from all three domains.

The goal of the present paper is to present the beginnings of a theoretical framework that begins to draw together these three disparate approaches. This will be accomplished within the structure provided by the Temporal Context Model (TCM, Howard & Kahana, 2002a), developed to explain experimental findings from free recall, an episodic recall task. TCM describes a set of rules that govern the behavior of a distributed representation of temporal context. We will show that the equation governing contextual drift, taken as a model of temporal-spatial context, can explain the primary features of the entorhinal place code, a phenomenon central to the MTL's support for spatial function. We will then demonstrate that the equation governing retrieved temporal context, a kind of plasticity postulated to explain properties of episodic association, can support a more general function in extracting the temporal structure of experience. This provides a framework for modeling the dissociation between relational learning and simple pairwise association.

Recency and Contiguity: TCM and Fundamental Properties of Episodic Recall

TCM was developed to describe two fundamental properties of episodic memory. The recency effect (Bjork & Whitten, 1974; Howard & Kahana, 1999; Murdock, 1963b; Rat-

cliff & Murdock, 1976) is the tendency for more recent items to be recalled better than less recent items. Associative effects (Howard & Kahana, 1999, 2002b; Kahana, 1996) describe the development of episodically-formed connections between items. This section will first review prior work on TCM, describing the structure and reasoning behind the model. Following this, we will describe a linking hypothesis between TCM and the brain, with a special emphasis on the medial temporal lobe.

An episodic representation

Context, in one form or another, has long been an important component of models of episodic memory performance (e.g. Anderson & Bower, 1972; Raaijmakers & Shiffrin, 1980; Mensink & Raaijmakers, 1988; Yntema & Trask, 1963). The basic approach of TCM has been to take a particular formulation of context, referred to as temporal context and use it as the sole cue for recall of item representations. Because context changes gradually over time TCM can predict forgetting over long time scales. Unlike some prior formulations (e.g. Mensink & Raaijmakers, 1988), however, TCM also explicitly models context that changes gradually *within* a list of items. This assumption enables a description of recency effects within lists, an effect which has often been attributed to short-term memory (e.g. Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1980). The most radical point of departure of TCM from prior models of episodic recall, however, is the assumption that context serves as the sole cue for episodic recall. In TCM, observed episodic associations between items are a consequence of effects items have on context, eliminating the need for direct item-to-item associations in describing episodically-formed associations. We will describe TCM in more detail in the following subsections. This treatment reviews prior work (Howard & Kahana, 2002a; Howard, Wingfield, & Kahana, In revision; Howard, 2004). Readers already familiar with TCM as a model of episodic recall may wish to advance to the subsection entitled “A mapping between TCM and the MTL.”

Temporal context and items. The central assumption of TCM is that there is a distinction between temporal context and to-be-recalled items. The current state of temporal context at time step i is referred to as \mathbf{t}_i . We assume that \mathbf{t}_i is a vector in a high-dimensional space; typically an infinite-dimensional space for simplicity. The item presented at time step i is referred to as \mathbf{f}_i . We assume that the item representations \mathbf{f} are vectors in a separate high-dimensional space, typically infinite for simplicity. We assume that item representations do not change over the course of a typical recall experiment and that they are orthonormal. That is, we assume that there is no overlap between item representations and that the length of each item vector is one.

Activation of an item representation corresponds to perception. The current state of the item vector corresponds to the item currently being experienced. For instance, an item representation may be activated on the basis of external stimuli during presentation of a list of items. Similarly, an item

representation may be activated by means of an “internal stimulus” during the recall process. No matter the source, the consequence of activating an item representation is the perception of the corresponding item. Howard and Kahana (2002a) assumed that only one item representation could be activated at any one time. Although not a fundamental assumption of TCM, we will also assume that at most one item representation is active at a time throughout the current ms.

Context cues items. In TCM, the current state of context, \mathbf{t}_i , is used to cue recall of items in semantic memory. Each item in semantic memory is activated by a state of context to the extent that that state of context resembles the contexts in which it was presented. This can be implemented using a Hebbian outer product matrix connecting states of context \mathbf{t} with patterns in semantic memory, \mathbf{f}_i

$$\mathbf{M}_i^{TF} = \mathbf{M}_{i-1}^{TF} + \mathbf{f}_i \mathbf{t}_i', \quad (1)$$

where the prime denotes the transpose. When \mathbf{M}^{TF} is multiplied from the right with a context vector, \mathbf{t} , this results in a superposition of patterns in semantic memory, each weighted by their similarity between their context and the cue context. That is

$$\mathbf{M}^{TF} \mathbf{t}_j = \sum_i [\mathbf{f}_i (\mathbf{t}_i' \mathbf{t}_j)], \quad (2)$$

which follows immediately from the definition of \mathbf{M}^{TF} (Eq. 1) and basic properties of vector arithmetic. The key here is the $\mathbf{t}_i' \mathbf{t}_j$ term. The transpose of a vector multiplied by another vector is a scalar referred to as the inner product. For the present purposes, this is the same as the dot product and can also be written $\mathbf{t}_i \cdot \mathbf{t}_j$.¹ We can see that when the item layer is cued by a state of context, the result is a combination of item representations. A particular item enters this combination in a way that is proportional to the similarity (quantified by the dot product) of the contexts it has been presented in to the cue state of context.

Mapping activations onto probability of recall. Howard and Kahana (2002a) assumed that this combination of item representations was unstable. Due to attractor dynamics, the superposition of item representations that results from cuing with a state of context would collapse into one particular item representation (or perhaps a null state in which all elements of the vector went to zero). Let us define the *activation* of a particular item i by a particular state of context \mathbf{t} as

$$a_i := \mathbf{f}_i' \mathbf{M}^{TF} \mathbf{t}. \quad (3)$$

¹ For the present purposes, we can define the inner, or dot product as

$$\mathbf{v} \cdot \mathbf{w} = \sum_i [\mathbf{v}]_i [\mathbf{w}]_i,$$

where the $[\cdot]_i$ operator refers to the i th element of the vector taken as its argument. The dot product is positive if the two vectors point in similar directions (if they are correlated). It is negative if they point in opposite direction. Importantly, the dot product is zero if the two vectors are orthogonal.

Using this definition (and the assumption that the item representations are orthonormal), then the scalar a_i just measures the extent to which the superposition points in the direction of the word corresponding to \mathbf{f}_i . The probability of recalling item i given \mathbf{t} can be given by the Luce choice rule:

$$P(\mathbf{f}_i|\mathbf{t}) = \frac{\exp\left(\frac{2a_i}{\tau}\right)}{\sum_j \exp\left(\frac{2a_j}{\tau}\right)}. \quad (4)$$

This can be conceived of as the probability of the superposition collapsing to a particular state. Howard and Kahana (2002a) took the sum in the denominator of Eq. 4 to be over potential recalls in the list. This equation is not a fundamental part of TCM. The important properties of this equation are simply that it provides a non-linear mapping between activations and recall, and that it is a competitive recall rule. That is, the probability of recalling item i depends not only on the activation a_i , but also the activation of the other items a_j . This makes it a useful equation for describing situations in which we are interested in the relative probability of recalling an item.

Items retrieve context. In much the same way that temporal context can be used to provide an input to the item-space, items provide the input to the context-space. Howard and Kahana (2002a) proposed that a matrix \mathbf{M}^{FT} provides a connection such that the input to the context layer at time step i , \mathbf{t}_i^{IN} , is a consequence of the item presented at time step i :

$$\mathbf{t}_i^{IN} = \mathbf{M}_{i-1}^{FT} \mathbf{f}_i. \quad (5)$$

The vector \mathbf{t}_i^{IN} will sometimes be referred to as the ‘‘context retrieved by item i ’’ to emphasize the effect of item representations on contextual states. The form of \mathbf{M}^{FT} was derived in such a way to implement a functional rule that will be introduced later (Eq. 9 below). The form of \mathbf{M}^{FT} is rather complicated and probably does not correspond simply to any single structure in the brain. For this reason we will not discuss it further here, but rather treat the functional rule as the basic description of retrieved context for the present ms. However, we strongly assert the central point of retrieved context that items cause contextual input patterns.

Recency across time scales

At each time step, the state of context at time step i , \mathbf{t}_i is formed from the prior state of context \mathbf{t}_{i-1} and an input vector \mathbf{t}_i^{IN} according to:

$$\mathbf{t}_i = \rho_i \mathbf{t}_{i-1} + \beta \mathbf{t}_i^{IN}, \quad \rho_i : \|\mathbf{t}_i\| = 1 \quad (6)$$

When applied to list-learning applications, we have previously assumed that the time-steps correspond to the times at which list items are presented. We will assume (for convenience) that the input vectors, \mathbf{t}_i^{IN} , are always of unit length ($\|\mathbf{t}_i^{IN}\| = 1$, for all i). The vector \mathbf{t}_i^{IN} is weighted by the scalar β . This parameter is generally estimated from the data and is constrained such that $0 \leq \beta \leq 1$. We can see that Eq. 6 adds

input vectors to the state of \mathbf{t} . To ensure that the length of \mathbf{t}_i does not grow without bound, we assume that the scalar ρ_i is chosen to ensure that the length of \mathbf{t}_i remains unity: $\|\mathbf{t}_i\| = 1$. This constraint means that \mathbf{t}_i changes as a function of input to the system, rather than the passage of time *per se* (Waugh & Norman, 1965). This can be seen clearly if one assumes that at some time step i , the input vector is empty, $\|\mathbf{t}_i^{IN}\| = 0$. In this case

$$\mathbf{t}_i = \rho_i \mathbf{t}_{i-1} + \beta \mathbf{0} \quad (7)$$

requires that $\rho_i = 1$. This is consistent with the findings of Baddeley and Hitch (1977), who argued that the recency effect was unaffected by addition of an unfilled delay at the end of the list.

If the system is presented with an infinitely long series of orthonormal \mathbf{t}^{IN} 's, then the value of ρ_i will stabilize at $\rho := \sqrt{1 - \beta^2}$.² Under these circumstances, it becomes possible to concisely describe the similarity relationships between \mathbf{t}_i and the state of context at some other time, j , \mathbf{t}_j :

$$\mathbf{t}_i \cdot \mathbf{t}_j = \rho^{|i-j|}. \quad (8)$$

From this it is clear that \mathbf{t} changes gradually over time. Any particular component of \mathbf{t}_i decays exponentially as long as orthonormal inputs are presented.

In sum, contextual evolution in TCM is characterized by several important properties:

1. For a given value of β , when given a series of orthogonal inputs, the similarity of the current state to an initial state decays exponentially.
2. The change in context depends on the input.
3. When no input is given, the state of context does not change.

Because \mathbf{t}_i is the functional cue for recall, and \mathbf{t}_j is an effective cue for recall of item j to the extent that \mathbf{t}_i overlaps with \mathbf{t}_j , the property that \mathbf{t}_i decays gradually naturally provides a basis for the principle of recency (Howard & Kahana, 2002a), which is observed in all of the major episodic memory paradigms (Howard & Kahana, 1999; Murdock, 1962, 1963b; Neath, 1993; Ratcliff & Murdock, 1976). Appendix A illustrates this principle with a worked example that demonstrates the recency effect.

Long-term Recency and Buffer Models of Short-term Memory. For many years, the conventional wisdom was that the recency effect in free recall reflected the operation of a

² Under these circumstance, $\mathbf{t}_i^{IN} \cdot \mathbf{t}_{i-1} = 0$ and

$$\|\mathbf{t}_i\|^2 = \rho_i^2 \|\mathbf{t}_{i-1}\|^2 + \beta^2 \|\mathbf{t}_i^{IN}\|^2.$$

Because $\|\mathbf{t}_i^{IN}\| = 1$ by assumption and $\|\mathbf{t}_{i-1}\| = 1$ because of the condition on ρ_{i-1} , we find that the condition that $\|\mathbf{t}_i\| = 1$ implies that

$$1 = \rho_i^2 + \beta^2,$$

which implies that $\rho_i = \sqrt{1 - \beta^2}$. More generally, when $\mathbf{t}_i^{IN} \cdot \mathbf{t}_{i-1} \neq 0$, a quadratic equation in ρ_i is obtained, which can be solved by elementary methods.

short-term memory buffer (Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1980). Indeed, detailed search models based on a short-term memory buffer can describe standard free recall in considerable detail (Kahana, 1996; Raaijmakers & Shiffrin, 1980, 1981; Sirotin, Kimball, & Kahana, submitted). The recency effect in immediate free recall is eliminated by a distractor at the end of the list (Glanzer & Cunitz, 1966; Postman & Phillips, 1965), presumably because the distractor removes items from the end of the list from STS. However, when a distractor is also presented between each list item, this results in an increased recency effect over delayed free recall (Bjork & Whitten, 1974; Glenberg et al., 1980; Glenberg, Bradley, Kraus, & Renzaglia, 1983; Howard & Kahana, 1999; Nairne, Neath, Serra, & Byun, 1997; Thapar & Greene, 1993; Watkins, Neath, & Sechler, 1989). This presentation schedule is referred to as continuous-distractor free recall; the recency effect observed in continuous-distractor free recall is referred to as the long-term recency effect. Howard and Kahana (2002a) fit TCM to the probability of first recall, a sensitive measure of the recency effect (Howard & Kahana, 1999; Laming, 1999), to data from immediate, delayed and continuous-distractor free recall (see Figure 1).³ TCM accurately predicts the existence of a recency effect in immediate free recall, the disruption of recency in delayed free recall and the recovery of recency in continuous-distractor free recall.

Although contextual drift in TCM can account for much of the function of STS in free recall, there is of course much more to the concept of short-term memory than a rehearsal buffer. Atkinson and Shiffrin (1968) emphasized the importance of control processes in strategically manipulating the information in the buffer. This theme has persisted not only in the emphasis of the working memory framework introduced by Baddeley and Hitch (1974) on executive function, but also in more recent models of executive functioning in prefrontal cortex (e.g. Rougier & O'Reilly, 2002; Braver et al., 2001, for an integrative review, see Miller & Cohen, 2001). Although we argue that t_i captures the critical storage processes of short-term memory essential for generation of the recency effect, we make no claim whatever that it describes control processes or executive function—these functions clearly require something external to TCM.

Retrieved temporal context and episodic association

In free recall, the canonical episodic memory task, subjects recall multiple words from a list without concern to word order. A great deal of evidence indicates that the order in which the items are recalled reflects the associative structure of memory. For instance, when a list of words from different natural categories is presented, words from the same category will tend to be recalled together, even if presentation order is randomized (e.g. Bousfeld, 1953; Pollio, Kasschau, & DeNise, 1968). This tendency for adjacent recalls to come from the same category can be taken as a measure of stronger associations between words from the same category than between words from different semantic categories.

In this case, output order in free recall presumably reveals something about the structure of semantic memory. In addition to semantic, or structural sources of association, associations can also be formed rapidly among items presented in temporal proximity. If free recall is indeed a consequence of an episodic representation, then temporally-defined output order relationships should reveal the properties of this episodic representation.

We can define the association between two items functionally as the tendency of one item to cause production of the other. To measure associations in episodic memory Kahana (1996) developed conditional response probability (CRP) curves. CRP curves measure the probability of making a transition from one item to another in free recall as a function of the distance between them in the list. CRPs have now been computed for data collected under a wide variety of situations (Howard & Kahana, 1999; Kahana, 1996; Kahana & Caplan, 2002; Kahana, Howard, Zaromb, & Wingfield, 2002; Klein, Addis, & Kahana, In press; Ward, Woodward, Stevens, & Stinson, 2003). Consideration of these data confirm two very general properties of episodically-formed associations among items in a series:

1. Contiguity. Stronger associations are formed between stimuli that occur near each other in time than between those that are separated by a greater interval.

2. Asymmetry. Forward associations are stronger than backward associations.

Both of these properties have been observed in immediate (Howard & Kahana, 1999; Kahana, 1996; Ward et al., 2003), delayed (Howard & Kahana, 1999; Kahana et al., 2002) and continuous-distractor free recall (Howard & Kahana, 1999), as well as serial recall (Kahana & Caplan, 2002; Raskin & Cook, 1937).

Because the current state of context is always the cue for episodic recall, associative effects in TCM are mediated by the effects items have on the state of context. This is possible because a central postulate of TCM is that the input to Eq. 6 is caused by the presentation of items.⁴ In TCM items cause an input, t_i^N , that is part of t_i . Because t is the cue for episodic recall, associative effects between items are mediated by the effect they have on t —by the contextual inputs those items evoke, and the similarity of those inputs to states of t in which other items were encoded. TCM produces contiguity effects because items retrieve contextual elements that were present when the items were initially presented. Because context changes gradually (Eq. 6), these contextual el-

³ Details of the procedure can be found in Howard and Kahana (2002a).

⁴ Although this might not seem such a radical assumption, several memory models have included mechanisms of contextual drift in which the change in contextual elements is assumed to be a consequence of stochastic fluctuations that are not under experimental control (Estes, 1955; Mensink & Raaijmakers, 1988; Murdock, Smith, & Bai, 2001). Another set of models developed to explain performance in short-term serial recall tasks have explicitly decoupled contextual representations from item recall, while not necessarily assuming that context fluctuates randomly (Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1992, 1999; Henson, 1998).

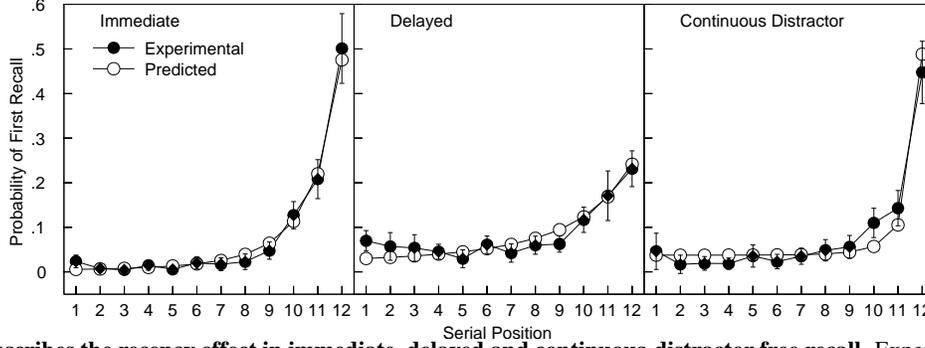


Figure 1. TCM describes the recency effect in immediate, delayed and continuous-distractor free recall. Experimental and predicted values of the probability of first recall, a sensitive measure of the recency effect across delay schedules. **a.** In immediate free recall, the recall test follows immediately after the presentation of the last item. **b.** In the delayed condition, sixteen seconds of a distractor task intervened between presentation of the last list item and the recall test. Accordingly, the recency effect, the advantage for recall of the last items in the list, was greatly reduced. **c.** In continuous-distractor free recall, sixteen seconds of distractor intervened between the last item of the list and the recall test, but also in between each item of the list, effectively “stretching out” the list while preserving the relative temporal spacing of the list. Under these circumstances, the recency effect was much larger than that observed in delayed recall. Because information that enters t_i decays gradually, TCM, when coupled with a competitive retrieval rule, can describe the persistence of the recency effect across time scales. Model results are from Howard and Kahana (2002a). The experimental data is taken from Howard and Kahana (1999).

ements will tend to overlap with “nearby” states of context. Because a state of context cues a given item for recall to the extent that it overlaps with the context(s) in which the item was presented (Eqs. 1, 3), these retrieved contextual elements will favor recall of nearby items. TCM predicts asymmetry because of the detailed assumptions about the nature of these retrieved contextual elements.

Two components of contextual retrieval. Because retrieved context provides the basis for associations between items, the form of \mathbf{M}^{FT} is clearly very important. Howard and Kahana (2002a) hypothesized that retrieved context should be a combination of prior contextual states and the context initially retrieved by an item. Let us refer to the i th time step at which stimulus A is presented as A_i . The input caused by stimulus A changes from presentation to presentation according to

$$\mathbf{t}_{A_{i+1}}^{IN} = \alpha_O \mathbf{t}_{A_i}^{IN} + \alpha_N \mathbf{t}_{A_i} \quad (9)$$

where α_O determines the level of retrieval of old contextual associations and α_N determines the level of new item-to-context associations.⁵ This is a critical further assumption beyond Eq. 5 that allowed the specification of a learning rule for \mathbf{M}^{FT} (Howard & Kahana, 2002a).⁶ The values of α_O and α_N are calculated on each learning trial such that the length of the retrieved context vector on subsequent presentations of A will be one (see Appendix B for details). Howard and Kahana (2002a) derived a learning rule for \mathbf{M}^{FT} to allow the model to simultaneously satisfy Eqs. 5 and 9. The matrix \mathbf{M}^{FT} probably does not correspond simply to a single brain structure, so here we will simply take the functional description of contextual retrieval, Eq. 9, as the basic level of description for changes in contextual retrieval. Equation 9 states that when item A , initially presented at time A_i is repeated later on at time A_{i+1} , the input to Eq. 6, $\mathbf{t}_{A_{i+1}}^{IN}$ will be a combination of two components:

1. The input from the original presentation, $\mathbf{t}_{A_i}^{IN}$, weighted by α_O .

2. The context, \mathbf{t}_{A_i} , that was present when the item was initially presented, weighted by α_N .

The ratio of these two components is controlled by a free parameter $\gamma := \alpha_N / \alpha_O$. These two components give rise to qualitatively different associative effects.

Two components describe episodic association. TCM describes asymmetric associations between stimuli in episodic recall (Howard & Kahana, 1999; Kahana, 1996; Kahana & Caplan, 2002) as a consequence of the combined effects of the two components of Eq. 9. One component, $\mathbf{t}_{A_i}^{IN}$, is the same input pattern that was evoked by A when it was initially presented. Because $\mathbf{t}_{A_i}^{IN}$ does not contribute to contextual states that preceded A_i , but does contribute to subsequent states of context (see Eq. 6), $\mathbf{t}_{A_i}^{IN}$ provides an asymmetric cue that favors forward recalls. The other retrieved context component, \mathbf{t}_{A_i} , is the context that was present when A was presented previously. Because each state of context in a list of non-repeated items is as similar to its predecessor as it is to the states that follow, \mathbf{t}_{A_i} provides a symmetric retrieval cue that favors nearby list items in both the forward and backward directions (see Eq. 8). In concert, these two retrieval cues provide an asymmetric retrieval cue that favors recall of

⁵ The notation used here is slightly different from that used in Howard and Kahana (2002a). There α_O was referred to as A_i and α_N was referred to as B_i . The notation used here is consistent with that used in Howard et al. (In revision).

⁶ In treating the effect of normal aging on episodic association, Howard et al. (In revision) introduced a third component, a noise vector weighted by a parameter η to Eq. 9. The function of this term was to provide an ineffectual retrieval cue that could trade off with the other two components to model the age-related deficit in associative processes. The interested reader should be aware that the version of Eq. 9 used here is not the most general case.

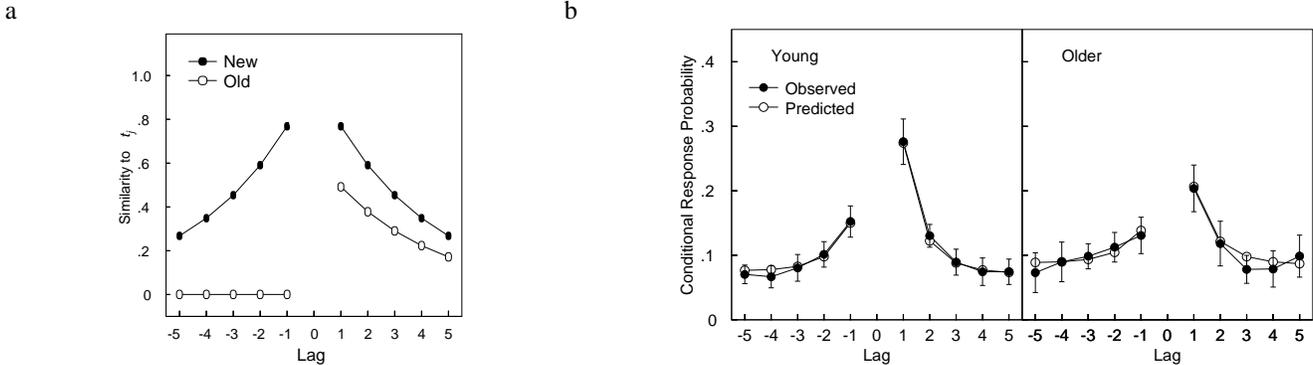


Figure 2. TCM provides a natural explanation of asymmetric association in free recall. **a.** In TCM there are two sources of associative effects. One source relies on the ability to retrieve contextual elements consistently from presentation to presentation of an item. The cue strength derived from these “old” item-to-context associations provides an asymmetric cue that only helps recall items forward in the list. The other source is the ability of an item to retrieve contextual elements that were already present when the item is presented. The cue strength derived from these “new” item-to-context associations provides a symmetric cue that helps both forward and backward recalls. The combination of these two cues leads to the characteristic shape of the CRP. After Howard and Kahana (2002a). **b.** The combination of an asymmetric retrieval cue and a symmetric retrieval cue is an asymmetric retrieval cue. This results in good quantitative fits to observed CRP curves. The left panel shows data from a delayed free recall study of younger adults along with predicted data from TCM. The right panel shows analogous curves from older adults. The decrease in associative tendencies for older adults was modeled as a result of including a noise term in Eq. 9. This data was originally presented in Kahana, et al (2002). The modeling of the older adults’ data is explained in greater detail in Howard et al (in revision).

nearby items.

Figure 2a shows a plot of the cue strength from the two components of context retrieved by an item at the center of the curve to its neighbors. The curve labeled “Old” shows the cue strength of the old context $t_{A_i}^{IN}$ to the neighbors of A . The cue strength is large for items that immediately followed A , and falls off with temporal distance. The old cue strength is zero for items that preceded A . This, combined with the non-zero cue strength to items that followed A leads to an associative asymmetry. The curve labeled “New” in Figure 2a shows the cue strength from the new context component t_{A_i} . This component provides a cue strength that contributes to both forward and backward recalls. Combining these two components in an appropriate ratio shows a strong correspondence to the shape of observed CRPs, a measure of temporally-defined associations observed in free recall (see Figure 2b). Appendix C shows a worked example of a simple calculation of episodically-formed associations that may help to illustrate in more detail why these properties arise from the model.

By varying the relative contributions of α_O and α_N to t^{IN} , we can modulate the directionality of association. When $\gamma = 0$, t^{IN} does not change from presentation to presentation. Under these circumstances, $\alpha_O = 1$ and $\alpha_N = 0$ at each time step. There is a strong forward association and no backward association. Of particular interest here is the fact that the backward association is completely dependent on the value of α_N . If we were somehow able to selectively disrupt new item-to-context learning so that $\alpha_N = 0$, we would observe temporally-defined associations with a form like the curve labeled “Old” in Figure 2a. This ability to dissociate forward from backward associations is consistent with neuropsychological results. Bunsey and Eichenbaum (1996) found that

rats with hippocampal damage were able to learn forward associations as well as control rats, but did not generalize to a backward association.

We saw in the previous subsection that TCM can describe the long-term recency effect. This is a consequence of a gradually decaying strength provided by a contextual cue and a competitive retrieval process. If recency effects and associative effects came from a common source, this would predict that associative effects, like recency effects, should persist across time scales. In a continuous-distractor experiment with great care taken to avoid inter-item rehearsal, Howard and Kahana (1999) observed no reliable change in the shape or magnitude of the CRP functions used to describe associations in free recall with inter-item distractor intervals up to 16 s. Howard and Kahana (2002a) showed that TCM predicts the persistence of both contiguity and asymmetry as the length of the inter-item distractor interval is increased. Howard (2004) provides a more complete set of quantitative predictions for the behavior of TCM coupled with Eq. 4 for calculating probability of recall as the time scale is increased.

A mapping between TCM and the MTL

TCM has been shown to describe fundamental properties of episodic recall performance. MTL damage is known to affect episodic recall (Graf et al., 1984). If TCM provides a realistic description of episodic recall performance, then it ought to be possible to make a mapping of TCM onto the anatomy of the MTL. In this subsection we present a coarse picture of such a mapping. The remainder of this paper evaluates this mapping by examining the ability of TCM with this linking hypothesis to explain the entorhinal place code and consequences of hippocampal lesions on relational memory performance in rats. It should be noted that the results in

these later sections provide much of the justification for the particular mapping proposed here.

Three stages of processing relevant to the functioning of the MTL. Here we briefly summarize the large-scale organization of the MTL and related structures. This presentation draws heavily on reviews by Burwell (2000) and Suzuki and Eichenbaum (2000). The hippocampus proper consists of the CA sub-fields and the dentate gyrus. The hippocampus receives subcortical input from the medial septum *via* the fornix. This input from the septum is essential for the normal operation of theta oscillations, which has an extremely important effect on the normal functioning of the hippocampus (e.g. Hölscher, Anwyl, & Rowan, 1997; Huerta & Lisman, 1993; Wyble, Linster, & Hasselmo, 2000). We will not explicitly model theta here, although theta is almost certainly essential for a detailed physiological description of many of the phenomena discussed here (Hasselmo, Bodelón, & Wyble, 2002; Hasselmo, Hay, Ilyn, & Gorchetnikov, 2002). However, the septo-hippocampal pathway is not believed to carry detailed information about to-be-remembered stimuli. Detailed stimulus representations are believed to be conveyed to the hippocampus *via* the perforant path from EC, which provides the primary informational input to the hippocampus proper.

The entorhinal cortex is reciprocally connected to perirhinal and postrhinal/parahippocampal cortex.⁷ These three regions, collectively referred to as the parahippocampal region, provide the cortical inputs to the hippocampus proper, and are, in turn, reciprocally connected to a wide variety of neocortical association areas. These neocortical association areas draw on every sensory system of the brain as well as higher-order multimodal association areas.

In summary, there are three stages of information processing relevant to the large-scale structure of the MTL. Cortical association areas gather higher-order information from the cortex and provide input to the MTL via parahippocampal regions. Parahippocampal regions, including entorhinal, perirhinal and postrhinal (parahippocampal) cortices are reciprocally connected and provide input to the hippocampus proper, primarily through EC. The hippocampus proper, then, receives input from essentially the entire brain in a small number of synapses.

Mapping TCM onto the three stages. We will argue that the three large-scale stages described above correspond to structures and functions within TCM. We will argue that item representations, \mathbf{f} , correspond to cortical association areas, that the context vector, \mathbf{t}_i , resides in parahippocampal regions, including in particular EC, and that a function of the hippocampus proper is to affect new item-to-context learning, corresponding to a nonzero value of α_N in Eq. 9. This corresponds to a reconstruction of patterns of activity in EC that were present when an item was initially presented.

Item representations are activated when an item is perceived, whether as a result of external stimulation or recall of an item by means of connections from the context layer. General perception and cognition is generally not affected by

even extensive MTL lesions (see Corkin, 2002, for a recent review). This leads us to hypothesize that the item representations, the \mathbf{f} vectors, reside outside of the MTL, in the cortical association areas that project to the parahippocampal region.

In this ms we advance the hypothesis that \mathbf{t}_i resides in parahippocampal regions. Before laying out the reasoning for this hypothesis, we first consider the evidence for the alternative hypothesis that \mathbf{t}_i resides in the prefrontal cortex. Changes in the context vector \mathbf{t}_i are associated with the recency effect, the recency effect is associated with short-term memory (e.g. Atkinson & Shiffrin, 1968). Short-term memory is associated with working memory (Baddeley, 1986; Baddeley & Hitch, 1974) and working memory is associated with prefrontal cortex (PFC). There is indeed ample evidence that the PFC is involved in working memory tasks (e.g. Goldman-Rakic, 1996; Rypma & D'Esposito, 1999; Smith & Jonides, 1999). Working memory involves a great many cognitive functions beyond those necessary to support a recency effect, notably executive and attentional functions. Although frontal regions participate in encoding and retrieval into episodic memory (for recent reviews see Rugg, Otten, & Henson, 2002; Simons & Spiers, 2003), this does not imply that the locus of \mathbf{t}_i is in PFC, even if one grants that TCM is an accurate description of episodic recall. For instance, encoding and retrieval related activations in PFC may reflect a gating function allowing selective access to \mathbf{t}_i . Indeed, a number of computational models have emphasized the executive and organizational properties of PFC in working memory tasks (Becker & Lim, in press; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Dehaene & Changeux, 1997; Rougier & O'Reilly, 2002).

There is good evidence (beyond the simulations of entorhinal place cells that will be reported in the following section) to support the hypothesis that \mathbf{t}_i resides in parahippocampal regions, including EC. As discussed above, \mathbf{t}_i functions very much like a short-term memory store in non-spatial tasks. There is strong evidence that extra-hippocampal MTL structures, including EC, have properties consistent with a role in non-spatial memory over the scale of tens of seconds. Given that animals cannot do free recall of words, the best analogue of the recency effect in free recall is the forgetting observed with recognition of non-spatial stimuli over tens of seconds.

There is evidence for a role of parahippocampal regions in such tasks from both single-unit and lesion studies. In a delayed match to sample (DMS) task using odor stimuli in the rat, Young, Otto, Fox, and Eichenbaum (1997) showed that responses of parahippocampal neurons, including those in the lateral EC, exhibited stimulus-specific firing that persisted into the delay interval. Suzuki, Miller, and Desimone (1997) extended this result to demonstrate that this stimulus-specific firing persisted across multiple intervening stimuli. Buffalo, Reber, and Squire (1998) showed that peo-

⁷ The nomenclature postrhinal cortex is used in rats, whereas the homologous region is referred to as parahippocampal cortex in monkeys.

ple with lesions to the perirhinal cortex showed deficits of recognition memory over delays as short as 6 s. Mumby and Pinel (1994) showed that rats with damage to entorhinal and perirhinal cortex were impaired on delayed non-match to sample (DNMS) of trial-unique object at delays as short as 15 s. Otto and Eichenbaum (1992) showed no deficit in a continuous delayed non-match task from fornix transection, but showed a deficit from combined perirhinal/entorhinal lesions at delays of 30 s. This not only points to a role for the parahippocampal regions in memory on the time scale of the recency effect in free recall, but argues against a role of the hippocampus in such processes. Murray and Mishkin (1998), showed that lesions to the amygdala and hippocampus that spared rhinal cortex did not have an effect on DNMS performance, whereas a comparable study showed a severe impairment from rhinal cortex lesions at delays as short as tens of seconds (Meunier, Bachevalier, Mishkin, & Murray, 1993).

States of context \mathbf{t}_i also include contextual input patterns \mathbf{t}_i^{IN} (see Eq. 6). The hypothesis that \mathbf{t}_i resides in parahippocampal regions brings with it the corollary that \mathbf{t}_i^{IN} also resides in parahippocampal regions. As we have seen, \mathbf{t}_i^{IN} is caused by the particular item presented to the network at time i (Eq. 5). In this way, we can think of \mathbf{t}_i^{IN} as a higher-order stimulus representation driven by item presentation. The newly-activated contextual elements \mathbf{t}_i^{IN} would depend on the item presented and its prior history. These elements would be present in a background of activity \mathbf{t}_{i-1} that in turn depends on the prior items presented and their history.

If \mathbf{t}_i resides in parahippocampal regions, then what is the function of the hippocampus proper? The first suggestion comes from the finding that hippocampal damage is associated with a disruption of memory for items from the early part of the serial position curve. Studies of epileptic patients who received anterior temporal lobectomies that included hippocampal resection show a deficit in memory that is largest for items from early and middle serial positions (Hermann, Seidenberg, Wyler, et al., 1996; Jones-Gotman, 1986). These studies both suggested that damage to the hippocampus itself was responsible for the deficit. Jones-Gotman (1986) showed that performance was related to the extent of the damage to the right hippocampus in memory for visual materials. Hermann et al. (1996) showed that memory for verbal materials was more affected by the lobectomy in patients who did not have hippocampal sclerosis in the left hippocampus, suggesting that the non-sclerotic hippocampus was contributing to recall of pre-recency items prior to the operation. Lesion studies in rats also support the view that memory for the early and middle items in a list depends on an intact hippocampus (Kesner, Crutcher, & Beers, 1988; Kesner & Novak, 1982). Although it is not as clear that the hippocampus in particular is implicated, studies of human amnesics have also argued for a dissociation between the recency portion and pre-recency portions of the serial position curve (Baddeley & Warrington, 1970; Carlesimo, Marfia, Loasses, & Caltagirone, 1996).

In TCM, recall of items from the end of the list is predom-

inantly a result of the recency effect caused by using end-of-list context as a cue. In contrast, recall of non-recency items is predominantly a consequence of contextual retrieval giving rise to temporally-defined associations. Indeed Kahana et al. (2002) showed that the mnemonic deficit in normal aging, which may be associated with MTL dysfunction (Grady et al., 1995) results in normal recency effects, accompanied by reduced temporally-defined associations, which can be explained within TCM as a disruption of the process of contextual retrieval (Howard et al., In revision).

If damage to the hippocampus proper resulted in a disruption of contextual retrieval, this would manifest as a deficit for pre-recency items. However, a complete disruption of contextual retrieval, with say $\mathbf{t}_i^{IN} = \mathbf{0}$, would result in a disruption of the recency effect as well, because the rate of contextual drift depends on the amount of input provided. In any event, the state of temporal context \mathbf{t}_i in parahippocampal regions should be able to be affected by input from item representations in neocortical association areas. These considerations lead us to hypothesize that the hippocampus is responsible for a more subtle aspect of contextual retrieval. In this manuscript we explore the hypothesis that the hippocampus is responsible for learning new item-to-context associations. Hippocampal lesions will be modeled by setting α_N to zero. More concretely, we hypothesize that the hippocampus functions to recover the state active in EC when an item was previously presented (Figure 3).

The hypothesis that the hippocampus affects associative memory by recovering states of activity in EC is consistent with the finding that hippocampal damage results in a deficit for backward associations. In the Bunsey and Eichenbaum (1996) experiment, rats learned something like a paired associate task. In a cue phase, the animals were presented with an odor. In a choice phase, they had to select which of two scented cups contained a food reward. The odor presented in the cue phase of the trial predicted which of the scents contained the reward. Correct performance depended on the formation of some sort of association between each cue odor and the correct choice odor. Animals with hippocampal damage were able to perform the choice as well as unlesioned animals. In a second phase of the experiment, animals were tested on their generalization to the backward association. In this phase, the odors from the choice phase were presented as cues to select among. Control-lesioned rats selected the odor consistent with the presence of a backward association. That is, after learning to choose B when cued with A , control rats chose A when cued with B . Despite their ability to learn the forward association as well as control rats, the hippocampal-lesioned rats showed no development of a backward association. In TCM, this finding of impaired backward associations and intact forward associations is what one would expect if the hippocampus was necessary to make $\alpha_N > 0$. If $\alpha_N = 0$, then this “lesioned” model would be able to make forward associations, but would not support backward association; the lesioned model would show associations like the curve labeled “old” in Figure 2a.

The mapping between TCM and the MTL describes a process of memory encoding and retrieval. Item presentation

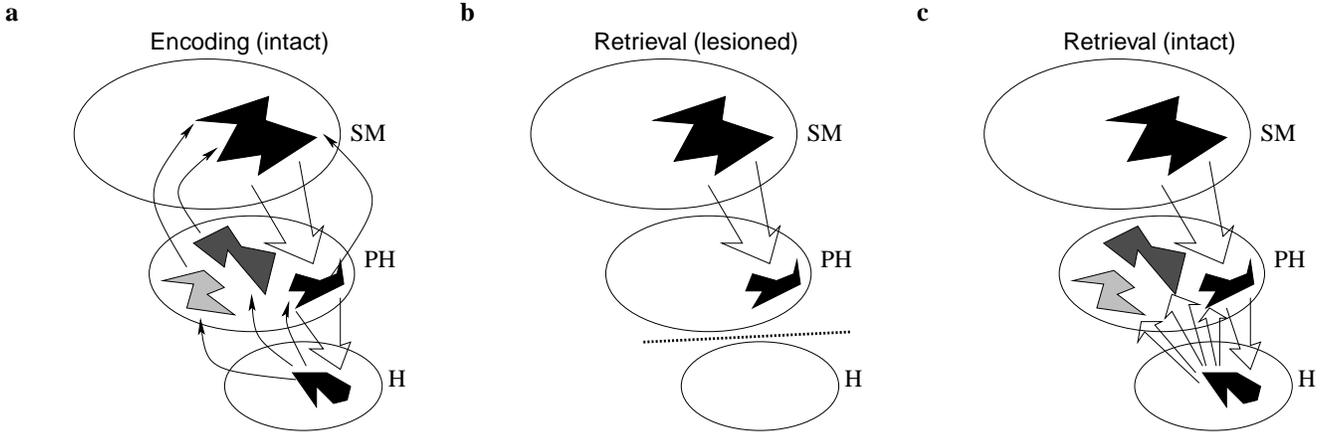


Figure 3. A linking hypothesis between TCM and the MTL. **a.** “Items” are patterns of activity in semantic memory (SM), which is presumed to reside in cortical association areas. These areas project to parahippocampal (PH) regions, including at least EC, which support a state of context \mathbf{t}_i which serves as the cue for episodic recall. Presentation of an item in semantic memory calls up a set of elements \mathbf{t}_i^{IN} in PH. The state of context also includes patterns activated by previous item presentations (the red and green patterns). The set of elements activated by the item causes a set of elements in the hippocampus (H) to be activated, perhaps biased by the other contextual elements active in EC and/or the prior state of activation in H. Hebbian association (indicated by the thin solid lines) takes place between the state of context in PH and the state in semantic memory to allow contextual states to cue the item in semantic memory. **b.** Repetition of the item in semantic memory reactivates the stimulus-specific elements in PH. Because the stimulus-specific elements remained active in PH following the initial presentation of the stimulus, their reactivation serves as a cue for items that followed the initial presentation. **c.** The proposed function of the hippocampus is to allow retrieval of contextual states upon re-presentation of an item. In this case, when the item is re-presented in semantic memory, it again activates the set of stimulus-selective elements in PH, as in **b.** However, H functions to reinstate the entire contextual state that obtained when the stimulus was originally presented. Because this state includes elements derived from items presented prior to the original item presentation, this “retrieved context” functions as a symmetric cue for recall of other stimuli.

corresponds to activation of an appropriate pattern in cortical association areas. These provide an input to EC and other parahippocampal regions. These newly-active patterns of activity decay over time as new items are presented, activating other patterns of input. At any time, the state of activity in parahippocampal regions is the cue for episodic retrieval. Repeating an item representation has an effect on the pattern of activity in parahippocampal regions. If the hippocampus is functioning properly, it enables repetition to result in the recovery of other patterns of activity that were present when the item was initially presented. Disruption of hippocampal function does not prevent an item from activating a pattern in parahippocampal regions. However, it does prevent item presentation from reconstructing other patterns of activity in parahippocampal regions. Figure 3 attempts to illustrate these properties. In this view, the hippocampus does not “contain” memories *per se*. Rather, it operates to change the pattern of activity in EC, which cues cortical regions. Successful activation of cortical regions corresponds to the act of remembering. Insofar as the function of the hippocampus and MTL is to draw together different transiently active cortical representations it bears a strong resemblance to hippocampal indexing theory (Teyler & DiScenna, 1985, 1986).

Preview. In the remainder of this ms, we will explore the value of the linking hypothesis described above by arguing that TCM describes location-specific firing characteristics of cells in EC and by showing that disrupting contextual learning can describe characteristic effects from relational learn-

Table 1

Parameters used in simulations. The parameters β and γ are intrinsic to TCM. β controls the rate of contextual drift (Eq. 6). γ controls the ratio of new to old retrieved context (the ratio α_N/α_O , Eq. 9). In the relational memory simulations, β was set to $\sqrt{1-\rho^2}$, with $\rho = 0.9$. The large difference between β in the spatial applications and the non-spatial applications is appropriate given the different time scale of contextual evolution (see text for details). σ is specific to the spatial applications and determines the width of the tuning curves for the head direction inputs to the place cells. The value of $\pi/6$ was taken to be coarsely consistent with experimental findings for the head direction system. τ is used in the transitivity simulations and determines the sensitivity of the recall rule (Eq. 4).

Simulation	Parameter			
	Intrinsic		Application-specific	
	β	γ	σ	τ
Open field	0.01	0	$\pi/6$	
W-maze	0.01	0	$\pi/6$	
Transitivity	0.435	0/1		1
Memory space	0.435	0/1		

ing experiments. Table 1 summarizes the values of the parameters used in the simulations. TCM itself contributes two parameters. The value of β , from Eq. 6, determines

how rapidly context changes given a particular set of inputs. Larger values of β mean that context changes rapidly; smaller values mean a more slowly-changing \mathbf{t}_i . The difference between the values of β across applications should not be too troubling given the difference in the time-steps. That is, β determines the change between time step $i - 1$ and time step i . In the spatial applications, the time steps come at 50 Hz (for the open field) and 30 Hz (for the W-maze data). In contrast, the time difference between \mathbf{t}_i and \mathbf{t}_{i+1} in the relational memory applications is much slower, corresponding to the time between sampling of odors, on the scale of seconds.⁸ The value of γ is just the ratio α_N/α_O ; this determines the rate of change of \mathbf{t}^{IN} across different presentations of the same item. The value of γ is different in the spatial compared to the non-spatial applications. This reasons for this are rather subtle and are discussed extensively in the General Discussion. The other two parameters are specific to the subject areas covered in this ms. The spatial applications include a parameter σ that controls the width of the tuning curves of simulated head direction cells. The value of this parameter was taken to be roughly consistent with published properties of actual cells (Taube, 1998). The parameter τ (Eq. 4) is necessary to map activity onto probability of recall. This was used previously in modeling free recall (Howard & Kahana, 2002a) and is used here in the simulation of transitive associations.

The Entorhinal Place Code and Contextual Drift

The most striking piece of data implicating the MTL in spatial navigation is the finding that cells in the hippocampus fire in response to the animal's location within an environment. This phenomenon was first reported by O'Keefe and Dostrovsky (1971) and has subsequently been explored extensively by numerous researchers. This research has centered on the responses of cells within subfield CA1 of the dorsal hippocampus (e.g. Muller & Kubie, 1987; O'Keefe & Burgess, 1996; Shapiro, Tanila, & Eichenbaum, 1997; Thompson & Best, 1989; Wilson & McNaughton, 1993), although other subfields and MTL regions have also been explored (Barnes et al., 1990; Frank et al., 2000; Gothard, Hoffman, Battaglia, & McNaughton, 2001; Jung, Wiener, & McNaughton, 1994; Phillips & Eichenbaum, 1998; Quirk et al., 1992; Sharp & Green, 1994; Skaggs, McNaughton, Wilson, & Barnes, 1996). Given the importance of the hippocampus in learning and memory and the replicability of the place cell phenomena, there have been several attempts to model the computational origin of the place code (e.g. Burgess & O'Keefe, 1996; Dobioli, Minai, & Best, 2000; Hartley, Burgess, Lever, Cacucci, & O'Keefe, 2000; Hetherington & Shapiro, 1993; Kali & Dayan, 2000; O'Keefe, 1991; Redish, 1999; Samsonovich & McNaughton, 1997; Sharp, 1991; Sharp, Blair, & Brown, 1996; Touretzky & Redish, 1996; Zipser, 1985, 1986). One obvious reason, however, that there is a place code in the hippocampus is that it receives input from the EC, which itself shows place-specific firing. The computational/cognitive origin of the hippocampal place code is apparently not in the hippocampus. If we

find a satisfactory explanation of the activity of EC cells, we will be one step closer to understanding the origin of the hippocampal place code.

EC place-specific activity in the open field. Cells in EC exhibit several properties that are not shared with hippocampal place cells. Hippocampal place cells typically show very compact, distinct place fields. Cells that fire robustly (> 10 Hz) in one location within the open field will typically be completely silent when the animal is outside the place field (Thompson & Best, 1989). In contrast, EC place cells typically fire throughout open environments. Firing for these entorhinal cells is reliably modulated by the animal's position (Quirk et al., 1992), but in a much more noisy way than hippocampal cells. In addition to this quantitative difference, qualitative differences are observed in the firing properties of entorhinal vs hippocampal place cells. After repeated exposure to multiple environments (Lever, Wills, Cacucci, Burgess, & O'Keefe, 2002), the hippocampal place code "remaps" from one environment to another. If an animal is observed after extensive experience in two distinct spatial environments, say, a cylindrical enclosure and a square enclosure, the place fields observed in the one environment will be uncorrelated with the place fields observed in the other environment. That is, if a particular hippocampal place cell shows a place field in the Northwest corner of the square enclosure, this does not predict its responsiveness in the cylindrical enclosure; in the cylindrical enclosure it may have a place field in a completely different location or stop firing altogether (Muller & Kubie, 1987). During the initial exposures to unfamiliar environments, the hippocampal place code, like the entorhinal place code, shows similar firing in both environments. In contrast, EC place cells show correlated firing across environments that persists even after extensive training (Quirk et al., 1992). That is, an EC place cell that is more likely to fire in the Northwest quadrant of the square enclosure will also be more likely to fire in the Northwest quadrant of the cylindrical enclosure.

EC place-specific activity on the linear track. A key feature of Eq. 6 is that \mathbf{t}_i is sensitive to the history of inputs leading up to time step i . To make this point more concretely, it is clear from Eq. 6 that \mathbf{t}_i includes \mathbf{t}_i^{IN} and \mathbf{t}_{i-1} . However, because \mathbf{t}_{i-1} contains \mathbf{t}_{i-1}^{IN} , this means that \mathbf{t}_i also contains \mathbf{t}_{i-1}^{IN} . We can continue this process of "unwinding" indefinitely. In this way we find that the context vector \mathbf{t}_i depends on the history of stimulus presentations leading up to time step i . Recent evidence from place cell studies indicates that the entorhinal place code also exhibits history-dependence. Frank et al. (2000) recorded from place cells in EC and CA1 while animals traversed a W-shaped maze. The animals' task was to repeatedly visit the arms of the maze in sequence (see Figure 4a). Of particular interest here is a phenomenon called retrospective coding.

⁸ Given the definition of β , it is also reasonable to assume that some classes of inputs, like odors, might produce a stronger response in EC cells than others.

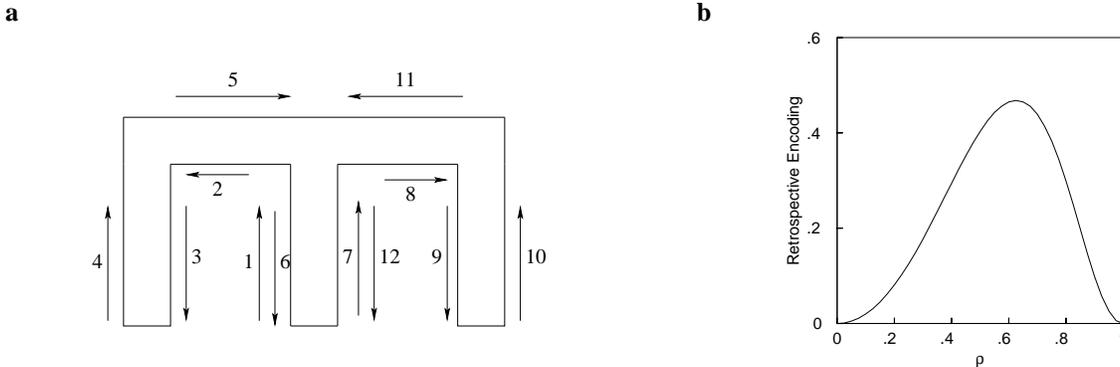


Figure 4. Retrospective encoding requires an imperfect spatial representation. **a.** Simple schematic model of the paths taken by the animal in the W-maze. The animal repeatedly traveled the path 1-2-3-4...-10-11-12-1-2-3... Initially the animal traveled from the center arm to the left arm, a center-left trip (steps labeled 1-3), followed by a left-center trip (4-6), followed by a center-right trip (7-9) and a right-center trip (10-12). A different representation on step 6 compared to step 12 is evidence for retrospective coding. **b.** A simplified version of the TCM context evolution equation was presented with velocity vectors corresponding to the series of movements to generate a positional representation \mathbf{p} . We defined retrospective encoding as $1 - (\mathbf{p}_6 \cdot \mathbf{p}_{12})$. This reflects the degree to which \mathbf{p}_6 and \mathbf{p}_{12} are different from each other. Retrospective coding is plotted as a function of ρ in a general integration scheme, where $\mathbf{p}_i = \rho \mathbf{p}_{i-1} + \mathbf{v}_i$. When $\rho = 0$, \mathbf{p} is just the most recent movement and the model provides a “pure head direction” representation. When $\rho = 1$, \mathbf{p} reflects the sequence of all prior movements and the model provides a perfect place representation. At both of these extremes, the model fails to show evidence for retrospective coding. In contrast, for intermediate values of ρ , the model shows retrospective coding, as seen in EC and the hippocampus (Frank, *et al*, 2000). Although this is an imperfect representation of Euclidean space, it is in some sense superior to a perfect representation, in that it discriminates different episodes that happen in the same location (Wood *et al*, 2000).

In the W-maze, the animal visits the middle arm following visits to either the left arm or the right arm (steps 6 and 12 in Figure 4a). In these situations, the animal’s location, and heading, as well as all available visual cues are presumably identical. These visits differ, however, in the history of movements leading up to them. This provides us an opportunity to distinguish between a “pure place code,” which would predict that cells should not distinguish between 6 and 12 and a “history-dependent pseudo-place code,” which would. Frank *et al.* (2000) found that some cells in EC reliably differentiated these visits, a phenomenon they referred to as retrospective coding. Wood, Dudchenko, Robitsek, and Eichenbaum (2000) observed a similar phenomenon. In their task, the animal repeatedly ran in a figure-8 pattern around an elevated track. As the animal ran up the central stem of the maze, the firing of some hippocampal cells depended on whether the animal was about to turn onto the left or the right arm. This finding provides clear evidence that “place cells” respond to variables other than physical location in the environment. In particular, this result shows that the hippocampal place code distinguishes among separable episodes occurring at the same location—a property that would certainly serve it well in memory more generally (Eichenbaum, 2001; Wood *et al.*, 2000). However, because the animal always alternated between “loops” of the 8, it was unclear from the task whether the cells were coding for the sequence of prior movements or the sequence of future movements in that experiment. Interestingly, Frank *et al.* (2000) observed retrospective coding in cells in superficial EC, which provides input to, but does not receive output from the hippocampus. This suggests that this history-dependence in the entorhinal place code does not depend on the functioning of the hippocampus proper.⁹ In

contrast, cells showing prospective coding that showed differential activity based on where the animal was going to go on trips up the center arm (see Figure 4a) were most robustly observed in deep layers of EC, that receive input from the hippocampus. Retrospective and prospective cells were further differentiated by the spatial distribution of differential firing. Retrospective cells were found that distinguished the prior history of movements along the length of the center arm. In contrast, prospective coding was most frequently seen close to the choice point where the paths diverged (Frank *et al.*, 2000). This suggests that perhaps some sort of postural realignment in preparation of a turn contributes to prospective coding. Recent studies have further illustrated the somewhat controversial relationship between retrospective and prospective coding (Lenck-Santini, Save, & Poucet, 2001; Ferbinteanu & Shapiro, 2003).

In this section we demonstrate that Eq. 6 is sufficient to describe key features of the entorhinal place code, given strictly a velocity, *i.e.* speed plus allocentric head direction, as input. In this section we will demonstrate that in the open field Eq. 6, when provided with velocity vectors as input, gives rise to simulated cells with noisy place fields that are consistent from environment to environment, in correspondence with available data (Quirk *et al.*, 1992). We will also demonstrate that this minimal model is sufficient to describe key features of the entorhinal place code in the W-maze, including the history-dependence illustrated by the phenomenon of retrospective coding. We start with some broad theoretical considerations before presenting a cellular simulation imple-

⁹ It is of course possible that superficial layers of EC acquire these properties as a consequence of indirect connections from the hippocampus.

menting the important properties of Eq. 6.

An imperfect integrator and retrospective coding. How do we keep track of our location as we move around our environment? One way might be to continuously update our position by orienting ourselves relative to salient landmarks. This is undoubtedly one way in which we, and other animals, know our position. But what about when no suitable landmarks are available. What if we are at sea on a cloudy night? Under these circumstances, we might, as ancient sailors did, adopt a strategy of dead reckoning.

Dead reckoning refers to the strategy of figuring out where we are based on the movements we have made. If we start out in a specific location and then make some movement we can figure out where we are after the movement if we add the movement to our initial location using vector addition. For instance, if we start out at some location \mathbf{p}_0 , and move due East along some vector \mathbf{v}_1 , defined in allocentric space, then our location after the movement is just $\mathbf{p}_1 = \mathbf{p}_0 + \mathbf{v}_1$. If we make another movement along some other vector \mathbf{v}_2 , then our new location is $\mathbf{p}_2 = \mathbf{p}_1 + \mathbf{v}_2$. In general, denoting the movement taken at time i as \mathbf{v}_i , and the position at the conclusion of that movement as \mathbf{p}_i , we can keep track of our position using

$$\mathbf{p}_i = \mathbf{p}_{i-1} + \mathbf{v}_i. \quad (10)$$

In this way, we can always keep track of our location relative to our starting point \mathbf{p}_0 . Although the precise form of our place representation will depend on the choice of starting location, the key feature is that the spatial relationships among the \mathbf{p} 's is perfectly preserved.¹⁰

Comparing the contextual evolution equation (Eq. 6) with the dead reckoning equation (Eq. 10), we see that the contextual evolution equation is also integrating its inputs, \mathbf{t}_i^{IN} ; the evolution equation, however, is performing an imperfect, leaky, integration. Because ρ_i is typically less than one, the contextual evolution equation gradually “forgets” inputs as more information is presented. For the sake of the following illustration, let us write an integrator equation similar to Eq. 6:

$$\mathbf{p}_i = \rho \mathbf{p}_{i-1} + \mathbf{v}_i. \quad (11)$$

This is similar to the context evolution equation (Eq. 6) except that ρ does not change from time-step to time-step to enable normalization and there is no β to parameterize the magnitude of the input. Let's consider the behavior of this model with various values of ρ . If $\rho = 1$, this model gives rise to the perfect path integrator described above. If $\rho = 0$, on the other hand, then the representation \mathbf{p} is identical to the current velocity vector: $\mathbf{p}_i = \mathbf{v}_i$. In this case, \mathbf{p} is more like a representation of head direction, if one ignores variation in the speed of movement. As ρ increases from zero, not only the current velocity vector contributes to \mathbf{p}_i , but previous velocity vectors contribute as well. That is, when ρ is intermediate between zero and one, \mathbf{p} is not the result of path integration, nor is it a representation of head direction. It lies somewhere in between, a weighted sum over recent movements, something more like a trajectory. These trajectories should be sensitive to the head direction of the current

movement, as well as to the direction at preceding time steps.

A weighted sum over recent movements is ideal for describing the phenomena of trajectory coding and retrospective coding, whereas neither a perfect path integrator ($\rho = 1$) nor a representation of head direction ($\rho = 0$) can accomplish this. To demonstrate this property Figure 4b shows the result of a simple calculation. Equation 11 was repeatedly presented with velocity vectors corresponding to the appropriate stage of the path through the W-maze. For instance, \mathbf{v}_1 was the same as \mathbf{v}_4 and reflected a movement to the North. We assumed that the velocity vectors were orthogonal to each other. To get an intuition as to what this means, we assumed that “South” is not the opposite of “North,” but rather an entirely different direction. The same holds true for East and West. After presenting many circuits around the maze, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{12}$, the similarity matrix of the \mathbf{p} vectors corresponding to the different stages of the path was constructed. We then took $1 - (\mathbf{p}_6 \cdot \mathbf{p}_{12})$ as a measure of retrospective coding. This reflects the degree to which \mathbf{p}_6 and \mathbf{p}_{12} are different from each other. Figure 4b shows this quantity as a function of ρ . Although there is no retrospective coding for the extremes where $\rho = 0$ or $\rho = 1$, there is retrospective coding for intermediate values of ρ .¹¹ We conclude from this that that a leaky, or “pseudo”-integrator is more appropriate for describing the phenomenon of retrospective coding than is a perfect integrator.

In spatial navigation tasks, we will assume that the dominant source of input to Eq. 6 is provided by information related to movement. Specifically, we will assume that the input to Eq. 6 consists exclusively of velocity vectors derived from the head direction system, modulated by the animal's speed

$$\mathbf{t}_i^{IN} = \mathbf{v}_i \quad (12)$$

where \mathbf{v}_i is the velocity vector at time step i . We implement this model using a cellular-level simulation that we will now describe.

Mechanisms of contextual evolution in EC

Anatomical and electrophysiological data indicate that EC has everything it would need to implement the source of the entorhinal place code postulated here. Three major components are necessary to accomplish this; access to a representation of velocity, the means to add vectors using vector addition, and a mechanism to normalize the context vector.

¹⁰ Actually, an additional necessity for “perfect” path integration is the presence of an additive inverse on the \mathbf{v} 's. Let's suppose you start at position $\mathbf{p}_{\text{start}}$. You make an easterly movement of one unit followed by a westerly movement. You end up in the same position. Now, what would a perfect path integrator model predict? Well, your position after the movements is $\mathbf{p}_{\text{end}} = \mathbf{p}_{\text{start}} + \mathbf{v}_E + \mathbf{v}_W$. The integration is only successful if $\mathbf{v}_E = -\mathbf{v}_W$. This need not be the case, as, for instance, in the simulations below.

¹¹ The values of ρ plotted in Figure 4b should not be directly compared to values of $\sqrt{1 - \beta^2}$ taken from the value of β used in the cellular simulation, later. The difference between the time steps of the cellular simulation is several orders of magnitude smaller than the time step between steps on the W-maze as defined by Figure 4a.

Head direction system. The MTL has access to a representation of heading from the head direction system (Taube, 1998). Cells in the head direction system respond preferentially when the animal's head is pointed in a particular direction in allocentric space. For instance, one head direction cell might respond best when the animal is pointed toward the North end of the room, independent of the animal's location. Another head direction cell might respond best when the animal is pointed toward the Southeast. A large number of such cells would provide very precise information about the animal's heading. If the inputs of these cells to the MTL were gated by information about running speed¹² this would provide the necessary velocity signal as input for Eq. 6. There is ample evidence to suggest that the head direction system contributes to the maintenance of the place code. Perhaps most compellingly, disruption of the vestibular sense disrupts the head direction system and also has a profound effect on the hippocampal place code (Russell, Hori, Smith, Darlington, & Bilkey, 2003; Stackman, Clark, & Taube, 2002).

Integrator cells. Cells in EC have precisely the electrophysiological properties necessary to implement Eq. 6. Egorov, Hamam, Fransén, Hasselmo, and Alonso (2002) observed cells in EC layer V that performed an integration on their inputs. These cells were able to adopt a stable firing rate in the absence of external inputs. In the absence of external inputs ($t_i^N = 0$), t remains constant. In this case, $\rho_i = 1$ and $t_i = t_{i-1}$. The existence of a stable firing rate in the absence of input observed by Egorov et al. (2002) provides the capability to implement this property. Further, Egorov et al. (2002) observed that these cells respond to subsequent suprathreshold inputs by adopting a new stable firing rate (see Figure 5). Depolarizing (positive) inputs resulted in a higher stable firing rate. Hyperpolarizing (negative) inputs resulted in a lower stable firing rate. This would enable the cells to perform the vector addition necessary to implement Eq. 6 when t_i^N is of non-zero length. It is worth noting that neuroanatomical studies have demonstrated that the pre-subiculum, which contains head-direction cells, projects to EC layer V cells (Haefen, Wouterlood, & Witter, 2000).

Normalizing gain modulation. The other main property of Eq. 6, is an exponential decay in the presence of additional inputs. This would require that the firing rate of a decaying cell be multiplied by a scalar less than one at each time step. This amounts to a gain control on the internal current that allows integrator cells to sustain firing in the slice. Gain modulation has been widely observed in diverse cortical systems (for a review, see Salinas & Thier, 2000). This in itself, however is not sufficient to enable us to implement the important properties of Eq. 6. A constant gain would cause t_i to decay even when there was no input provided, in contrast to one of the main properties of Eq. 6. To implement Eq. 6, the gain should be inversely related to the total network activity. That is, when the network is more active, the gain should be lower; when the network is less active, the gain should be higher. Chance, Abbott, and Reyes (2002) measured the gain of cultured somatosensory cells. They injected a con-

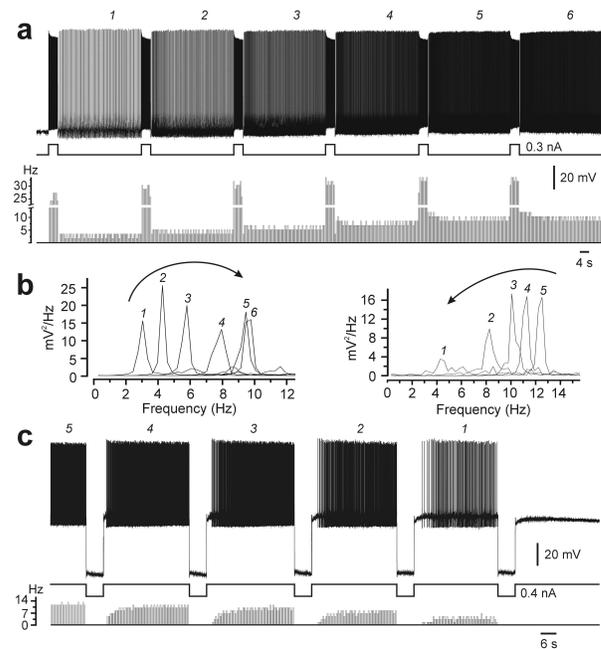


Figure 5. Cells in layer V EC integrate their inputs. Recordings were made from slices bathed in a solution including low concentrations of the cholinergic agonist muscarine. **a.** Cells from layer V, when presented with a depolarizing input, began firing at a stable rate (epoch 1). As subsequent depolarizing inputs were presented, the cell adopted a new, higher, stable firing rate. **c.** Analogously, when the cell is at a high firing rate, hyperpolarizing inputs cause the cell to adopt a lower, stable firing rate. **b.** Shows power spectra for the epochs labeled in **a** (left) and **b** (right). From Egorov, *et al* (2002).

stant amount of current and measured the cell's firing rate. They took the slope of output firing rate to input current as a measure of the cell's gain. In addition to the driving current Chance et al. (2002) also injected a current designed to mimic synaptic currents from some number of other cells. These inputs balanced excitatory and inhibitory input, so that the net current was zero. As the number of simulated synaptic inputs was increased, simulating a higher level of overall network activity, the target cells' gain factor was reduced.

¹² One candidate for this "movement gating signal" is the hippocampal theta rhythm. The presence or absence of type I theta during navigation is closely yoked to the animal's movement. The signal formed from movement direction information derived from the head direction system, coupled with theta-derived speed information would provide a representation of velocity. In fact, Vertes, Albo, and Viana Di Prisco (2001) have pointed out that the regions with head direction cells are always adjacent to regions which contain theta firing cells. Vertes et al. (2001) note further that these populations don't appear to have reciprocal connections, as if their function was to cooperatively represent velocity to downstream regions.

With a population of integrator cells, gain modulation of this type could cause the network to maintain a stable level of activity, implementing something like normalization.¹³ In addition to providing new insight into a basic principle of cortical information processing, the result of Chance et al. (2002) provides evidence for a process that should enable us to implement the key properties of Eq. 6.

Cellular Simulation Methods

Here we will introduce methods for the cellular simulation, which will be applied to the open field and the W-maze. Thus far we have used subscripts to refer to the time step. For instance, we have used \mathbf{t}_i to refer to the temporal context vector at time step i . It is necessary to introduce some new notation in order to talk about individual cells, analogous to the elements of a vector. In these settings, we will denote the time step s as an argument and use the subscript to refer to the cell number. Using this notation, the firing rate of each simulated cell at time step s was calculated as

$$t_i(s) = \rho(s) [t_i(s-1) + \beta t_i^{IN}(s)], \quad (13)$$

where $t_i^{IN}(s)$ is the input to cell i at time s . The form of the input will be discussed below. The quantity $\rho(s)$ is here a gain modulation factor. We have assumed that $\rho(s)$ is a function of the total network activity, not too dissimilar to the *in vitro* results of Chance et al. (2002). At each time step $\rho(s)$ was calculated according to

$$\rho(s) = \left\{ \sum_i [t_i(s-1)]^2 \right\}^{-1/2}. \quad (14)$$

Inclusion of the factor $\rho(s)$ constitutes a form of divisive inhibition (e.g. Chance & Abbott, 2000). Equations 13 and 14 bear more than a passing resemblance to Eq. 6. As in Eq. 6, $\rho(s)$ functions to keep the length of \mathbf{t}_i (nearly) constant.

The cellular-level simulation captures key properties of Eq. 6. This can be seen in Figure 6. Figure 6a shows the firing rate of one cell as a function of time step. After time 0, each of the other cells in the network was turned on one at a time with an input of β for one time step each. The activity of the cell decays exponentially as a function of time. The time constant of this decay depends on β . Figure 6b shows explicitly that the amount of decay depends on the input to the network. When no input is given to the network, there is no decay (time-steps 50-100 and after time-step 200).

In the simulations reported here, we used 220 integrator cells. Each cell received an input given by a Gaussian function representing a head direction cell with a preferred direction ϕ_i and standard deviation σ , weighted by β . To do this we first took the minimum absolute difference between the actual head direction at time s , $\phi(s)$, and the preferred direction of cell i , ϕ_i :

$$\phi_i^{\text{diff}}(s) := \begin{cases} |\phi(s) - \phi_i|, & |\phi(s) - \phi_i| < \pi \\ 2\pi - |\phi(s) - \phi_i|, & |\phi(s) - \phi_i| \geq \pi \end{cases}. \quad (15)$$

This defines the minimum angular distance between the actual head direction and the cell's preferred direction. Preferred directions for the different cells were evenly spaced each $2\pi/220$ radians. We can write an expression for the input to cell i at time s :

$$t_i^{IN}(s) = \|\mathbf{p}(s+1) - \mathbf{p}(s)\| \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-[\phi_i^{\text{diff}}(s)]^2}{2\sigma^2}, \quad (16)$$

where $\mathbf{p}(s)$ is the rat's observed position at time s and $\|\mathbf{p}(s+1) - \mathbf{p}(s)\|$ is just the distance the rat moved between successive observations. The value of σ was set to $\pi/6$ for each cell. This value was chosen to be roughly consistent with observed head direction cells, which have been shown to have a tuning curve that falls to baseline levels with a width of about 100° (Taube, 1998). The Gaussian expression generates a tuning curve for each cell as a function of that cell's preferred direction ϕ_i . This direct input from head direction cells predicts that entorhinal place cells should show selectivity in the open field. Across cells this input is sensitive to speed and head direction in such a way that it can be referred to as a velocity vector. In this way, the cellular simulation can be said to implement Eq. 6 with velocity vectors provided as input (Eq. 12).

In the following two subsections, we will use the cellular simulation to demonstrate that Eq. 6, coupled with a velocity vector as input (Eq. 12), is able to describe characteristic properties of the entorhinal place code observed in two broad domains of experiments. First we will treat the properties of entorhinal place cells in the open field. After that, we will treat phenomena observed on the W-maze.

Applications: The open field

Entorhinal cells exhibit several key properties while animals move through the open field. First, EC cells do show place-specific firing, although the place-modulation is considerably weaker than hippocampal cells in comparable tasks (Quirk et al., 1992). The place-modulated firing of cells in EC is also comparable across similar environments. This means that if an entorhinal cell tends to fire in, say, the Northwest corner of a square enclosure, it will also tend to fire in the Northwest quadrant of a circular enclosure (Quirk et al., 1992).¹⁴

Equation 6 with Eq. 12 predicts both of these properties. The existence of a place code is a consequence of the fact that the set of paths the animal takes to get to a given place should depend on where the animal is within an enclosure.

¹³ Gain control like that described in Chance *et al* (2002) operating on a set of integrator cells should keep some measure of network activity nearly constant over time. However, precise Euclidean normalization would require a very specific relationship between gain and inputs.

¹⁴ This property is also observed for hippocampal cells early in the animal's experience with different environments (Lever et al., 2002), although with a sufficient amount of experience firing becomes uncorrelated across enclosures of different type (Muller & Kubie, 1987).

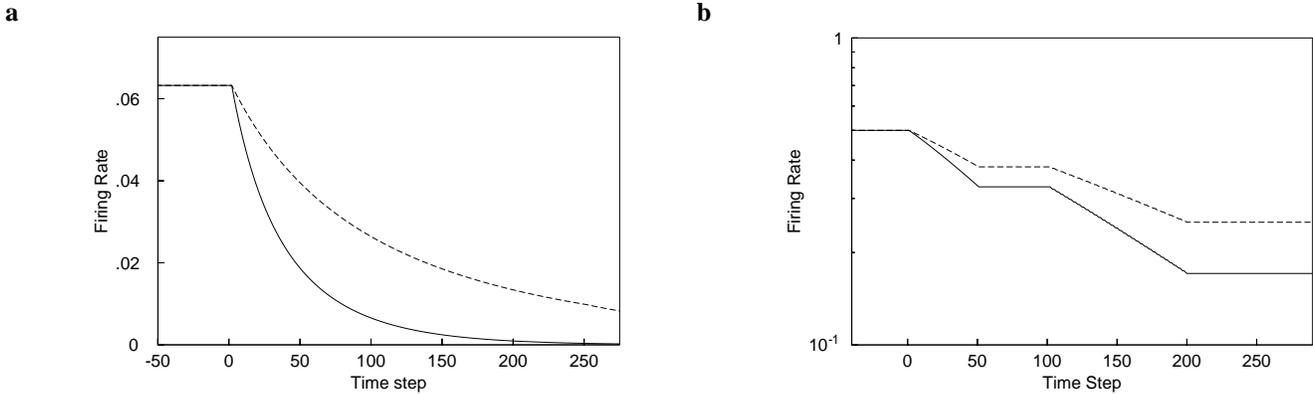


Figure 6. The cellular simulation implements key properties of Eq. 6. A network of 300 integrator cells was prepared. These figures show the firing rate of cell 1 as a function of time. **a.** At time 0, an input of β was provided sequentially to each of the other cells in the network, one at each time step. The solid line was calculated with $\beta = 0.2$. The dashed line was calculated with $\beta = 0.1$. The firing rate of cell 1 decays exponentially, with a time constant that depends on the value of β . **b.** At time steps 0-50 and 100-200, input was provided sequentially to each cell in the network, one at each time step. At the other times (time steps 50-100 and > 200), no input was provided. As in **a**, cell 1's firing rate decays exponentially when the other cells are being driven (note the logarithmic scale which makes exponential decay appear linear). However, the decay of cell 1's firing rate stopped when no input was provided.

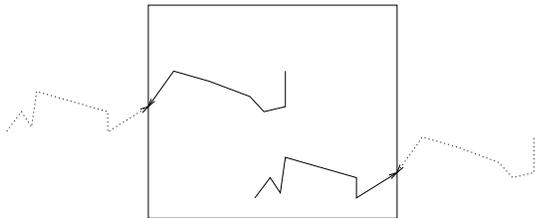


Figure 7. A weighted sum over recent movements predicts place-specific coding as a consequence of the kinematic constraints of the enclosure. The set of paths that lead to a position on the Western wall of the enclosure is different from the set of paths that lead to a point on the Eastern wall of the enclosure.

The similarity of the place representations across different environments follows if the set of paths leading to analogous positions in analogous enclosures are similar (see Figure 7).

Open Field Methods. The cellular simulation was presented separately with a series of positions and head directions collected by Lever et al. (2002) in a cylindrical environment and in a square environment.¹⁵ Position and head direction were sampled at 50 Hz for ten minutes. Place field maps were then calculated for the cylindrical and square enclosures using the simulated firing rates.

Simulated cells showed location-specific firing. Figure 8 shows place maps for four representative simulated cells in both the circular and square environments. These figures represent firing rate as a function of position averaged over the times the rat spent exploring the environment. Darker areas indicate higher average firing rates. Simulated cells showed regions of place-specific firing that extended over large sections of the environment. These regions were irregularly defined and apparently considerably more noisy than hippocampal place fields. This finding of noisy place-specific

firing is consistent with findings regarding place cells in EC in the open field (Quirk et al., 1992).

Topologically similar place fields across enclosures. The other primary finding of the simulation was that cells showed place fields in similar locations of topologically similar environments. That is, if a cell showed elevated firing in the Northwest quadrant of the circular environment, it also showed elevated firing in the Northwest quadrant of the square environment. This property has also been reported for cells in EC (Quirk et al., 1992). Topologically similar place fields across environments have also been observed in the hippocampus early in training (Lever et al., 2002). After a sufficient amount of experience, however, hippocampal cells will show place fields that are uncorrelated across enclosures (Lever et al., 2002; Muller & Kubie, 1987; Quirk et al., 1992). The reasons for this are not clear, but entorhinal cells with topologically similar place fields have been observed under conditions that also produce remapped hippocampal cells (Quirk et al., 1992), suggesting that remapping does not take place in EC.

The simulated cells showed place fields in a wide variety of locations. The only difference between cells was the preferred direction of their input. In all cases the preferred direction of the cell pointed in the direction of the cell's place field. For instance simulated cell 170 shown in Figure 8a has a place field along the Eastern edge of the circular and square environments, and a preferred direction that points toward the East. This is a consequence of the kinematic constraints of the environment. The Western wall of the environment cannot be reached using Easterly movements—this would require the animal to walk through the wall of the enclosure. This property depends to some extent on the value of β . If

¹⁵ As a check on the recorded head directions, we redid the simulations with head direction calculated from sequential movements and obtained the same pattern of results.

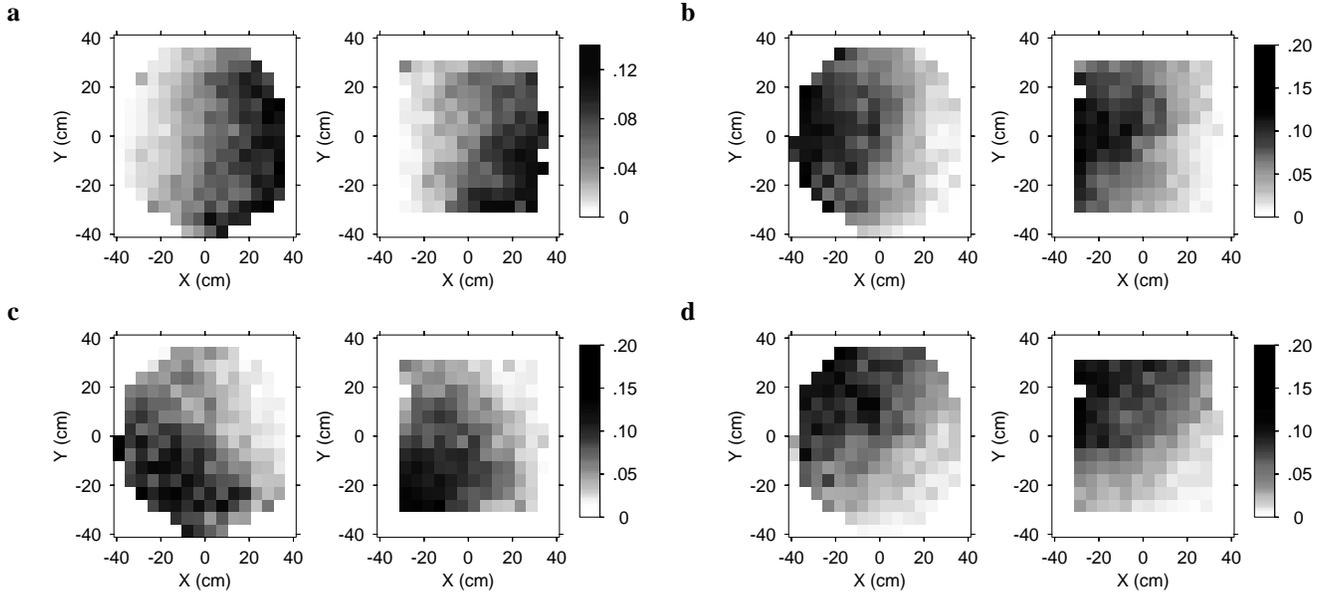


Figure 8. The cellular simulation shows place fields that are topologically similar in similar environments. Four representative cells from the simulation of motion in the open field. Paths were generated using positions and head directions from an experimental session presented in Lever *et al* (2002). The vast majority of cells showed apparent location-specific firing in both the circular and square enclosures. Like cells in EC described by Quirk *et al* (1992), the simulated cells showed large irregular place fields with a definite spatial correlate. Further, the simulated cells, like entorhinal cells observed by Quirk *et al* (1992), showed a high correlation between the place fields observed in the circular environment and in the square environment. The location of the place field is determined by the preferred direction of the input to the cell and the movements taken within the environment. **a.** Simulated cell 170 (preferred direction East). This cell fires preferentially in the East of the circular and square environments. **b.** Simulated cell 48 (W) fired preferentially in the Western edge in both environments. **c.** Simulated cell 20 (SW) fired in the southwest of both environments. **d.** Simulated cell 75 (NW) fired in the northwest of both environments.

β goes to one, the model should behave like a set of head direction cells. In this case the cells would fire preferentially in whatever location the animal assumed a particular head direction. Presumably, the animal would be less likely to point toward the East when positioned along the Eastern wall.

Applications: The W-maze

In the previous subsection, we saw that a representation of temporal context (Eq. 6), if driven by self-motion information as inputs, can capture key properties of the entorhinal place code in the open field. The “place code” derived from Eq. 6 did not directly represent place, *per se*, but rather reflected a sensitivity to the sequence of movements leading up to the current position. This treatment of the open field leads to strong predictions when the sequence of movements leading to a particular position is carefully controlled. These predictions can be readily tested within the maze paradigm used by Frank *et al.* (2000).

The W-maze (see Figure 4a) enables one to examine situations where location (and heading) are controlled but the sequence of movements leading up to that position are varied. Under these circumstances, a large proportion of entorhinal cells show *retrospective coding*, differentiating these two cases. It is also possible to compare situations in which a similar series of movements occur in different spatial locations. A sizable proportion of entorhinal cells exhibit *tra-*

jectory coding, showing similar firing in response to similar sequences of movements. Both of these effects are consistent with an entorhinal representation that responds to the sequence of movements leading up to the present location, rather than place, *per se*. Here we make explicit that these experimentally observed phenomena are indeed predictions of Eq. 6.

W-maze Methods. The cellular simulation was driven with positions and head directions from a segment of data lasting a little over twenty minutes, sampled at 30 Hz.¹⁶ This data was included as part of the study of Frank *et al.* (2000). After the simulation was completed, we calculated the firing rate maps separately for four types of trips named on the basis of the arm of the W-maze the trip started and finished on: center-left, left-center, right-center and center-right. Trip identity was provided at each time step, so that the values were identical to those used in Frank *et al.* (2000). Three trips in which the animal started on the center arm and crossed over to the wrong arm before reversing were eliminated from the path analyses (although not from the simulation itself). Two center-left trips and one center-right trip were excluded

¹⁶ This sampling rate is different than that used in the open field data. Although we might have down-sampled the open field data to equalize the sampling rate across simulations, this is not a concern because the change in the context (place) vector is driven by the animal’s movements, rather than time *per se*.

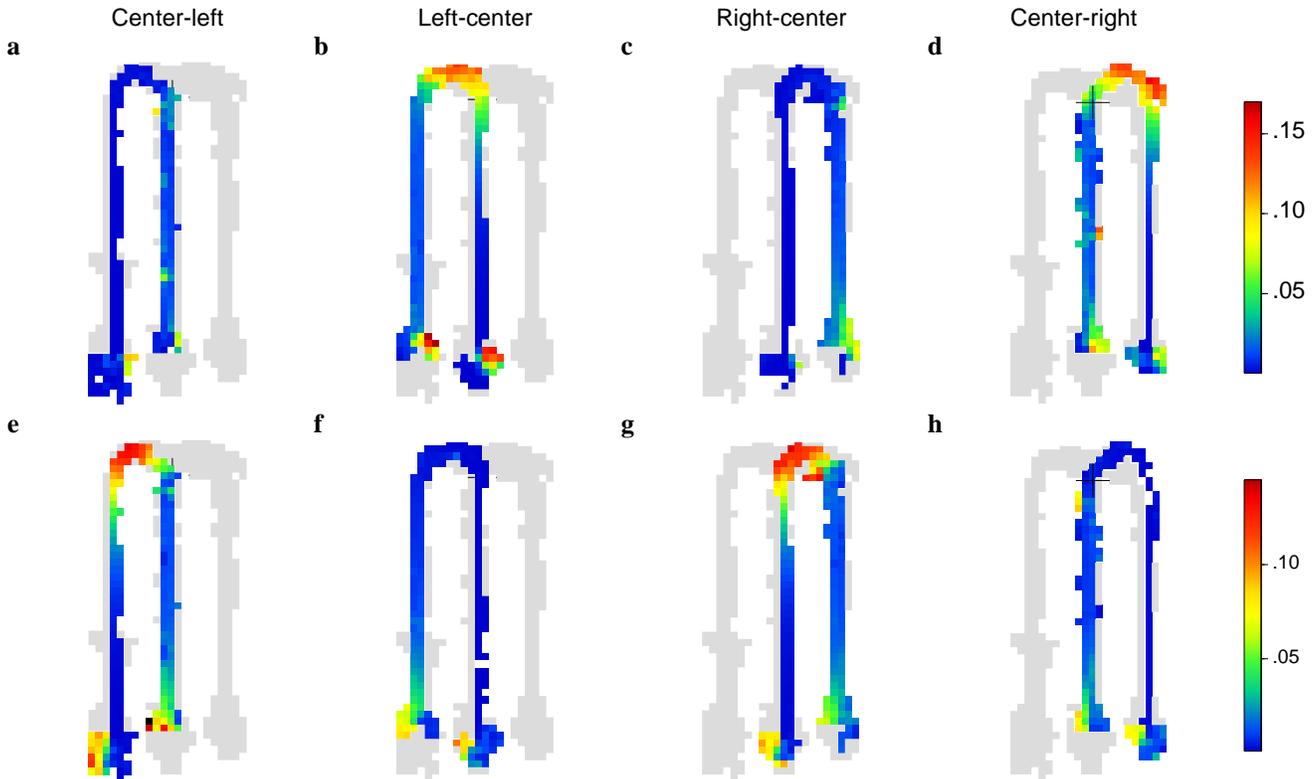


Figure 9. Trajectory coding in the cellular simulation. Average firing rate as a function of position for different trips in the W-maze. The trip each map corresponded to is indicated at the top of each column. Positions the animal did not visit on a particular trip are colored grey. **a-d.** Simulated cell 64 (preferred direction east) fired preferentially on left-center and center-right trips. The scale bar for this cell is shown to the right of **d.** **e-h.** Simulated cell 155 (preferred direction Northwest) fired preferentially on right-center and center-left trips. The scale bar for this cell is shown to the right of **h.**

in this way. In addition, there were some small gaps (typically one or two samples) in the position record. These were filled in with linear extrapolation of both position and head direction.

Trajectory coding. Numerous simulated cells showed evidence for trajectory coding. Figure 9 shows two examples. The figure shows firing rate maps separately for four different trips. The cell in the top row (**a-d**) had a preferred direction close to due East, such that it fired on left-center and center-right trips. The cell on the bottom (**e-h**) had a preferred direction toward the Southwest and fired on right-center and center-left trips. In general, almost all the cells observed showed some type of trajectory coding. Cells with preferred directions toward the North or South showed place fields that extended along the length of the arms on appropriate paths. This finding is consistent with the observation that place fields observed in EC are longer than those in the hippocampus (Frank et al., 2000), and that elongated fields tended to be observed on the long arms of the W-maze (L. Frank, personal communication).

Retrospective/prospective coding. To determine if the model showed retrospective and prospective coding in a way that is comparable to the available data, we also undertook

an analysis closely analogous to that used by Frank et al. (2000) in classifying cells as retrospective or prospective. In addition to actual position, we were provided with position projected onto a linear path along the track. We first made a firing rate map with bins of 6 cm as a function of linear position, using the firing rates from the actual navigation data. We then plotted each mean firing rate as a function of distance from the start of the trip. In accordance with methods used in Frank et al. (2000), we constructed an analogue of the Frank et al. (2000) study's Figure 4. That study used a Gaussian kernel with a standard deviation of one bin to smooth the curves. We smoothed using a moving window of 4 bins. Figure 10 shows representative retrospective and prospective cells from this analysis. All of the cells that showed retrospective or prospective coding also showed evidence for trajectory coding.

Many cells showed retrospective coding. Many of these cells only showed a difference in firing in regions of the center arm near the choice point. Figure 10a shows an example of such a cell. This cell had a tuning curve with a preferred direction that pointed toward the South East. The animal's head frequently pointed in this direction just before the choice point on the left-center path (top). This elevated firing persists along the center arm because of the exponential decay of activity. The tuning curve of this cell was such that

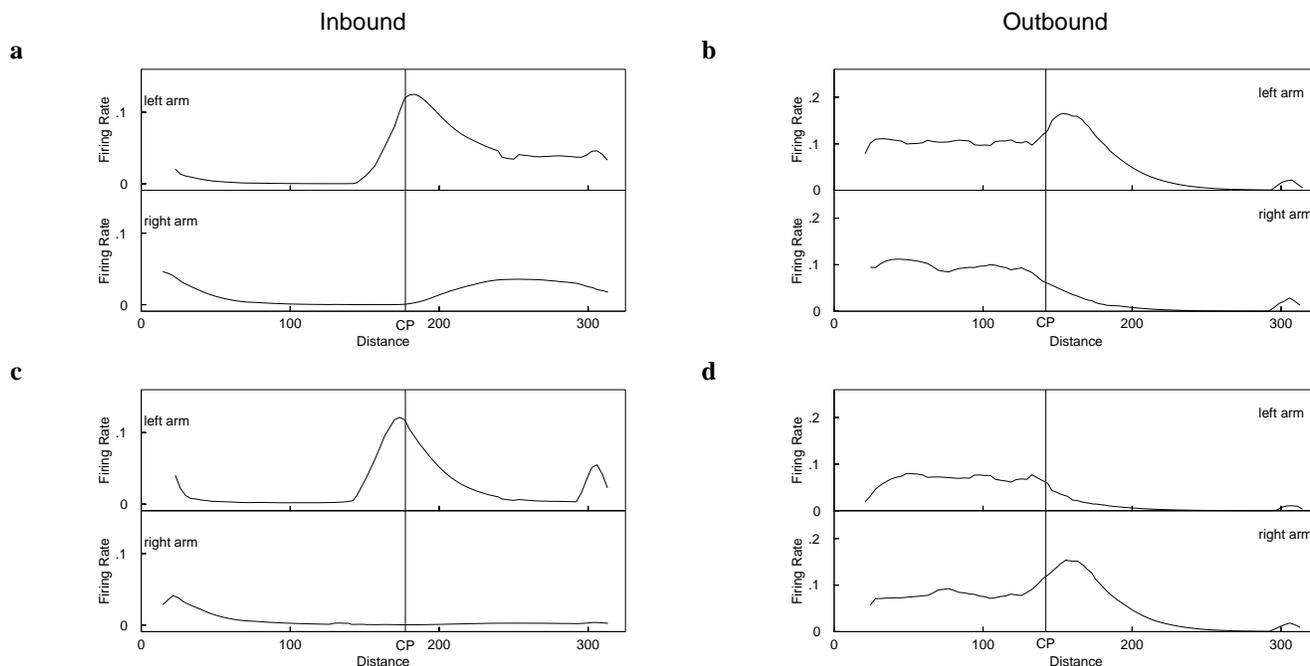


Figure 10. The cellular simulation showed retrospective and prospective cells. Firing rate is shown as a function of distance from the end of the center arm for different trips in the W-maze. The label “CP” indicates the location of the choice point defined by Frank *et al* (2000). Panels **a** and **c** show retrospective cells. Panels **b** and **d** show prospective cells. **a**. Simulated cell 30 (preferred direction E) showed a peak in firing just after the choice point for left-center trips. This cell showed differential firing after the choice point, but comparable firing toward the end of the center arm. **b**. Simulated cell 130 (NNW) showed elevated firing along most of the length of the center arm on both journeys to the right arm and journeys to the left arm. On center-left trips the cell showed elevated firing that began shortly before the choice point. In contrast, the cell showed depressed firing just before the choice point on center-right trips. The curves diverge about 10 cm before the choice point. **c**. Simulated cell 50 (ESE) showed a peak firing rate shortly before the choice point for left-center paths. Elevated firing lasted most of the length of the center arm. **d**. Simulated cell 83 (NNE) showed a pattern comparable to cell 130, except the elevation in firing came on center-right trips.

it overlapped slightly with the head direction associated with “Southward” travel down the center arm. This resulted in a gradual buildup of firing rate on the right-center path (bottom). As a consequence, the firing rate was similar across paths toward the food end of the center arm (toward the right end of the figure). The cell in Figure 10b shows a cell that also showed retrospective coding, but with a somewhat different profile. The preferred direction of this cell was typically obtained slightly further from the choice point than the cell in Figure 10a. As a consequence, there was much less overlap with the head directions typically obtained on the center arm, so that there was no visible elevation in firing on the right-center paths. As a consequence, this cell showed a higher firing rate for left-center paths than right-center paths over most of the length of the center arm. A smaller number of cells showed this type of retrospective coding compared to the pattern illustrated in Figure 10a. We observed comparable numbers of retrospective cells that preferred left-center trips and right-center trips. This was predictable given the even spacing of preferred directions across cells.

A small number of cells also showed some evidence for prospective coding. The cells shown in Figure 10c-d are particularly strong examples of these cells. These cells show

a peak in firing along one of the two arms after the choice point. However, the increase in firing leading to the peak starts reliably before the choice point when approaching the area of the peak, and decreases reliably when approaching the arm without the peak in firing. We only observed prospective coding immediately prior to the choice point.

Prospective coding is something of a misnomer for this model. The TCM evolution equation (Eq. 6) contains no information about future events. It is somewhat paradoxical that these cells can show firing that diverges based on what is about to happen. These simulated cells are actually responding to small variations in head direction that happen shortly before the choice point. These simulated cells have preferred directions that point along the “curve” in the path between arms. This is why the peak of firing is observed on one arm. As can be seen in these cells, there is also firing on the center arm for both paths. This is so because the tuning curve overlaps somewhat with the “due north” direction associated with moving up the center arm. Relatively small changes in head direction prior to the choice point come in a part of the tuning curve that is relatively steep, resulting in relatively large changes in simulated cellular activity from small changes in head direction.

Trajectory coding and retrospective coding were robustly observed in the simulation. In contrast, the prospective coding we observed was considerably more fragile and even the best prospective cells we found (Figure 10c-d) were not nearly as impressive as the striking prospective coding shown by the deep EC cell shown in Figure 4 of the Frank et al. (2000) paper. A larger β resulted in more cells showing prospective coding, as β controls the time constant for the rise in firing as well as the decay. Also, cells with more robust prospective coding were observed when the tuning curves were made more narrow. This was not adopted to keep with experimental findings about the width of the tuning curves in head direction cells (Taube, 1998). However, some sort of lateral inhibition process could result in a sharper effective tuning curve for entorhinal cells. If so, this would amplify the “prospective” coding generated by this purely retrospective model.

Discussion

We showed that the same equations that govern the evolution of temporal context in a model of human episodic memory performance also describe the activity of cells in EC during spatial navigation. We showed a consistency between the simulated cells and entorhinal cells during navigation through the open field and in the W-maze. In the open field, simulated cells had large, noisy place fields that were consistent across topologically similar environments (Figure 8). This is consistent with what is known about entorhinal cells in the open field (Quirk et al., 1992). In the W-maze, simulated cells showed evidence for trajectory coding (Figure 9), as well as retrospective and prospective coding (Figure 10). These are consistent with observations of entorhinal cells in the W-maze (Frank et al., 2000).

This theoretical connection between human memory and the place cell literature is especially timely in light of recent findings that suggest place cells exist in humans. Ekstrom et al. (2003) examined the activity of single units at various locations in patients being treated for pharmacologically resistant epilepsy during performance of a virtual navigation task. A number of cells showed virtual-place-specific firing. Notably, these cells were clustered in the hippocampus and the rhinal cortex. These findings support the view that there is more than a computational similarity between the EC function of rats and humans.

Ways in which the simulation could be more realistic. The cellular simulation is remarkably simple—it only includes enough detail to implement the important properties of Eq. 6 given velocity vectors as input. Nonetheless, it is apparently sufficiently rich to robustly describe the major phenomena demonstrated for entorhinal cells during navigation through spatial environments. Here we discuss some straightforward additions that would make the simulation more realistic.

Our implementation of integrator cells was limited in at least two respects. First, Egorov et al. (2002) showed that their integrator cells did not initiate sustained firing with a sufficiently small input. Similarly, if sufficiently hyperpolarized, integrator cells shut off and remained off. The simu-

lated integrator cells used here did not include this type of thresholding behavior. If thresholding were included in the simulations it would add to the realism of the model by preventing situations where a long-lasting period of low activation was produced (e.g. the firing on the center arm in Figure 10c-d). Another aspect of the Egorov et al. (2002) study that we neglected was the rate of change of firing rate in response to an input. Here we assumed that the response of the integrator cells to an input was essentially instantaneous. While this probably would not have had a negative effect on the open field results, or on the phenomena of trajectory or retrospective coding, this could have had a negative effect on the ability of the model to generate prospective coding.

Prospective coding could be a consequence of hippocampal inputs to EC, which were completely neglected in the current treatment. Consistent with this view, Frank et al. (2000) suggested that prospective coding was more frequent in EC layer V (which receives hippocampal input) than in superficial layers (which do not). Indeed, Muller and Kubie (1989) argued that the hippocampus does not actually code for the animal’s current position, but rather its position approximately 120 ms in the future (but see Breese, Hampson, & Deadwyler, 1989). Prospective coding could have been easily and robustly implemented if the inputs to the simulated entorhinal cells included information from cortical areas that contained information about future movements.

Another simplification in the current simulation was the lack of any inputs other than velocity information. It is clearly within the framework of the current model that non-spatial stimuli should contribute to firing in EC. For instance, we assume that t_i is driven by non-spatial inputs t_i^N in both episodic applications (Howard & Kahana, 2002a; Howard et al., In revision; Howard, 2004) and in the relational memory simulations presented in the next section of this ms. Although the fact that the current, highly simplified, treatment of the EC did remarkably well in describing the basic phenomena of the entorhinal place code, there are several situations where including other types of stimuli could have improved the model’s performance. For instance, the simulated trajectory coding cells occasionally showed elevated firing near the food wells (e.g. Figure 9). Although not explicitly addressed in Frank et al. (2000), none of the representative place field maps presented in that study showed such an effect. In the model, firing near the food well is a consequence of the wide range of directions the animal assumes as it turns around. This elevated firing rate would have been attenuated if we had included cells tuned for proximity to food reward. These “chocolate milk cells” would have been strongly activated and inhibited (divisively) other cells receiving head direction input when near the food wells. Along these lines, Gothard, Skaggs, and McNaughton (1996) reported “goal box cells” in the hippocampus that fired as the rat approached a movable goal location on a linear track.

Another, extremely important aspect of the functioning that we have ignored here is the basis of the head direction system. The model’s description of topologically similar place fields in circular and square environments is only valid insofar as cells in the head direction system maintain

the same preferred direction across enclosures. The empirical situation with regards to this is somewhat unclear. Taube, Muller, and Ranck (1990) reported that 3/8 head direction cells studied in both the cylinder and the square enclosures showed a change in their preferred direction of more than 48° across enclosures. Golob and Taube (1997) reported that only 2/11 cells (in animals with lesions to the hippocampus) reported a change of greater than 18°, and concluded that in general minimal changes take place across cylindrical and square enclosures, a conclusion that they regarded as consistent with the Taube et al. (1990) results. In any event, the present treatment of the place code predicts that the entorhinal place code should rotate in register with the head direction cells that provide its input.

For the present model to make any predictions at all, it is necessary to first specify the activity of the head direction system. This requires that the head direction system accurately integrates from moment to moment within an environment, a process believed to rely on attractor dynamics (Redish, Elga, & Touretzky, 1996; Zhang, 1996). This is in sharp contrast to the “pseudo” integrator proposed here to support the entorhinal place code. Further, the head direction system is reoriented by manipulations of visual stimuli (Taube et al., 1990), a process believed to result from a sufficiently divergent visual input causing a reset of the attractor network into a state distant from its predecessor (Skaggs, Knierim, Kudrimoti, & McNaughton, 1995, see chapter 5 of Redish, 1999, for a review and discussion).

Deep vs superficial cortical layers. We implemented the cellular simulation of Eq. 6 using known properties of cells in EC layer V (Egorov et al., 2002), the deep cortical layer. These cells receive input from, but do not project to, the hippocampus proper. In addition to layer V cells, EC also contains principal cells in layers II/III, the superficial layers of EC. Layer II cells project directly to the hippocampus. Although they do not receive input from the hippocampus directly, they do receive input from layer V cells in EC, so they are indirectly connected to hippocampus.

The cells reported in the open field by Quirk et al. (1992) were identified as being from superficial layers of EC. In that study, a small number of layer V cells from deep layers of EC were observed, but the spatial firing characteristics of these cells were not described. There are several issues that have bearing on the validity of “mixing layers” across experiments. There is no published data on EC layer V cells in the open field. However, Frank et al. (2000) showed similar qualitative properties for deep and superficial layers of EC, although superficial cells showed less positional information (were noisier) and may have shown less prospective coding than deep cells. This qualitative similarity suggests that perhaps similar properties would obtain for deep and superficial cells in the open field as well.

Similarities in the firing properties of deep and superficial cells could reflect a direct physiological connection or a parallel computational function. As pointed out earlier, EC layers II/III receive input from layer V, so perhaps this is the origin of the spatial properties of superficial EC. Although

cells in the superficial layers of EC do not show the striking integrator cell behavior observed in layer V *in vitro*, cells in superficial EC do show plateau potentials in response to inputs that persist for a relatively long time (Klink & Alonso, 1997). These plateau potentials have been argued to support cellular responses observed in DNMS tasks (Fransén, Alonso, & Hasselmo, 2002) and could be sufficient to support something sufficiently similar to Eq. 6 to result in similar place-specific activity. There are may also be other mechanisms by which a leaky integrator could be implemented in superficial EC.

Can Eq. 6 support the hippocampal place code? The hippocampus is said to support a representation of place insofar as cells in the hippocampus correlate with the animal’s location. If the activity of these cells correlated perfectly with the animal’s location in allocentric space, then this representation could be said to be a perfect representation of place. The representational scheme pursued here is not a “perfect” representation of place, but then again neither is the hippocampal place code. Directional firing of cells on the linear track (McNaughton, Barnes, & O’Keefe, 1983) are a clear example of a situation in which place cells’ responses are different despite the animal being in the same place. Path-dependence (Frank et al., 2000; Wood et al., 2000), including retrospective coding, is another such example of a situation where position is not sufficient to predict the firing of “place cells.” Similarly, the finding that the responses of place cells depend on the behavioral context (Markus, Qin, Leonard, et al., 1995) and the finding that similar responses take place in different environments (Lever et al., 2002) argue against a perfectly accurate hippocampal place code.

Having said that, the hippocampus can show remarkable spatial precision. The location of the animal in a familiar open environment can be reconstructed from examining place cell activity to a precision comparable to the error in recording the animal’s position (Wilson & McNaughton, 1993). Exceptionally good positional reconstruction can be found when recording from cells during navigation on the linear track (Jensen & Lisman, 2000). Can the precision of the hippocampal place code be derived from the systematically-imperfect representation of place that results from Eq. 6? A definitive answer must await further experimental and theoretical investigation. However, the present treatment of the entorhinal place code predicts that it should be possible to reconstruct position to sufficient precision using the history-dependent firing scheme presented here. On a linear track where movements are relatively constrained, it should be possible to get very good precision. In the open field, very good reconstruction is theoretically possible if the decay of velocity information is sufficiently slow.

Although position is a correlate of the cells in the simulation presented here, it would be fair to say that Eq. 6 doesn’t really support a positional representation at all. The weighted sum over recent movements of Eq. 6 should retain sensitivity to head direction specifically, and trajectory more generally in the open field. To be explicit, the present model predicts that the firing of entorhinal place cells should be modulated

by not only the head direction, but preceding head directions as well. At the values of β used here, we also observed that cells' preferred directions point in the direction of their place fields. This would be a marker of a history-dependent pseudo-place code like the one we have hypothesized resides in entorhinal cortex.

A number of models of hippocampal function assume that entorhinal place cells should be directional in the open field (e.g. Brunel & Trullier, 1998; Kali & Dayan, 2000). These properties for entorhinal place cells would need to be reconciled with the lack of a strong directionally selective signal in hippocampal place cells in the open field. Although hippocampal place cells show directional selectivity (e.g. Skaggs, McNaughton, Gothard, & Markus, 1993), this can be accounted for by taking into account the different amount of time the animal spends in different locations with different head directions (Muller et al., 1994). One possibility is that the hippocampus transforms directional inputs in such a way that it shows omnidirectional place fields in the open field. Mechanisms for this have been proposed by other authors (Brunel & Trullier, 1998; Kali & Dayan, 2000; Sharp, 1991).

An intriguing possibility is that the hippocampal place code really is dependent on head direction, but that it is not reflected in firing rate. Directionality could be retained at the ensemble level if theta phase is taken into account. Theta phase precession (O'Keefe & Recce, 1993) has been observed in the open field (Skaggs et al., 1996). Phase precession refers to the finding that when the animal initially enters a cell's place field, it fires at a later phase relative to the hippocampal theta rhythm than it does as it moves through the place field. This can, in principle at least, be used to reconstruct velocity as well as position, as the following thought experiment will illustrate.

Imagine two hippocampal cells with place fields in an open enclosure. Cell A and cell B have symmetric overlapping place fields. The center of field A is due west of field B. Burgess, Recce, and O'Keefe (1994) showed that place cells in the open field fire at a late theta phase when the cells' field center is in front of the rat, and at early phases when the cells' field center is behind the rat. Let's assume that the animal moves West to East on a path that crosses through the center of field A and then the center of field B. Consider the theta phase of A and B at the halfway point. Cell A should fire at an early phase, because the center of its place field is behind the animal. On the other hand, cell B should fire at a late phase because its field is in front of the animal. What if the animal makes the trip in the opposite direction? Now, the moves from East to West, passing first through the center of field B and then the center of field A. In this case the phase of the cells at the midpoint will be reversed. Now, cell A should fire at a late phase because the center of field A is in front of the animal, whereas B should fire at an early phase because the center of its field is behind the animal. The phase of firing of these cells is reversed relative to the situation in which the animal moved West to East, despite the fact that the animal is in the exact same position. What differs in these two cases is the animal's velocity. We conclude that theta phase could in principle be used to reconstruct velocity in the open field.

It should be pointed out, however, that the mechanism by which integrated head direction inputs in entorhinal cortex could give rise to theta phase coding of movement direction is not at all clear at this time.

We mentioned previously that the model presented here predicts that firing of entorhinal place cells should depend on the recent history of movements in the open field. An analogous prediction can be made regarding non-spatial stimuli. In a homogeneous list of to-be-remembered non-spatial stimuli, firing of entorhinal cells should depend not only on the current stimulus, but also on prior stimuli as well. In fact, Suzuki et al. (1997) showed that stimulus-specific entorhinal cells fired across several intervening stimuli in a working memory task, supporting at least the general thrust of the prediction. The predictions of the model, however, can be quantified and extend to experimental situations in which the stimulus in question would not be expected to be actively maintained in working memory.

The role of the hippocampus in relational memory

The previous section argued that a component of TCM, Eq. 6, describes a key computational function of the entorhinal cortex, and perhaps other extra-hippocampal MTL structures as well. In this section we argue that new item-to-context learning is supported by the hippocampus. As mentioned earlier, this process results in reinstatement of patterns in parahippocampal regions in response to the item being repeated (Figure 3). We will see that disrupting new item-to-context learning predicts neuropsychological dissociations observed with hippocampal damage. In the model, new item-to-context learning also causes representational changes that have been directly observed in extrahippocampal MTL areas and that may result from hippocampal function. We will discuss the utility of representations that result from new item-to-context learning in capturing relationships between temporally disparate stimuli. This corresponds to the development of a higher-order stimulus representation in parahippocampal regions.

Modeling transitive association using TCM

Eichenbaum and colleagues have argued that the hippocampus supports relational memory (Cohen et al., 1997; Eichenbaum, 2001). In contrast to extrahippocampal areas that are said to be capable of forming simple pairwise associations, the hippocampus supports the ability to discover and encode higher-order relationships among stimuli. The canonical example of this proposed hippocampal function is the formation of *transitive associations* between items that were never paired during training (Bunsey & Eichenbaum, 1996).

Bunsey and Eichenbaum (1996) examined the effect of hippocampal damage on transitive associations. In their task, animals were first presented with a cue odor. The identity of the cue odor predicted which of two choice odors would be paired with reward. There were two cue odors, each of which predicted reward for one of the two choice odors (see

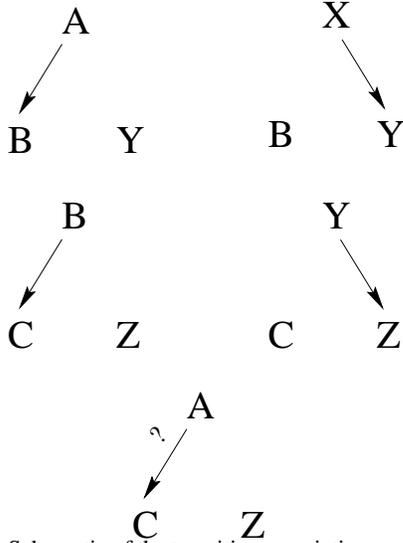


Figure 11. Schematic of the transitive association experiment used by Bunsey and Eichenbaum (1996). In an initial learning phase, animals learned to choose between choice odors (B and Y) depending on which cue odor was presented. The effect of this training was to form an association within each of the pairs, indicated by the arrows. In a second learning phase, the choice odors from the first learning phase became cues used to discriminate between another pair of odors. In a third phase, the animals were tested for their generalization across learning phases. In this transfer phase, animals were given cues from the first stage and choices from the second stage. Animals were tested, in the absence of reward, for their preference of the choice that would result if they formed a transitive association across phases (arrow with question mark). Although animals with hippocampal damage learned as well as controls on each of the learning phases, they were impaired at the transfer stage.

Figure 11). Two associations, $A \rightarrow B$ and $X \rightarrow Y$ were thus simultaneously established during the first stage of learning. Following the first stage of learning, the choice odors became cue odors for a second pair of associations. In this second stage, associations $B \rightarrow C$ and $Y \rightarrow Z$ were trained. In a final stage, transitive association was probed; the animals were presented with a cue from the first stage, and tested for their preference for the choices from the second phase. This probe phase tested for the existence of an association that “bridged” across B from A to C . Although rats with hippocampal damage learned each of the premise pairs, $A \rightarrow B$ and $B \rightarrow C$, they showed no evidence for a transitive association from $A \rightarrow C$. This is consistent with the hypothesis that the hippocampus, while not required for simple pairwise associations, is required for higher-order transitive associations. The hippocampus was apparently important in learning the relationship between A and C , which were never actually presented together, but were presented in the same temporal context, B .

Here we will show that using the theoretical framework offered by TCM, transitive associations can be selectively impaired, while leaving the ability to learn pairwise associations intact. This is accomplished by disrupting the ability of the model to bind items to their temporal context; by setting

α_N to 0.

Of key interest is the effect of the relative contribution of old and new context to Eq. 9. We will examine two extreme values for the ratio $\gamma := \alpha_N / \alpha_O$. In the “intact” case, $\gamma = 1$. For the intact case, old and new retrieved contextual components contribute equally to \mathbf{t}^{IN} . This is in the range of values that have been used in the past to describe human episodic recall data.¹⁷ In the “lesioned” case, representing the hypothesized effect of hippocampal lesions, $\gamma = 0$. Although the magnitude of \mathbf{t}^{IN} is the same in both cases, they differ in that the intact case allows new item-to-context learning ($\alpha_N > 0$), whereas the impaired case does not ($\alpha_N = 0$).

Previously we argued that simulating a hippocampal lesion by setting $\alpha_N = 0$ would selectively impair backward associations (see Fig. 2a). In fact Bunsey and Eichenbaum (1996) found that hippocampal lesions do selectively impair backward associations. In this section, we are interested in the ability of the model to develop and utilize transitive associations. To ensure that neither recency effects nor across-pair temporal associations enter into these analyses and simulations, we will assume that an infinitely long delay intervenes between pairs, and between study and test, effectively isolating the pairs from the rest of experience.

Consider the case in which a pair of stimuli A, B is presented, then a pair B, C is presented. If α_N is greater than zero, then when B is presented the second time, it will retrieve elements of the context retrieved by A . As a consequence, when learning B, C , the model is also in effect learning A, C as well. If $\alpha_N = 0$, however, \mathbf{t}_A^{IN} can still be associated to B , and \mathbf{t}_B^{IN} can be associated to C . However, there will be no transitive association between A and C . Appendix D explicitly derives the cue strength between A and C when A is presented during a recall test after presentation of A, B and B, C . From Appendix D we find that after both stages of learning the cue strength to item C given A is given by:

$$a_C = \alpha_N \rho^2 \beta (\alpha_O \beta + \alpha_N). \quad (17)$$

We can see from this expression that the cue strength is zero if α_N is zero. The transitive association from A to C , items that were never presented together, depends on a non-zero value of α_N , which we hypothesize corresponds to an intact hippocampus. In contrast, as derived in Appendix D, we find that the cue strengths from A to B and from B to C do not depend on a non-zero value of α_N . This is possible because forward associations do not depend on new item-to-context learning (see the curve labeled “old” in Figure 2a).

As a complement to the derivation presented in Appendix D, we also carried out a simulation. The goals of the simulation are to demonstrate the ability of this theoretical framework to describe the dissociation between learning of pair-wise associations from transitive associations more detail and under more realistic conditions.

¹⁷ In some cases, γ has been fixed at one (Howard & Kahana, 2002a), whereas in other cases, γ has been allowed to vary as a free parameter (Howard et al., In revision).

Transitivity Simulation Methods.

The equations for \mathbf{t}_i and \mathbf{t}_i^{IN} are typically assumed to describe infinite-dimensional vectors. How should we go about implementing an infinite-dimensional vector space? On the one hand, we might have chosen some large number to represent the dimensionality of the space and chosen random vectors to describe the \mathbf{t}_i^{IN} 's when items are first presented. These vectors would have been asymptotically orthogonal if the number of dimensions had been much larger than the number of vectors. To eliminate any concerns that might arise from random variability in choosing patterns, we adopted an alternative approach that has been used in previous simulations applying TCM to human data (Howard & Kahana, 2002a; Howard et al., In revision). The true dimensionality of the space is the dimensionality of the actual input vectors, which can be infinite. However, if the initial input vectors \mathbf{t}_i^{IN} are orthogonal, then they can be used as basis vectors to span the relevant parts of the space. In the simulations, we express \mathbf{t}_i and \mathbf{t}_i^{IN} as vectors of coefficients of the basis vectors. This greatly reduces the dimensionality of the simulations. It also makes it particularly easy to introduce an infinite delay. To introduce an infinite delay, all that needs to be done is set \mathbf{t}_i to one times a basis vector that has not yet been used.

Matrices corresponding to \mathbf{M}^{TF} and \mathbf{M}^{FT} were maintained. The matrix \mathbf{M}^{TF} was updated when a particular item was presented simply by adding the current state of \mathbf{t}_i to the appropriate column of \mathbf{M}^{TF} . The matrix \mathbf{M}^{FT} was somewhat more complicated. First α_N and α_O were calculated according to the procedure in Appendix B. Then, after \mathbf{t}_i was calculated \mathbf{M}^{FT} was updated according to

```
for ( i = 0 ; i < currentdim ; i++){
  MFT[item][i] = aold * MFT[item][i] + anew * t[i];
}
```

where “item” is the index of the stimulus presented, “currentdim” is the number of basis vectors that have been presented up to that time and “anew” and “aold” are calculated according to the assumptions of the simulation and the constraint that the Euclidean length of “MFT[item]” should be one after the updating. This enables \mathbf{M}^{FT} to implement Eq. 9. Note that “synapses” that do not connect to the current item are unaffected.

Learning. The model was presented with two phases of learning. During the first phase, A or X were presented randomly, and the model had to choose either B or Y as a response. If A was the cue stimulus, then B was considered the correct response. Similarly, if X was the cue stimulus, then Y was considered the correct response. Although only one cup was baited with a food reward, Bunsey and Eichenbaum (1996) allowed the animal to dig in the other cup if it initially dug in the incorrect cup (although the first cup was counted as the response for that trial). Using a similar procedure, if the model made the correct choice, then the stimulus corresponding to the correct choice was presented to the model and the trial ended. If the model made an incorrect choice, then the stimulus corresponding to that choice

was presented to the model. Before the trial ended, however, the correct choice stimulus was presented to the model. This simulates the experimental method that allowed the animal to dig in the correct cup after choosing incorrectly (Bunsey & Eichenbaum, 1996). For A trials, there are therefore two possibilities. If the animal chose correctly, A was presented, followed by B . If the animal chose incorrectly, A was presented, followed by Y and then B . In both cases, there is an increment to the cue strength from A to B , but only when there is an incorrect response is there an increment to the cue strength from A to Y . As a consequence, as long as the animal chooses more or less randomly during the initial of learning trials, $A \rightarrow B$ develops more strongly than $A \rightarrow Y$. Similar reasoning describes the development of $X \rightarrow Y$ over $X \rightarrow B$, and also applies to the second stage of learning.

Choice situations. The model was presented with choices during each trial of learning and during probe trials. At each choice, the probability of recalling the choice stimuli was calculated using Eq. 4. The sum in the denominator went over the two choice stimuli. The two choices were B and Y in the first stage of learning and C and Z in the second phase and in the probe trials.

Probe trials. After each ten learning trials in phase two, ten probe trials were presented. In these probe trials, \mathbf{t}^{IN} was calculated given either A or X as a stimulus. That is, \mathbf{t} was reset with an infinite delay and then updated with \mathbf{t}^{IN} set to $\mathbf{M}^{FT}\mathbf{f}_A$ or $\mathbf{M}^{FT}\mathbf{f}_X$, as appropriate. A choice was then made between C and Z . However, neither of the associative matrices, \mathbf{M}^{TF} nor \mathbf{M}^{FT} , were updated either when the cue stimulus was presented, nor when the response was selected. In this way, a probe trial would not affect either subsequent learning of the premise pairs, nor subsequent probe trials. Nonetheless, we could observe the process of learning in this situation, rather than just recording a single value at the end of each simulation run.

Results.

For each set of parameters, we repeated the simulation for 1000 random presentation orders. There was no systematic search of the parameter space. Rather, an informal search was undertaken to find a set of values that showed reasonable learning curves for $A \rightarrow B$ and $B \rightarrow C$. When this condition was met, the intact model always outperformed the lesioned model on $A \rightarrow C$ and the lesioned model never deviated significantly from chance. If τ was set too low, the model remembered whatever choice it happened to make on the first trial, even if it was incorrect. The parameter values used the simulation were listed in Table 1. Figure 12 shows results of the simulation. Figure 12a-b shows performance for the intact and lesioned model on first and second stage learning. Both the intact and the impaired model showed good learning on the premise pairs, $A \rightarrow B$ and $B \rightarrow C$, with responses tending toward perfect performance for both stages and both models. Figure 12c shows performance in the probe trials. Whereas the intact model showed generalization to

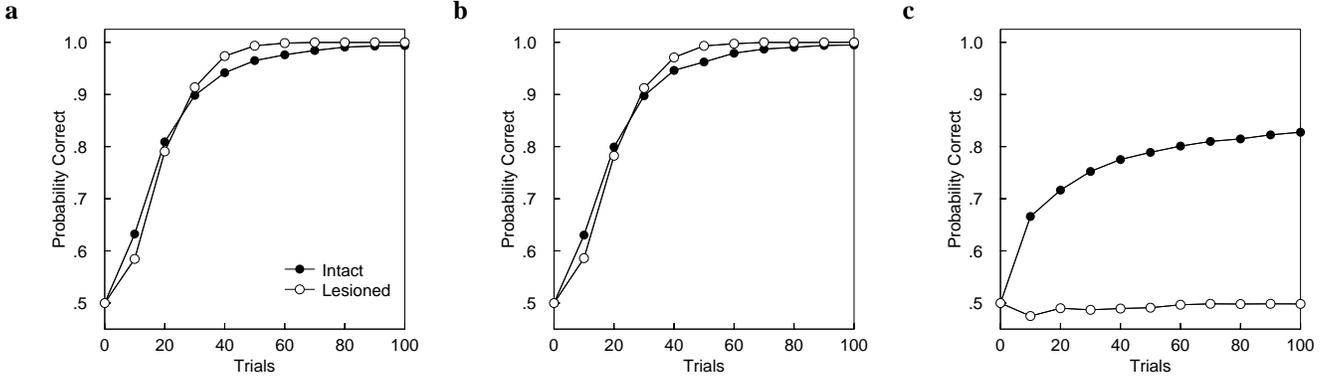


Figure 12. Impairment of new item-to-context learning specifically affects development of transitive associations. Performance as a function of learning is shown for the different stages of the Bunsey and Eichenbaum (1996) study. In all three panels, the intact model, with $\gamma = 1$ is shown with solid symbols; the lesioned model, with $\gamma = 0$ to simulate hippocampal damage, is shown with open symbols. **a.** Probability of a correct response during the first phase of learning. Both the intact and lesioned models learned the $A \rightarrow B$ and $X \rightarrow Y$ pairs. **b.** Probability of a correct response during the second phase of training. Both the intact and lesioned model learned $B \rightarrow C$ and $Y \rightarrow Z$. **c.** Performance on the probe trials, $A \rightarrow C$ and $X \rightarrow Z$. Probe trials were performed at each stage of learning, but in a way that neither subsequent second phase trials nor subsequent probe trials were affected. While the intact model develops a transitive $A \rightarrow C$ association, the lesioned model does not. This is consistent with the effects of hippocampal lesion observed by Bunsey and Eichenbaum (1996).

$A \rightarrow C$, the lesioned model did not. Whereas the intact model showed a dramatic improvement in the transitive association, the impaired model did not deviate significantly from chance, even with enough learning trials to acquire near-perfect performance on the premise pairs. From this we conclude that TCM provides a means to dissociate simple pairwise learning from relational learning, as evidenced by the phenomenon of transitive associations. This result also supports our hypothesis that the function of the hippocampus is to allow repetition of an item to allow the recovery of entorhinal activity patterns that were present when the item was previously presented.

Rapid development of an intermediate representation

Eichenbaum (2001, 2000) hypothesized that the hippocampus could accomplish many of the functions ascribed to it by forming a “memory space.” If the hippocampus could support the rapid development of a stimulus representation that captures the temporal and contextual relationships among stimuli, this representation would presumably be extremely useful in the “flexible re-expression” of memory (Eichenbaum, Otto, & Cohen, 1994; Cohen & Eichenbaum, 1993). Here we show that binding item representations to their temporal context, shown in the previous subsection to subserve backward associations and transitive associations, results in the rapid development of an intermediate representation that captures higher-order relationships among the stimuli. The mapping between TCM and the MTL argues that this intermediate representation should be located in parahippocampal regions.

Recovery of contextual states in parahippocampal regions results in a compressed stimulus representation.

In TCM, the inputs to Eq. 6, \mathbf{t}_i^{IN} , are caused by the particular item presented at time step i . We can think of \mathbf{t}_i^{IN} as an in-

termediate representation of the nominal stimulus presented at time step i (e.g. the word ABSENCE). We will explore the development of this representation in capturing higher-order relationships among stimuli. As before, we will consider two extreme cases. In the lesioned case, we will let $\alpha_O = 1$ and $\alpha_N = 0$. In the intact case, as in the previous subsection, $\alpha_O = \alpha_N$.

In the lesioned model,

$$\mathbf{t}_{A_{i+1}}^{IN} = \mathbf{t}_{A_i}^{IN}; \quad (18)$$

the input evoked by an item never changes. In the lesioned case, \mathbf{t}^{IN} is like a mirror that simply reflects the item currently being presented, \mathbf{f}_i . In the intact case, however, $\mathbf{t}_{A_{i+1}}^{IN}$ is composed of both $\mathbf{t}_{A_i}^{IN}$ and \mathbf{t}_{A_i} ; rather than simply mirroring the stimulus being presented, $\mathbf{t}_{A_i}^{IN}$ changes over time to reflect the temporal contexts in which item A is presented. This results in a “mixing” of the representations of the study items with learning.

The binding of items to the temporal contexts in which they were presented enables \mathbf{t}^{IN} to become a representation that can capture higher-order relationships among stimuli. To demonstrate this, we calculated stimulus similarities, e.g. $\mathbf{t}_A^{IN} \cdot \mathbf{t}_B^{IN}$ after the model was presented with a set of stimuli that included chains of transitive associations, e.g. $A \rightarrow B$, $B \rightarrow C$, $\dots E \rightarrow F$. This list structure, referred to as a double function list, because items serve as both cues and responses, was first introduced to the study of memory by Prieff (1938). Performance on double function lists is worse than on regular lists of paired associates. Slamecka (1976) argued that this is due to backward and remote associations among the items. TCM shares this prediction, which has been directly observed in final free recall of double function lists (Howard & Jing, 2003).

Despite the random order of presentation of the pairs, dou-

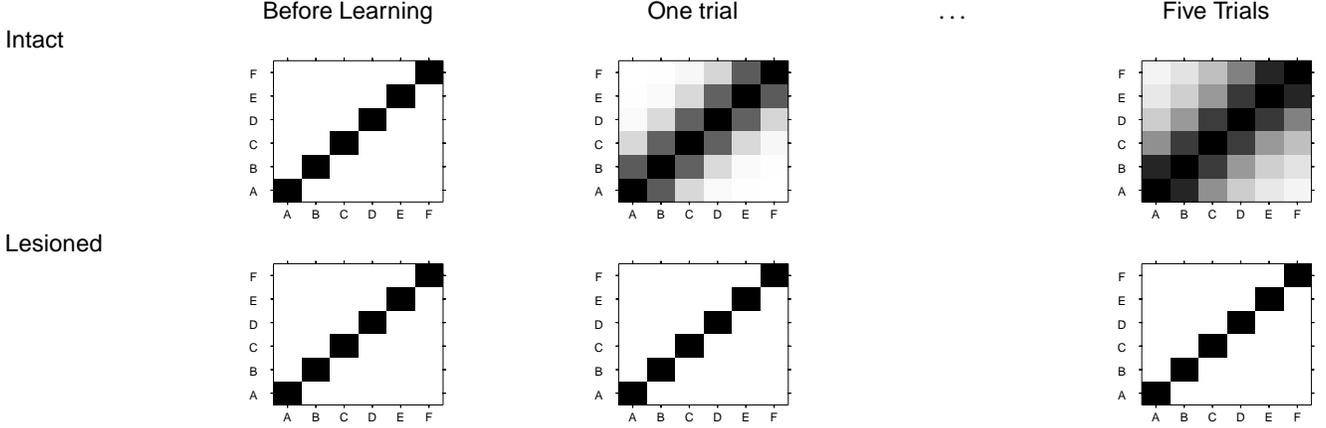


Figure 13. Rapid development of a “memory space” in TCM with an intact hippocampus. The network was presented with the double function pairs $A \rightarrow B$, $B \rightarrow C$ etc in 1,000 different random orders, intermixed with a parallel series of $X \rightarrow Y$, $Y \rightarrow Z$, etc, pairs. Each panel shows the similarity matrix among \mathbf{t}_A^{IN} , \mathbf{t}_B^{IN} , etc., at various stages of learning. For instance, the color of cell D, F indicates the value of $\mathbf{t}_D^{IN} \cdot \mathbf{t}_F^{IN}$. Black indicates a value of one, white indicates a value of zero. The top row of matrices is for the “intact” model. The bottom row is for the lesioned model. Parameter values are the same used in Figure 12. The intact and impaired model both start with an orthonormal representation of the \mathbf{t}^{IN} ’s. This can be seen by the value of one for all the cells along the diagonal and zeros for all off-diagonal cells on the left-most panels. With learning, however, the intact model develops a similarity structure that comes to reflect the “distance” within the double function list. This can be seen by the development of non-zero off-diagonal cells whose magnitude falls off with distance from the diagonal. This is what enables the intact model to “generalize” associations to pairs that were never presented together. In contrast, the lesioned model always evokes the same \mathbf{t}^{IN} in response to each stimulus. This prevents the lesioned model from generalizing, although the orthonormal \mathbf{t}^{IN} representation may be associated to any presented stimulus to support a forward association.

ble function lists induce a higher-order structure:

$$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F. \quad (19)$$

In this structure, B is closer to D than it is to E . If the \mathbf{t}^{IN} s have come to capture this higher order structure, then after learning we should observe that

$$\mathbf{t}_B^{IN} \cdot \mathbf{t}_D^{IN} > \mathbf{t}_B^{IN} \cdot \mathbf{t}_E^{IN}. \quad (20)$$

In general, if \mathbf{t}^{IN} is a representation that reflects higher-order relationships among the stimuli, then the similarity between the \mathbf{t}^{IN} s evoked by any two stimuli ought to be inversely proportional to their distance in the structure illustrated by Eq. 19.

Memory space simulation methods. We examined the effect of learning on the similarity relationships among 5 pairs structured according to Eq. 19. The pairs were presented in a random order, with presentation of another parallel series of pairs (*i.e.* an $X \rightarrow Y$ series) interspersed randomly. For each level of learning, 1000 replications with a different random presentation order were averaged. The value of β was the same as those used previously in the simulation shown in Figure 12. Both the lesioned model and the intact model were run for 1-5 trials. In both cases, we assumed that initially the \mathbf{t}^{IN} s were orthonormal prior to learning: $\mathbf{t}_{A_0}^{IN} \cdot \mathbf{t}_{B_0}^{IN} = \delta_{AB}$.

Results. The stimulus similarities for the intact and lesioned model at various stages of learning are illustrated in

Figure 13. On the left, we can see that before learning both the lesioned model and the intact model start with an orthonormal stimulus representation. This is just an expression of our assumptions about the initial conditions used in the simulation. With repeated presentations of the linked lists, the lesioned model does not change its stimulus representation. This is a consequence of Eq. 18; the similarity relationships among the \mathbf{t}^{IN} s do not change for the lesioned model because $\mathbf{t}_{A_i}^{IN} = \mathbf{t}_{A_{i+1}}^{IN}$. The intact model, however shows a more interesting pattern of results. First, we note that the stimulus representation of members of the same pair become similar to each other; although \mathbf{t}_C^{IN} and \mathbf{t}_D^{IN} are initially completely dissimilar, they quickly come to have some similarity. Comparing the rightmost panel with the middle panel, we see that this similarity increases with subsequent learning for the intact model.

Moreover, the intact model develops a stimulus representation that reflects the higher order structure of the linked list. Looking at the right of the figure, we see that after five learning trials the similarity of \mathbf{t}_B^{IN} to \mathbf{t}_D^{IN} is higher than it was at the start of learning. Stimuli B and D were never presented together, but were both presented with C . The model shows stimulus generalization among arbitrary stimuli as a function of the similarity of the temporal contexts in which they were presented. This stimulus generalization is the property that allows the development of transitive associations seen in the simulations of the Bunsey and Eichenbaum (1996) experiment (Figure 12). In addition to allowing associations between stimuli that were never presented together, this stimulus generalization also comes to reflect the higher-order structure of the list. For example \mathbf{t}_B^{IN} and \mathbf{t}_D^{IN} are more

similar to each other than are t_B^{IN} and t_E^{IN} . Similarly, t_B^{IN} and t_E^{IN} are more similar to each other than are t_B^{IN} and t_F^{IN} . The similarity between any two input patterns comes to reflect their “distance” in the linked-list structure.

Discussion

Transitive associations link items that were not actually paired together during study, but rather are associated by means of having been presented in the context of some other, common element. We showed that one component of retrieved context, weighted by α_N , is responsible for backward and transitive association, hallmarks of relational learning (Figure 12). We also showed that in TCM this ability is a consequence of the development of an intermediate stimulus representation that comes to reflect the temporal context in which items were presented.

TCM developed a two-component account of associations to describe the characteristic shape of CRP curves (Figure 2). It is striking that this two-component account also turns out to provide an account of the dissociations between transitive and pairwise associations that result from hippocampal damage (Bunsey & Eichenbaum, 1996). The two-component account also predicts that hippocampal function is important in proper development of intermediate representations necessary for relational learning.

TCM and “transitive inference”.

If you tell a school age child that Alexander is taller than Betsy, and Betsy is taller than Catherine, that child should be able to tell you, without being explicitly instructed so, that Alexander is also taller than Catherine. In this example of transitive inference, the child is able to infer from her experience with the world that the property of height obeys a transitive relationship; if $A > B$ and $B > C$, then $A > C$. The cognitive process that enables one to reach the conclusion that $A > C$ is referred to as a transitive inference.

In the animal cognition and neuropsychology literature there has been considerable attention paid to a related task, in which animals learn preference relations between arbitrary stimuli. This has become an issue in describing hippocampal function because of the finding that MTL damage selectively disrupts “transitive inference.” Dusek and Eichenbaum (1997) trained rats on a series of conditional discriminations. When presented with a pair of odors A and B , one of the odors, A , was always paired with reward and the other was not. To receive a food reward, the animal would choose A when presented with the pair A, B . Several such pairs, e.g. B, C and so on, up to D, E were presented with the stimulus with the label appearing earlier in the alphabet paired with reward. After learning all of these premise pairs, the animals were tested on novel stimulus pairings. The novel end-anchored pairing A, E should be relatively easy; A was always rewarded and E never was rewarded. However, the pairing of B, D cannot be solved simply on the basis of reward valence. Control animals preferred B when presented with the B, D pairing, as if they had learned relationships like $A > B$, $B > C$ and so on from the premise pairs and

performed a transitive inference when presented with B, D . Interestingly, Dusek and Eichenbaum (1997) found that animals with lesions intended to disrupt hippocampal function (either fornix lesions or entorhinal lesions intended to deafen the hippocampus) learned the pairwise discriminations as well as intact animals. Lesioned animals also selected A as often as control animals when presented with the end-anchored pairs. However, unlike the control animals, they selected B and D equally often when presented with B, D . Hippocampal lesions specifically disrupted performance on the novel stimulus pairings that were presumably solved by means of transitive inference.

Referring to performance on the B, D pair as an inference may be something of a misnomer; it is not necessary to assume that the animal has actually performed a logical inference to explain this behavior as the task can be performed on a purely associative basis. Recently Van Elzakker, O’Reilly, and Rudy (2003) did an experiment that they argued contradicted an inferential explanation of the transitive inference findings of Dusek and Eichenbaum (1997). Rather than presenting four pairs, as in the study of Dusek and Eichenbaum (1997), they presented five pairs, referring to the additional pair as E, F . This enabled them to compare transitive choices when the animal was presented with novel pair combinations of differing lags. For instance, only one item intervenes between B and D , whereas two items intervene between B and E . The logic of their experiment was that if the choice on novel pairs was made on the basis of a logical inference, then B, D should be easier than B, E , because fewer premises must be combined to make the judgment. In fact, Van Elzakker et al. (2003) found that performance was better on B, E than on B, D . This finding is consistent with an associative account. In the experiment of Van Elzakker et al. (2003), stimulus A was always rewarded, whereas F never was. If a stimulus similarity gradient is established (as in Figure 13), then stimuli closer in the chain to A would be more strongly associated to food than items further away in the chain.

The finding that hippocampal damage selectively disrupts performance on novel stimulus pairings that could be solved on the basis of a transitive inference has been extensively covered recently by models of hippocampal function (Frank, Rudy, & O’Reilly, 2003; O’Reilly & Rudy, 2001; Levy, 1996; Wu & Levy, 1998, 2001). For the most part, models of the role of the hippocampus in transitive inference hypothesize that the hippocampus supports overlapping stimulus representations that can be used to perform the task.¹⁸ This is a role that is wholly consistent with the role for the hippocampus proposed here. Frank et al. (2003) hypothesized that there were two stages in making a response when presented with a pair of stimuli in a choice situation. In a first stage, the animal selected which of the two odors to approach based on an associative gradient from reward to each of the stimuli. After selecting a stimulus to approach, the animal then ei-

¹⁸ Although it has not been directly applied to the transitive inference task, this property is also shared by the Gluck and Myers (1993) model of hippocampal associative learning (see Gluck & Myers, 1997, for a review).

ther selected that odor on the basis of a recall-like process or it switched to the other odor. Transitive inferences were a consequence of an associative gradient across the stimuli to the “dig” response. The Complementary Learning Systems model (O’Reilly & Rudy, 2001) postulates that transitive performance is a consequence of overlapping hippocampal stimulus representations. In that model, however, correct responding depends on network dynamics to affect pattern completion. As a consequence, transitive performance is sensitive to the detailed structure of the learning episode. Similarly, in the Wu and Levy (1998, 2001) model of transitive inference performance, the extent to which the hippocampal representation evoked by the *B, D* probe overlaps with the representation evoked by *C* corresponds with network performance on the transitive inference problem (Wu & Levy, 1998). This apparently supports a representation that captures the “distance” between the stimuli in the higher order structure, resulting in the network showing a symbolic distance effect (Wu & Levy, 2001).

An intermediate stimulus representation like that described here could be used to construct an associative gradient to perform the transitive inference task, in much the same way that the Frank et al. (2003) model did. However, the intermediate stimulus representation does not necessarily imply a purely associative account of the transitive inference task. Quite the contrary, if an intermediate stimulus representation is developed that places the stimuli in order along a relevant, albeit abstract, dimension, then this information could be used to inform a logical inference, in much the same way that an inference about a physical dimension, like location or height, can be performed. For instance, a different levels of association between stimuli and a food reward could be used to generate an abstract dimension like “foodliness.” It is clear from Figure 13 that the similarity of items nearby in the higher-order structure is higher than for items far apart in the higher-order structure. From this it is clear that this representation has extracted a dimension analogous to “position” from the higher-order list structure. This could, in principle at least, be used as the basis for a non-associative, logical decision.

The hippocampus and semantic learning.

In contrast to episodic memory, semantic memory refers to general knowledge about the world without reference for specific events. For instance our knowledge about bananas must have been learned as a result of some instruction or experience, but it is not necessary to remember any one of those learning events to remember that bananas are yellow, or that they are good to eat. The default hypothesis, until quite recently, has been that semantic memory depends on episodic memory. The idea is that we experience a number of specific episodes pertaining to the same subject (bananas in this case). Perhaps the brain manages to gradually build up a representation that extracts the commonalities of these experiences so that it no longer requires any of the individual episodes (e.g. Marr, 1971; McClelland, McNaughton, & O’Reilly, 1995).

The belief that learning of semantic memory depends

on episodic memory is consistent with findings showing that some MTL amnesics have not learned the meanings of words that entered the lexicon after the incident that caused their amnesia (Ostergaard, 1987; Gabrieli, Cohen, & Corkin, 1988). More recently, the dependence of semantic memory on episodic memory has been cast into doubt by the finding that patients with substantial hippocampal damage acquired at a very early age show no evidence for any episodic memory, but nonetheless have acquired enough semantic memory to perform at a normal level in school (Vargha-Khadem et al., 1997). Subsequent studies have purported to show some acquisition of post-morbid vocabulary in adult amnesics (Kitchener, Hodges, & McCarthy, 1998; Linden et al., 2001; Schmolck, Kensinger, Corkin, & Squire, 2002). These findings have led some to propose alternative relationships between episodic and semantic memory (Tulving & Markowitsch, 1998; Vargha-Khadem, Gadian, & Mishkin, 2001). Others have argued that, even if the data is to be taken at face value, the observed semantic knowledge of these patients is a consequence of some preserved episodic memory, or is perhaps the result of some reorganization available to the developing brain that does not reflect normal adult function (e.g. Squire & Zola, 1998). This position is supported by evidence that severe damage limited to the hippocampus results in measurable deficits in post-morbid vocabulary acquisition (Cipolotti et al., 2001; Nadel & Moscovitch, 2001; Spiers, Maguire, & Burgess, 2001; Verfaellie, Koseff, & Alexander, 2000). Others note that while MTL amnesics can acquire familiarity for new words, and even learn to recite their definitions, their semantic knowledge for these materials lacks the inter-related richness of normal subjects (Westmacott & Moscovitch, 2001).

Vocabulary acquisition can be seen as a special case of semantic learning. A dictionary describes the meaning of each word simply in terms of other words. Learning the meaning of a word can in some sense be described as a process of placing the word in the proper relationship to the other words in the lexicon. TCM describes episodic association and transitive associations on the basis of retrieved context. Recent models of vocabulary acquisition using realistic databases of naturally occurring text describe semantic relationships among words by extracting information about the words’ contextual relationships (Griffiths & Steyvers, 2002; Landauer & Dumais, 1997). In much the way that associations in TCM can be seen as a retrieved context model of episodic association, these models can be seen as retrieved context models of semantic association.

Latent Semantic Analysis (LSA Landauer & Dumais, 1997) is a well-studied computational model that has been shown to describe something of human vocabulary acquisition. It expresses a representation of the semantic structure of the language by extracting useful information from the temporal co-occurrence properties of the language, as measured by large bodies of naturally-occurring text (for instance, an encyclopedia). This is possible because of regularities in the use of language. Words that are similar to each other tend to occur in the same context. For instance, words that refer to similar objects, like “table” and “chair,” will tend to occur

together in discussions of, say, seating arrangements or furniture. It is easy to extract this type of information—it can be accessed using a simple co-occurrence matrix. This information is analogous to pairwise associations between words in a transitive association experiment. But LSA goes further. Words that refer to the same object are not necessarily likely to occur in the same context, but will tend to appear in *similar* contexts. This is often the case with synonyms. If “sofa” and “couch” mean very nearly the same thing, an author is likely to choose one or the other, but not both, for a given passage. LSA is able to extract the similarity that can be inferred in this way by means of dimensional reduction. This process is analogous to the transitive associations described here. The end result of these computations is that the representation of the words in the corpus comes to reflect with some fidelity the semantic structure of English. As evidence for this claim, LSA can achieve a passing score on the Test of English as a Foreign Language (TOEFL Landauer & Dumais, 1997).

To summarize, LSA provides a description of semantic relationships that relies on two processes: one process that associates items based on their temporal co-occurrence and a second process that discovers transitive associations between items based on the contexts in which they occur. These are analogous to the two components giving rise to associations in TCM. One process can support associations between items that actually co-occur, like *A, B* in the Bunsey and Eichenbaum (1996) experiment. The other can support transitive associations between items that never occurred together, but that occurred in similar contexts, like *A* and *C* in the Bunsey and Eichenbaum (1996) experiment. If TCM can provide a description of semantic learning, and if the mapping between hippocampal function and TCM is the way we have hypothesized here, then this suggests a way to reconcile the conflicting data regarding hippocampal involvement in new semantic learning. Perhaps the preserved semantic learning with hippocampal damage can be described largely by a series of pairwise relationships.

Physiological evidence for a stimulus representation that reflects temporal context.

In this ms we have argued that the hippocampus functions to reconstruct the state of activity in entorhinal cortex when an item is repeated (Figure 3). We have shown that this ability to make new item-to-context associations leads to an intermediate stimulus representation that reflects the temporal contexts in which an item is presented (Figure 13) and argued that this representation can support transitive associations (Figure 12). We have argued that this representation should result from hippocampal function and should be located in parahippocampal regions. There is strong physiological evidence that the MTL in fact does in fact support the development of an intermediate stimulus representation that comes to reflect temporal context with learning.

Miyashita (1988) used abstract visual patterns as stimuli in a delayed match to sample (DMS) experiment. In his experiment, monkeys were presented with many learning sessions. In each session, the order of sample stimuli remained constant. The sample stimuli evoked sustained firing in some

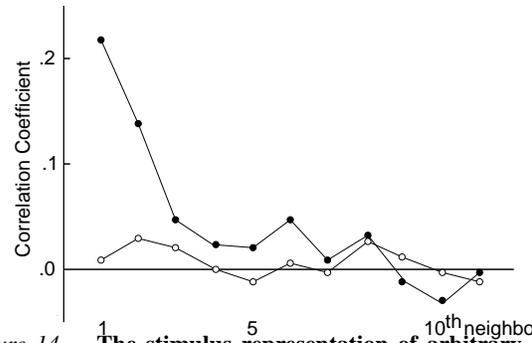


Figure 14. The stimulus representation of arbitrary stimuli comes to reflect temporal context in the inferior temporal cortex of monkeys. Single units were recorded from area TE of the inferior temporal cortex of macaque monkeys while they performed a delayed match to sample (DMS) task using abstract visual stimuli. The y-axis shows the correlation coefficient calculated for pairs of stimuli. High values of the correlation coefficient mean that neurons tended to fire selectively in response to both stimuli. The correlation coefficient, then, provides a measure of the overlap between the patterns of neural activity corresponding to different stimuli. The cues constituting the DMS task were presented in a fixed order. The filled symbols show the correlation coefficients for stimuli as a function of their distance within a “list” of DMS trials that was presented many times. The open symbols are for an unfamiliar list. The correlation coefficient falls off with distance for the familiar list such that remote pairs are no more correlated than by chance, or for pairs from the new list. For stimuli that were presented many times, the representation of stimuli that were presented in similar temporal contexts becomes more similar. Graph based on data from Miyashita (1988).

subset of the neurons in area TE, an inferotemporal area reciprocally connected to the perirhinal cortex, an extrahippocampal MTL region. According to the mapping between TCM and the MTL set out at the beginning of this ms, TE could be part of an item representation. Miyashita (1988) found that after many sessions of learning, but not after a single session, neurons that responded to the *i*th sample in the session also tended to respond to samples that were presented at nearby positions in the session (see Figure 14). Subsequent experimental work extended this finding to show “pair-selective” neurons that responded to both members of a pair of stimuli that were repeatedly presented together in an analogue of a paired-associate task (Sakai & Miyashita, 1991).

There is good evidence that this effect, first observed in TE, is in fact a consequence of MTL functioning. Pair-coding neurons are observed in perirhinal cortex (Erickson & Desimone, 1999; Messinger, Squire, Zola, & Albright, 2001), which, like the entorhinal cortex is an extrahippocampal MTL area. Further, the time course of activity following an individual stimulus presentation shows associative effects in perirhinal cortex about 100 ms earlier than in TE (Naya, Yoshida, & Miyashita, 2001). Naya, Yoshida, and Miyashita (2003) showed that pair-coding neurons are more prevalent in perirhinal cortex. These data suggest that the temporal stimulus generalization effect observed in TE is ac-

tually a consequence of MTL functioning.

The finding of pair-selective cells is perfectly consistent with the results for the intact model ($\alpha_N > 0$) shown in Figure 13 (top). Similarly the lesioned model would not show such an effect (Figure 13, bottom). Higuchi and Miyashita (1996) trained monkeys on a set of paired associates to a criterion. The pairs were each presented several hundred times. After training, the monkeys received ibotenic lesions to the entorhinal and perirhinal cortices, disconnecting TE from the backward signal from the MTL. After the lesion the monkeys were trained on a new set of stimuli. Pair coding was abolished in TE for both the old and new stimuli after the lesion, while general firing properties of the neurons were unchanged. Similar results have been found in another study (Miyashita, Kameyama, Hasegawa, & Fukushima, 1998). This result, in conjunction with the data reviewed above, argues strongly that the pair-coding phenomenon depends on input from MTL. The fact that pair-coding was abolished by lesion, even after several hundred trials suggests that the pair-coding phenomenon does not result from a change in the item representation *per se*, but rather from direct inputs from an activated MTL representation. That is, the observed pair coding in TE could result from input analogous to the mixture of item representations that results from $\mathbf{M}^{TF} \mathbf{t}$. This mapping predicts that pair-coding should be dependent on hippocampal lesions, and the pair-coding should be observed in parahippocampal MTL regions after relatively little training compared to extra-MTL regions.

General Discussion

TCM describes a distributed representation of temporal context that was argued to mediate performance in free recall, an episodic memory task. By demonstrating that the same equation used for contextual drift, Eq. 6, can be used to describe the entorhinal place code when provided with appropriate inputs, the model becomes one of a joint temporal-spatial context. Indeed, if episodic memory is defined to be memory that refers to a specific event in time and place, it is reasonable to hypothesize that a joint representation of temporal-spatial context contributes to this cognitive function.

A key component of TCM (Howard & Kahana, 2002a) is a form of short-term memory, \mathbf{t}_i , that varies according to a simple equation (Eq. 6). We implemented the key features of Eq. 6 (Figure 6) using a model intended to represent EC. The simulation was populated of integrator cells modeled after those in EC layer V (Egorov et al., 2002) and provided with input from the head direction system (Taube, 1998), which are known to synapse on EC layer V (Haeflten et al., 2000). Normalization of the integrator cell population was accomplished by means of a gain modulation where the gain varied inversely with the activity in the network (Chance et al., 2002).

This cellular simulation was essentially just Eq. 6 with input from velocity movements. This simple model described much of the place code observed in EC. In the open field, these features include a representation that correlated with

spatial position and was consistent across different environments (Figure 8). In the W-maze, we showed that this representation naturally accounts for history-dependent phenomena, including retrospective (Figure 10) and trajectory coding (Figure 8), observed in entorhinal place cells. This close correspondence between the predictions of Eq. 6 and the activity of entorhinal cells during spatial navigation is consistent with the hypothesis that \mathbf{t}_i resides in parahippocampal regions, including EC.

We explored the ability of TCM to organically explain neuropsychological dissociations associated with hippocampal damage. We hypothesized that a primary function of the hippocampus was to allow repetition of an item to reconstruct the state of \mathbf{t}_i in EC that was present when that item was initially presented (Figure 3). In TCM, a parameter, α_N , describes this ability. We showed that setting α_N to zero, corresponding to no reconstruction, prevents transitive and backward associations while pair-wise associations remain intact (Figure 12). These dissociations have been reported with hippocampal damage (Bunsey & Eichenbaum, 1996), and have been taken to be hallmarks of relational memory. We then illustrated that the ability to reconstruct states of \mathbf{t}_i in EC allows the development of an intermediate stimulus representation that captures the higher-order structure of the stimuli, consistent with the “memory space” idea advanced by Eichenbaum (2000, 2001) (Figure 13). We also argued that a memory space could be useful in describing performance in so-called transitive inference tasks in a way broadly consistent with existing models of the hippocampus and transitive inference performance. Neurophysiological results from primate studies have shown direct evidence for a stimulus representation that comes to reflect the temporal context in which items were presented (Erickson & Desimone, 1999; Messinger et al., 2001; Miyashita, 1988; Sakai & Miyashita, 1991). The development of this intermediate stimulus representation is also known to be a consequence of MTL function (Higuchi & Miyashita, 1996; Miyashita et al., 1998; Naya et al., 2001).

Our hypotheses regarding the entorhinal place code and relational memory were both supported by substantive physiological evidence. We argued that EC supports a leaky integrator functioning like short-term memory (Eq. 6). We argued for the plausibility of this hypothesis using detailed intracellular experiments (Egorov et al., 2002), neuroanatomy (Haeflten et al., 2000), and physiology (Chance et al., 2002). Using this implementation we demonstrated a close correspondence between simulated neurons and data from single entorhinal units during spatial navigation (Quirk et al., 1992; Frank et al., 2000). In treating relational memory, we argued that the MTL, in particular the hippocampus proper, causes the development of an intermediate stimulus representation that reflects temporal context (Eq. 9). There is considerable evidence for just this phenomenon in the primate (Higuchi & Miyashita, 1996; Messinger et al., 2001; Miyashita, 1988; Naya et al., 2001).

Relationship to other work

The present work draws together thought on MTL function in apparently disparate domains. In doing so, it builds on extensive work in each of these domains. The relationship of the model presented here to other models of relational memory, in particular models developed to describe the effect of hippocampal lesions on the so-called transitive inference task, was discussed above. We discuss the relationship of TCM to other models of episodic memory and place cells here.

Retrieved context models of memory. In the domain of episodic recall TCM can be seen as a descendant of the stimulus-sampling model (Estes, 1950, 1955), which was subsequently cast as a model of temporal effects and forgetting in paired associate learning (Mensink & Raaijmakers, 1988, 1989). The important difference between TCM and these prior models is the nature of contextual drift. Whereas those other works assumed that contextual drift was a random process, TCM assumes that contextual drift is a consequence of elements retrieved by the nominal stimuli presented during learning.

TCM's focus on contextual processing in describing episodic recall has parallels in other aspects of memory research. As mentioned earlier, retrieved context models have also made considerable headway in describing the structure of semantic memory (Griffiths & Steyvers, 2002; Landauer & Dumais, 1997). Retrieved context has also been proposed as the basis for episodic recognition decisions (Dennis & Humphreys, 2001). When presented with a probe item, Dennis and Humphreys (2001) proposed that it is used to retrieve a superposition of context vectors corresponding to the state of contexts in which the item was previously presented. This retrieved context is then compared to a representation of list context. This approach, which successfully explains the bulk of the extant recognition memory data, represents a departure from many previous models of recognition memory (e.g. Murdock, 1982; Shiffrin & Steyvers, 1997). Recent years have seen the development of a neuroanatomical model of two-process recognition memory in which the hippocampus proper is responsible for episodic recollection, whereas cortical regions within the MTL are responsible for a scalar familiarity signal (Davachi, Mitchell, & Wagner, 2003; Norman & O'Reilly, 2003; Rugg & Yonelinas, 2003; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998; Yonelinas et al., 2002). This view of the hippocampus in recognition memory is quite consistent with the view expressed here—that the hippocampus is responsible for reconstructing patterns of context present in entorhinal cortex. Reconstruction of these patterns is a plausible candidate for recollection (Polyn, Norman, & Cohen, 2002). If this is the case, and context changes gradually in entorhinal cortex, as hypothesized here, then one would expect to see associative effects as a consequence of successful recollection during a recognition test.

It is striking that retrieved context has been proposed in the cognitive literature, more or less independently, as a mechanism for performance in three diverse classes of

tasks: episodic recall (Howard & Kahana, 2002a), recognition memory (Dennis & Humphreys, 2001), and semantic learning (Griffiths & Steyvers, 2002; Landauer & Dumais, 1997). The similarities of these three classes of models represent a unique opportunity for theoretical convergence. The present work suggests that a unification would have relevance for understanding the function of the medial temporal lobe.

Path integration models of the place code. TCM describes the entorhinal place code as a joint expression of temporal-spatial context. That this might provide an explanation of the MTL's importance in both episodic memory and spatial navigation has been proposed by other authors (e.g. Levy, 1989). Our emphasis on inputs corresponding to information about physical motion in space places the present treatment in the tradition of "path integration" models of the place code (McNaughton, Barnes, Gerrard, et al., 1996; Samsonovich & McNaughton, 1997; Redish & Touretzky, 1997). Much like the present treatment, these models postulate that the place code results from updating a representation of position by operating on input from the head direction system. In particular, the treatment of Redish and Touretzky (1997, see also Redish, 1999) postulated that path integration takes place in the EC. In the present treatment, we have argued that a leaky, "pseudo" integrator resides in EC.

The most obvious difference between prior path integration place cell models and the present treatment is the level of neural sophistication those models brought to bear on the problem. The relative simplicity of the present treatment is a consequence of several factors. One is the relatively limited scope of the current treatment, restricting our attention to the properties of the entorhinal place code and neglecting such important factors as the means of operation of the head direction system and the hippocampal place code. Another is the recent discovery of "integrator cells" in the EC that integrate their inputs in the absence of synaptic connections (Egorov et al., 2002). This remarkable finding simplifies considerably the neural hardware required to implement an integrator. Of course the intracellular machinery that supports the properties of these cells is of tremendous interest (Fransen, Egorov, Hasselmo, & Alonso, 2003).

On a computational level, the current treatment differs from prior work on path integration models of place cell formation by postulating that path integration is "leaky"— ρ_i is less than one (see Figure 4), meaning that integration is not perfect. In contrast, prior models hypothesized that integration was not leaky, but perfect. The "leakiness," or forgetting, in the current treatment was originally introduced to TCM as a way of modeling recency and contiguity effects in episodic memory performance. However, the assumption of forgetting in dead reckoning simplifies considerably the computational requirements of the system.

In dead reckoning, the current position is derived from the prior position combined with the current movement. If there is any error in the estimation of the current movement, this will lead to an error in the subsequent estimate of position. This error will accumulate in a perfect integrator—with

more movements, the amount of uncertainty in position will grow without bound as more and more movements are integrated. Previous path integration models have devoted considerable effort to error-correcting mechanisms to counteract this tendency (e.g. Redish & Touretzky, 1997). However, when $\rho < 1$, \mathbf{t} is not subject to cumulative error. The amount of “error,” while decidedly non-zero, is stable with time. It is an open question whether the systematic discrepancies between the model’s representation and a perfect representation of place are reasonable given the observed data. At least in the case of retrospective coding, an “error-free” representation of place is unable to describe the observed data (Figure 4).

Receptive field place cell models. The other large class of models of place cell formation can be referred to as “receptive field models” (Brunel & Trullier, 1998; Burgess & O’Keefe, 1996; Hartley et al., 2000; Kali & Dayan, 2000; Sharp, 1991; Sharp et al., 1996). These models make two broad assumptions about the basis of the place code. One is that the hippocampus receives inputs from the EC that have a spatial-geometric character. The second assumption is that the hippocampus supports a conjunctive coding of these inputs, resulting in a sharper, more focused spatial representation.

In one popular theory (Burgess & O’Keefe, 1996; Hartley et al., 2000), entorhinal cells are assumed to code for the distance to a particular landmark, such as a wall, within the environment. Hippocampal cells receive input from a number of entorhinal cells, resulting in a relatively focused place field. For instance, one entorhinal cell might respond preferentially whenever the animal is 10 cm from the Eastern wall of an enclosure, resulting in a place field shaped like a “strip” running North-South 10 cm from the Eastern wall. Another entorhinal cell might respond preferentially whenever the animal is 8 cm from the Northern wall of the enclosure. In other treatments (Brunel & Trullier, 1998; Kali & Dayan, 2000), the entorhinal inputs are assumed to retain directionality, as well as sensitivity to the distance of landmarks. The inclusion of directionality is consistent with an encoding of “local view” information.

Receptive field models rely on a conjunctive code of entorhinal representations. For instance, Brunel and Trullier (1998) and Kali and Dayan (2000) showed that by means of conjunctive coding, broad, directionally-sensitive place fields in EC can give rise to focused, non-directional fields in the hippocampus. Conjunctive coding from multiple, non-specific entorhinal cells can give rise to a more specific hippocampal representation. To use the example above, a hippocampal cell might receive input from these two entorhinal cells and have a place field that is in the North-East quadrant of the enclosure, 10 cm from the Eastern wall and 8 cm from the Northern wall. In this example, the hippocampus provides a more focused spatial representation than EC by means of a conjunctive representation. This is quite consistent with recent findings of Anderson and Jeffery (2003) that some hippocampal place fields were modulated by the presence of non-spatial environmental stimuli in a conjunc-

tive fashion. The combination of spatial-geometric input and conjunctive encoding leads to some very specific predictions. For instance, if the inputs to hippocampal place cells are coding for distances to the boundary in an environment, then hippocampal place cells should deform in a very specific way as the environment is stretched. These predictions have been directly observed in quite dramatic fashion (O’Keefe & Burgess, 1996).

The present treatment, in focusing exclusively on the entorhinal place code, is completely mute on the issue of whether or not the hippocampus generates a conjunctive code of its inputs. The other main assumption of receptive field models of the hippocampus is that cells in EC provide a spatial-geometric code as input to the hippocampus. At first glance, it might seem that this is in direct contrast to the weighted sum over recent movements explored here. This contrast could be more apparent than real. It is possible that the weighted sum over recent movements postulated here approximates the spatial assumptions of the receptive field models sufficiently closely to result in comparable predictions if similar assumptions about the hippocampus are made. For instance, a weighted sum over recent movements should weight recent movements strongly, resulting in a directional selectivity, as assumed by some recent receptive field models (Brunel & Trullier, 1998; Kali & Dayan, 2000). Similarly, a weighted sum over recent movements might be able to approximate the specification that entorhinal cells code for distance to a wall of an enclosure.

This paper has tried to explain the entorhinal place code using solely self-motion information as input to Eq. 6. This should not be taken as a statement that \mathbf{t}_i^N should contain only self-motion information in spatial applications. Because \mathbf{t}_i reflects a temporal-spatial integrator, a joint representation of temporal-spatial context, we would expect that exposure to salient non-spatial stimuli during exploration would contribute to \mathbf{t}_i . There is therefore no fundamental difficulty in modeling “receptive fields” defined by a relationship to a landmark. Inputs \mathbf{t}_i^N corresponding to landmark stimuli should be able to be “dropped in” to \mathbf{t}_i in the same way as retrieved temporal context from words are. In this way, arguments advanced in the current ms is not necessarily inconsistent with hippocampal place cells that appear to be bound to landmarks or conjunctions of landmarks (e.g. Gothard, Skaggs, Moore, & McNaughton, 1996).

Contextual retrieval in spatial navigation

In treating relational memory we emphasized the importance of new item-to-context learning in establishing an intermediate stimulus representation. We argued that a non-zero value of α_N meant that the hippocampus was functioning normally and allowed an item to reconstruct the state of context in EC that was present when the item was initially presented. In that section, we argued that setting α_N to zero provided a good model of hippocampal lesions in transitive association (Figure 12). In contrast, when we were treating the entorhinal place code, we set α_N to zero throughout. How is it that the activity of cells in EC during spatial navigation,

believed to be the most characteristically hippocampal function (O’Keefe & Nadel, 1978) can be described under circumstances that corresponded to a hippocampal lesion in our treatment of relational memory?

Although we initially set γ to zero in treating the entorhinal place code out of convenience, it is clear that including item-to-context learning in spatial navigation would require some elaboration of the model. What would happen if γ was set to a non-zero value in the spatial navigation applications? The first decision that needs to be made is what constitutes an “item” to define item-to-context learning (as in Eq. 9). If we simply define each head direction as an “item” this leads to a very interesting, but suboptimal situation. A thought experiment should suffice to illustrate.

A thought experiment in which head directions retrieve context. Consider the situation in which we have a series of four movements that repeat in sequence as the animal runs around a square maze. We have four orthonormal “items,” \mathbf{v}^0 , $\mathbf{v}^{\pi/2}$, \mathbf{v}^{π} and $\mathbf{v}^{3\pi/2}$ corresponding to movements in the four cardinal directions. These are repeatedly presented in order. We can then describe the behavior of \mathbf{t} in terms of these four basis vectors. If there is no item-to-context learning, there will be something like a place code— \mathbf{t}_i will be different on the four sides of the square. Let us denote the activity on the i th segment of the N th traversal of the maze as \mathbf{t}_{iN} . Let us “turn on” new item-to-context learning with $\gamma = 1$ and consider the asymptotic behavior as N gets large. After a sufficiently long time, \mathbf{t}_{iN}^{IN} should no longer change with N , so that $\mathbf{t}_{i(N+1)}^{IN} = \mathbf{t}_{iN}^{IN}$. For this to be the case, substituting into Eq. 9 tells us that

$$\mathbf{t}_{i(N+1)}^{IN} \propto \mathbf{t}_{iN} + \mathbf{t}_{i(N+1)}^{IN}. \quad (21)$$

For this to be true, \mathbf{t}_{iN} must lie in the same direction as $\mathbf{t}_{i(N+1)}^{IN}$. But \mathbf{t}_{iN} includes a term $\mathbf{t}_{(i-1)N}^{IN}$. This means that $\mathbf{t}_{(i-1)N}^{IN}$, the input vector from the previous direction has to lie in the same direction as \mathbf{t}_{iN}^{IN} . The steady state of this system is for the \mathbf{t} vectors corresponding to all four stages of the path and all four input vectors \mathbf{t}^{IN} to point in the same direction.¹⁹ This means that the space spanned by \mathbf{t}_i after learning has collapsed into a single point. There is no longer any place-specific firing under these circumstances. After sufficient experience, the place field for every simulated cell would cover the entire maze. We conclude from this thought experiment that self-motion information can not retrieve context in the same way that non-spatial items do in Eq. 6.

Elaborations of the contextual retrieval rule. We just saw that it is insufficient to treat velocity vectors as “items” in engaging the contextual retrieval rule (Eq. 9). How might TCM be elaborated to account for hippocampal function during spatial navigation? One possibility is that for an “item” to engage the new item-to-context learning rule (Eq. 9) it must have certain properties that are not met by input from the head direction system, but that are met by words in a randomly assembled list and other non-spatial stimuli. There

are several properties that distinguish these classes of stimuli. For instance, it is possible that the anatomy and/or physiology of the MTL is such that head direction inputs cannot engage new item-to-context learning whereas non-spatial stimuli can. It is possible that to engage new item-to-context learning it is necessary to have a rapid change in the item representation. This is more plausible for non-spatial stimuli than for head direction—physics and the inherent overlap in the tuning curves of head direction cells means that you can’t “turn on” one particular head direction all at once. If this is the case, then a high-pass filter at the input end of the hippocampus (perhaps the dentate gyrus) could accomplish this task. Another possibility is that new item-to-context learning could only be engaged by items with sufficiently low frequency. The head direction system is active more or less all the time, whereas the types of non-spatial stimuli typically used in memory experiments are infrequently encountered.

If the hippocampus does not associate head directions to positional representations during spatial navigation, then what does it do? Redish (1999) has suggested that the hippocampus plays a role in spatial navigation by retrieving context to help orient the animal when it enters a new environment. Another possibility is that the hippocampus does perform new item-to-context learning during spatial navigation, but that this process is restricted to salient environmental stimuli. This could be important in associating non-spatial stimuli to spatial locations (Burgess, Maguire, & O’Keefe, 2002; Gilbert & Kesner, 2002). Recently, Burgess and colleagues (Burgess, 2002; Burgess et al., 2002) have hypothesized that when presented with items encountered in a virtual environment, the hippocampus plays a key role in retrieving the spatial context the item was learned in. This hypothesis is supported by both neuropsychology (Spiers et al., 2001) and functional imaging (Burgess, Maguire, Spiers, & O’Keefe, 2001). Contextual retrieval of salient stimuli could also be important in supporting behavioral path integration. To return to the home cage, the rat must presumably recover the spatial representation of the home cage’s location. If presenting the item “home cage” as a probe, then retrieved context would be the location of that object. This interpretation is consistent with lesion studies of behavioral path integration, which show that animals with hippocampal damage cannot return directly to their starting position (Maaswinkel, Jarrard, & Whishaw, 1999; Whishaw, McKenna, & Maaswinkel, 1997; Whishaw & Maaswinkel, 1998 but see Alyan & McNaughton, 1999). The finding that hippocampal place fields in blind rats only become aligned after the first experience with a distinctive landmark object (Save, Cressant, Thinus-Blanc, & Poucet, 1998) is also quite consistent with the idea that contextual learning and retrieval only engages sufficiently distinctive stimuli.

At the very least the mesoscopic computational approach taken here has enabled us to frame the question of hippocampal function in a way that, if satisfied, will be simultaneously consistent with considerations from multiple domains. If one

¹⁹ If we have four orthonormal input vectors initially, the steady state is the vector with all four components set to $1/2$.

can model hippocampal place cell behavior in a way that enables the hippocampus to reconstruct the state of EC when presented with a repeated non-spatial stimulus, then the resulting physiological model would be able to explain data from a broad variety of cognitive memory tasks.

Conclusions

The Temporal Context Model (TCM), developed to describe essential properties of episodic recall, captures key properties of both the entorhinal place code and relational memory. It does so by proposing the existence of a leaky integrator, and changes in stimulus representations, respectively. Both mechanisms are consistent with observed cellular-level data. TCM can address data across a wide variety of tasks, providing a first step toward a unified computational account of MTL function.

References

- Alyan, S., & McNaughton, B. L. (1999). Hippocampectomized rats are capable of homing by path integration. *Behavioral Neuroscience*, *113*(1), 19-31.
- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97-123.
- Anderson, M. I., & Jeffery, K. J. (2003). Heterogeneous modulation of place cell firing by changes in context. *Journal of Neuroscience*, *23*(26), 8827-35.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, p. 89-105). New York: Academic Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, p. 47-90). New York: Academic Press.
- Baddeley, A. D., & Hitch, G. J. (1977). Recency reexamined. In S. Dornic (Ed.), *Attention and performance VI* (p. 647-667). Hillsdale, NJ: Erlbaum.
- Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 176-189.
- Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W., & Lin, L. H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Progress in Brain Research*, *83*, 287-300.
- Becker, S., & Lim, J. (in press). A computational model of prefrontal control in free recall: strategic memory use in the California Verbal Learning Task. *Journal of Cognitive Neuroscience*, *15*(6).
- Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*, 173-189.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624-52.
- Bousfeld, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229-240.
- Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., Janowsky, J. S., Taylor, S. F., Yesavage, J. A., Mumenthaler, M. S., Jagust, W. J., & Reed, B. R. (2001). Context processing in older adults: evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, *130*(4), 746-63.
- Breese, C. R., Hampson, R. E., & Deadwyler, S. A. (1989). Hippocampal place cells: stereotypy and plasticity. *Journal of Neuroscience*, *9*(4), 1097-1111.
- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127-181.
- Brunel, N., & Trullier, O. (1998). Plasticity of directional place fields in a model of rodent CA3. *Hippocampus*, *8*(6), 651-665.
- Buffalo, E. A., Reber, P. J., & Squire, L. R. (1998). The human perirhinal cortex and recognition memory. *Hippocampus*, *8*(4), 330-9.
- Bunsey, M., & Eichenbaum, H. B. (1996). Conservation of hippocampal memory function in rats and humans. *Nature (London)*, *379*(6562), 255-257.
- Burgess, N. (2002). The hippocampus, space, and viewpoints in episodic memory. *Quarterly Journal of Experimental Psychology*, *55*(4), 1057-80.
- Burgess, N., & Hitch, G. J. (1992). Toward a network model of articulatory loop. *Journal of Memory and Language*, *31*, 429-460.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551-581.
- Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, *35*(4), 625-41.
- Burgess, N., Maguire, E. A., Spiers, H. J., & O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *NeuroImage*, *14*(2), 439-53.
- Burgess, N., & O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus*, *6*(6), 749-62.
- Burgess, N., Recce, M., & O'Keefe, J. (1994). A model of hippocampal function. *Neural Networks*, *7*(6-7), 1065-1081.
- Burwell, R. D. (2000). The parahippocampal region: corticocortical connectivity. *Annals of the New York Academy of Sciences*, *911*, 25-42.
- Carlesimo, G. A., Marfia, G. A., Loasses, A., & Caltagirone, C. (1996). Recency effect in anterograde amnesia: Evidence for distinct memory stores underlying enhanced retrieval of terminal items in immediate and delayed recall paradigms. *Neuropsychologia*, *34*, 177-184.
- Chance, F. S., & Abbott, L. F. (2000). Divisive inhibition in recurrent networks. *Network*, *11*(2), 119-29.
- Chance, F. S., Abbott, L. F., & Reyes, A. D. (2002). Gain modulation from background synaptic input. *Neuron*, *35*(4), 773-82.
- Cipolotti, L., Shallice, T., Chan, D., Fox, N., Schill, R., Harrison, G., Stevens, J., & Rudge, P. (2001). Long-term retrograde amnesia...the crucial role of the hippocampus. *Neuropsychologia*, *39*(2), 151-72.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature (London)*, *386*(6625), 604-608.

- Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: The MIT Press.
- Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews, Neuroscience*, 3(2), 153-60.
- Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Science, USA*, 100(4), 2157-62.
- Dehaene, S., & Changeux, J. P. (1997). A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Science, USA*, 94(24), 13293-8.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108(2), 452-78.
- Doboli, S., Minai, A. A., & Best, P. J. (2000). Latent attractors: a model for context-dependent place representations in the hippocampus. *Neural Computation*, 12(5), 1009-1043.
- Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Science, USA*, 94(13), 7109-7114.
- Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., & Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature (London)*, 420(6912), 173-8.
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews, Neuroscience*, 1(1), 41-50.
- Eichenbaum, H. (2001). The hippocampus and declarative memory: cognitive mechanisms and neural codes. *Behavioural Brain Research*, 127(1-2), 199-207.
- Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. London: Oxford University Press.
- Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, 17, 449-518.
- Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature (London)*, 425(6954), 184-8.
- Erickson, C. A., & Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *Journal of Neuroscience*, 19(23), 10404-10416.
- Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, 57, 94-107.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62, 145-154.
- Ferbinteanu, J., & Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron*, 40(6), 1227-39.
- Fernandez, G., Effer, A., Grunwald, T., et al.. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science*, 285, 1582-1585.
- Frank, L. M., Brown, E. N., & Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27(1), 169-178.
- Frank, M. J., Rudy, J. W., & O'Reilly, R. C. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. ii. a computational analysis. *Hippocampus*, 13(3), 341-54.
- Fransén, E., Alonso, A. A., & Hasselmo, M. E. (2002). Simulations of the role of the muscarinic-activated calcium-sensitive nonspecific cation current INCM in entorhinal neuronal activity during delayed matching tasks. *Journal of Neuroscience*, 22(3), 1081-1097.
- Fransen, E., Egorov, A., Hasselmo, M., & Alonso, A. (2003). Model of graded persistent activity in entorhinal cortex neurons. *Society for Neuroscience Abstracts*, 557.6.
- Gabrieli, J. D., Cohen, N. J., & Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition*, 7(2), 157-77.
- Gilbert, P. E., & Kesner, R. P. (2002). Role of the rodent hippocampus in paired-associate learning involving associations between a stimulus and a spatial location. *Behavioral Neuroscience*, 116(1), 63-71.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 5, 351-360.
- Glenberg, A. M., Bradley, M. M., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 413-418.
- Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., & Gretz, A. L. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 355-369.
- Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus*, 3(4), 491-516.
- Gluck, M. A., & Myers, C. E. (1997). Psychobiological models of hippocampal function in learning and memory. *Annual Review of Psychology*, 48, 481-514.
- Goldman-Rakic, P. S. (1996). Regional and cellular fractionation of working memory. *Proceedings of the National Academy of Science, USA*, 93(24), 13473-13480.
- Golob, E. J., & Taube, J. S. (1997). Head direction cells and episodic spatial information in rats without a hippocampus. *Proceedings of the National Academy of Science, USA*, 94(14), 7645-50.
- Gothard, K. M., Hoffman, K. L., Battaglia, F. P., & McNaughton, B. L. (2001). Dentate gyrus and cal ensemble activity during spatial reference frame shifts in the presence and absence of visual input. *Journal of Neuroscience*, 21(18), 7284-92.
- Gothard, K. M., Skaggs, W. E., & McNaughton, B. L. (1996). Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *Journal of Neuroscience*, 16(24), 8027-40.
- Gothard, K. M., Skaggs, W. E., Moore, K. M., & McNaughton, B. L. (1996). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *Journal of Neuroscience*, 16(2), 823-35.
- Grady, C. L., McIntosh, A. R., Horwitz, B., Maisog, J. M., Ungerleider, L. G., Mentis, M. J., Pietrini, P., Schapiro, M. B., & Haxby, J. V. (1995). Age-related reductions in human recognition memory due to impaired encoding. *Science*, 269(5221), 218-21.
- Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 164-78.
- Griffiths, T., & Steyvers, M. (2002). A probabilistic approach to semantic representation. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*.

- Haeften, T. van, Wouterlood, F. G., & Witter, M. P. (2000). Pre-subicular input to the dendrites of layer-v entorhinal neurons in the rat. *Annals of the New York Academy of Science*, *911*, 471-3.
- Hartley, T., Burgess, N., Lever, C., Cacucci, F., & O'Keefe, J. (2000). Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus*, *10*(4), 369-79.
- Hasselmo, M. E., Bodelón, C., & Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation*, *14*, 793-817.
- Hasselmo, M. E., Hay, J., Ilyn, M., & Gorchetchnikov, A. (2002). Neuromodulation, theta rhythm and rat spatial navigation. *Neural Networks*, *15*(4-6), 689-707.
- Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*, 73-137.
- Hermann, B. P., Seidenberg, M., Wyler, A., et al. (1996). The effects of human hippocampal resection on the serial position curve. *Cortex*, *32*, 323-334.
- Hetherington, P. A., & Shapiro, M. L. (1993). A simple network model simulates hippocampal place fields: II. computing goal-directed trajectories and memory fields. *Behavioral Neuroscience*, *107*(3), 434-43.
- Higuchi, S., & Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proceedings of the National Academy of Science, USA*, *93*(2), 739-743.
- Hölscher, C., Anwyl, R., & Rowan, M. J. (1997). Stimulation on the positive phase of hippocampal theta rhythm induces long-term potentiation that can be depotentiated by stimulation on the negative phase in area CA1 *in vivo*. *The Journal of Neuroscience*, *17*, 6470-6477.
- Howard, M. W. (2004). Scaling behavior in the temporal context model. *Journal of Mathematical Psychology*, *48*, 230-238.
- Howard, M. W., & Jing, B. (2003). Mediated associations in paired associate learning. *Poster presented at 44th annual meeting of the Psychonomic Society*.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923-941.
- Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.
- Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*(1), 85-98.
- Howard, M. W., Wingfield, A., & Kahana, M. J. (In revision). Dissociations between recency and association in the Temporal Context Model: A description of the mnemonic deficit observed in cognitive aging.
- Huerta, P. T., & Lisman, J. E. (1993). Heightened synaptic plasticity of hippocampal CA1 neurons during a cholinergically induced rhythmic state. *Nature*, *364*(6439), 723-725.
- Jensen, O., & Lisman, J. E. (2000). Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding. *Journal of Neurophysiology*, *83*, 2602-2609.
- Jones-Gotman, M. (1986). Right hippocampal excision impairs learning and recall of a list of abstract designs. *Neuropsychologia*, *24*(5), 659-70.
- Jung, M. W., Wiener, S. I., & McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *Journal of Neuroscience*, *14*(12), 7347-56.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*, 103-109.
- Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition*, *30*(6), 841-9.
- Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag-recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 530-540.
- Kali, S., & Dayan, P. (2000). The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model. *Journal of Neuroscience*, *20*(19), 7463-7477.
- Kesner, R. P., Crutcher, K., & Beers, D. R. (1988). Serial position curves for item (spatial location) information: role of the dorsal hippocampal formation and medial septum. *Brain Research*, *454*(1-2), 219-26.
- Kesner, R. P., & Novak, J. M. (1982). Serial position curve in rats: role of the dorsal hippocampus. *Science*, *218*(4568), 173-5.
- Kitchener, E. G., Hodges, J. R., & McCarthy, R. (1998). Acquisition of post-morbid vocabulary and semantic facts in the absence of episodic memory. *Brain*, *121*(7), 1313-27.
- Klein, K. A., Addis, K. M., & Kahana, M. J. (In press). A comparative analysis of serial and free recall. *Memory & Cognition*.
- Klink, R., & Alonso, A. (1997). Muscarinic modulation of the oscillatory and repetitive firing properties of entorhinal cortex layer II neurons. *Journal of Neurophysiology*, *77*(4), 1813-1828.
- Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, *34*(5/6), 419-426.
- Landauer, T. K., & Dumais, S. T. (1997). Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.
- Lenck-Santini, P. P., Save, E., & Poucet, B. (2001). Place-cell firing does not depend on the direction of turn in a y-maze alternation task. *The European Journal of Neuroscience*, *13*(5), 1055-8.
- Lever, C., Wills, T., Cacucci, F., Burgess, N., & O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature (London)*, *416*(6876), 90-4.
- Levy, W. B. (1989). A computational approach to hippocampal function. In R. D. Hawkins & G. H. Bower (Eds.), *Computational models of learning in simple neural systems* (p. 243-305). New York: Academic Press.
- 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.
- Linden, M. Van der, Cornil, V., Meulemans, T., Ivaniou, A., Salmon, E., & Coyette, F. (2001). Acquisition of a novel vocabulary in an amnesic patient. *Neurocase*, *7*(4), 283-93.
- Maaswinkel, H., Jarrard, L. E., & Wishaw, I. Q. (1999). Hippocampectomized rats are impaired in homing by path integration. *Hippocampus*, *9*(5), 553-561.
- Markus, E. J., Qin, Y. L., Leonard, B., et al. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *Journal of Neuroscience*, *15*(11), 7079-7094.
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society B*, *262*(841), 23-81.

- 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*(3), 419-57.
- McNaughton, B. L., Barnes, C. A., Gerrard, J. L., et al. (1996). Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *Journal of Experimental Biology*, *199*(1), 173-85.
- McNaughton, B. L., Barnes, C. A., & O'Keefe, J. (1983). The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Experimental Brain Research*, *52*(1), 41-9.
- Mensink, G.-J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-55.
- Mensink, G.-J. M., & Raaijmakers, J. G. W. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, *33*, 172-186.
- Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Science, USA*, *98*(21), 12239-12244.
- Meunier, M., Bachevalier, J., Mishkin, M., & Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *Journal of Neuroscience*, *13*(12), 5418-32.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.
- Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature (London)*, *335*(6193), 817-820.
- Miyashita, Y., Kameyama, M., Hasegawa, I. I., & Fukushima, T. (1998). Consolidation of visual associative long-term memory in the temporal cortex of primates. *Neurobiology of Learning and Memory*, *70*(1/2), 197-211.
- Muller, R. U., Bostock, E., Taube, J. S., & Kubie, J. L. (1994). On the directional firing properties of hippocampal place cells. *Journal of Neuroscience*, *14*(12), 7235-51.
- Muller, R. U., & Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, *7*(7), 1951-68.
- Muller, R. U., & Kubie, J. L. (1989). The firing of hippocampal place cells predicts the future position of freely moving rats. *Journal of Neuroscience*, *9*(12), 4101-10.
- Mumby, D. G., & Pinel, J. P. (1994). Rhinal cortex lesions and object recognition in rats. *Behavioral Neuroscience*, *108*(1), 11-8.
- Murdock, B., Smith, D., & Bai, J. (2001). Judgments of frequency and recency in a distributed memory model. *Journal of Mathematical Psychology*, *45*(4), 564-602.
- Murdock, B. B. (1962). The serial position curve in free recall. *Journal of Experimental Psychology*, *64*, 482-488.
- Murdock, B. B. (1963a). Interpolated recall in short-term memory. *Journal of Experimental Psychology*, *66*, 525-532.
- Murdock, B. B. (1963b). Short-term memory and paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, *2*, 320-328.
- Murdock, B. B. (1963c). Short-term retention of single paired associates. *Journal of Experimental Psychology*, *65*, 433-443.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609-626.
- 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*(16), 6568-82.
- Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217-27.
- Nadel, L., & Moscovitch, M. (2001). The hippocampal complex and long-term memory revisited. *Trends in Cognitive Science*, *5*(6), 228-230.
- Nairne, J. S., Neath, I., Serra, M., & Byun, E. (1997). Positional distinctiveness and the ratio rule in free recall. *Journal of Memory and Language*, *37*, 155-166.
- Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science*, *291*(5504), 661-664.
- Naya, Y., Yoshida, M., & Miyashita, Y. (2003). Forward processing of long-term associative memory in monkey inferotemporal cortex. *Journal of Neuroscience*, *23*(7), 2861-71.
- Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition*, *21*, 689-698.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review*, *110*(4), 611-46.
- O'Keefe, J. (1991). An allocentric spatial model for the hippocampal cognitive map. *Hippocampus*, *1*(3), 230-5.
- O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature (London)*, *381*(6581), 425-428.
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*, 171-175.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.
- O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*, 317-30.
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311-345.
- Osgood, C. E. (1949). The similarity paradox in human learning. *Psychological Review*, *56*, 132-143.
- Ostergaard, A. L. (1987). Episodic, semantic and procedural memory in a case of amnesia at an early age. *Neuropsychologia*, *25*(2), 341-57.
- Otto, T., & Eichenbaum, H. (1992). Complementary roles of the orbital prefrontal cortex and the perirhinal-entorhinal cortices in an odor-guided delayed-nonmatching-to-sample task. *Behavioral Neuroscience*, *106*(5), 762-75.
- Phillips, R. G., & Eichenbaum, H. (1998). Comparison of ventral subicular and hippocampal neuron spatial firing patterns in complex and simplified environments. *Behavioral Neuroscience*, *112*(3), 707-13.
- Pollio, H. R., Kasschau, R. A., & DeNise, H. E. (1968). Associative structure and the temporal characteristics of free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 190-197.

- Polyn, S. M., Norman, K. A., & Cohen, J. D. (2002). Connectionist modeling of source memory phenomena. *Society for Neuroscience Abstracts*, 582.5.
- Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 17, 132-138.
- Primoff, E. (1938). Backward and forward associations as an organizing act in serial and in paired-associate learning. *Journal of Psychology*, 5, 375-395.
- Quirk, G. J., Muller, R. U., & Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience*, 10(6), 2008-17.
- Quirk, G. J., Muller, R. U., Kubie, J. L., & Ranck, J. B., Jr. (1992). The positional firing properties of medial entorhinal neurons: Description and comparison with hippocampal place cells. *Journal of Neuroscience*, 12(5), 1945-1963.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, p. 207-262). New York: Academic Press.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93-134.
- Raskin, E., & Cook, S. W. (1937). The strength and direction of associations formed in the learning of nonsense syllables. *Journal of Experimental Psychology*, 20, 381-395.
- Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83(3), 190-214.
- Redish, A. (1999). *Beyond the cognitive map: From place cells to episodic memory*. Cambridge, MA: MIT Press.
- Redish, A. D., Elga, A. N., & Touretzky, D. S. (1996). A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems*, 7(4), 671-685.
- Redish, A. D., & Touretzky, D. S. (1997). Cognitive maps beyond the hippocampus. *Hippocampus*, 7, 15-35.
- Rougier, N. P., & O'Reilly, R. C. (2002). Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cognitive Science*, 26(4), 503-520.
- Rugg, M. D., Otten, L. J., & Henson, R. N. (2002). The neural basis of episodic memory: evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society B*, 357(1424), 1097-1110.
- Rugg, M. D., & Yonelinas, A. P. (2003). Human recognition memory: a cognitive neuroscience perspective. *Trends in Cognitive Science*, 7(7), 313-319.
- Russell, N. A., Horii, A., Smith, P. F., Darlington, C. L., & Bilkey, D. K. (2003). Long-term effects of permanent vestibular lesions on hippocampal spatial firing. *Journal of Neuroscience*, 23(16), 6490-6498.
- Rypma, B., & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proceedings of the National Academy of Science, USA*, 96, 6558-6563.
- Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature (London)*, 354(6349), 152-155.
- Salinas, E., & Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27(1), 15-21.
- Samsonovich, A., & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17, 5900-5920.
- Save, E., Cressant, A., Thinus-Blanc, C., & Poucet, B. (1998). Spatial firing of hippocampal place cells in blind rats. *Journal of Neuroscience*, 18(5), 1818-1826.
- Schmolck, H., Kensinger, E. A., Corkin, S., & Squire, L. R. (2002). Semantic knowledge in patient h.m. and other patients with bilateral medial and lateral temporal lobe lesions. *Hippocampus*, 12(4), 520-33.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11-21.
- Shapiro, M. L., Tanila, H., & Eichenbaum, H. (1997). Cues that hippocampal place cells encode: dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, 7(6), 624-42.
- Sharp, P. E. (1991). Computer simulation of hippocampal place cells. *Psychobiology*, 19(2), 103-115.
- Sharp, P. E., Blair, H. T., & Brown, M. (1996). Neural network modeling of the hippocampal formation spatial signals and their possible role in navigation: a modular approach. *Hippocampus*, 6(6), 720-34.
- Sharp, P. E., & Green, C. (1994). Spatial correlates of firing patterns of single cells in the subiculum of the freely moving rat. *Journal of Neuroscience*, 14(4), 2339-2356.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM — retrieving effectively from memory. *Psychonomic Bulletin and Review*, 4, 145.
- Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, 4, 637-648.
- Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (submitted). Going beyond a single list: Modeling the effects of prior experience on episodic free recall.
- Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1995). A model of the neural basis of the rat's sense of direction. *Advances in Neural Information Processing Systems*, 7, 173-80.
- Skaggs, W. E., McNaughton, B. L., Gothard, K., & Markus, E. J. (1993). An information-theoretic approach to deciphering the hippocampal code. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing 5* (p. 1030-1037). San Mateo, CA: Morgan Kaufman.
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6, 149-172.
- Slamecka, N. J. (1976). An analysis of double-function lists. *Memory & Cognition*, 4, 581-585.
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283, 1657-1661.
- Spiers, H. J., Burgess, N., Maguire, E. A., Baxendale, S. A., Hartley, T., Thompson, P. J., & O'Keefe, J. (2001). Unilateral temporal lobectomy patients show lateralized topographical and episodic memory deficits in a virtual town. *Brain*, 124(12), 2476-89.
- Spiers, H. J., Maguire, E. A., & Burgess, N. (2001). Hippocampal amnesia. *Neurocase*, 7(5), 357-82.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195-231.

- Squire, L. S., & Zola, S. M. (1998). Episodic memory, semantic memory and amnesia. *Hippocampus*, 8, 205-211.
- Stackman, R. W., Clark, A. S., & Taube, J. S. (2002). Hippocampal spatial representations require vestibular input. *Hippocampus*, 12, 291-303.
- Stern, C. E., Corkin, S., Gonzalez, R. G., et al. (1996). The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Science, USA*, 93(16), 8660-8665.
- Suzuki, W. A., & Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the New York Academy of Science*, 911, 175-91.
- Suzuki, W. A., Miller, E. K., & Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *Journal of Neurophysiology*, 78(2), 1062-1081.
- Taube, J. S. (1998). Head direction cells and the neurophysiological basis for a sense of direction. *Progress in Neurobiology*, 55(3), 225-256.
- Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. ii. effects of environmental manipulations. *Journal of Neuroscience*, 10(2), 436-47.
- Teyler, T. J., & DiScenna, P. (1985). The role of hippocampus in memory: a hypothesis. *Neuroscience and Biobehavioral Reviews*, 9(3), 377-89.
- Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100(2), 147-54.
- Thapar, A., & Greene, R. L. (1993). Evidence against a short-term-store account of long-term recency effects. *Memory & Cognition*, 21, 329-337.
- Thompson, L. T., & Best, P. J. (1989). Place cells and silent cells in the hippocampus of freely-behaving rats. *Journal of Neuroscience*, 9, 2382-90.
- Touretzky, D. S., & Redish, A. D. (1996). Theory of rodent navigation based on interacting representations of space. *Hippocampus*, 6(3), 247-70.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, 53, 1-25.
- Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus*, 8, 198-204.
- Tulving, E., & Schacter, D. L. (1991). Priming and human memory systems. *Science*, 247, 301-305.
- Van Elzaker, M., O'Reilly, R. C., & Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. i. an empirical analysis. *Hippocampus*, 13(3), 334-40.
- Vargha-Khadem, F., Gadian, D. G., & Mishkin, M. (2001). Dissociations in cognitive memory: the syndrome of developmental amnesia. *Philosophical transactions of the Royal Society of London. Series B: Biological sciences.*, 356(1413), 1435-40.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connely, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, 277, 376-380.
- Verfaellie, M., Koseff, P., & Alexander, M. P. (2000). Acquisition of novel semantic information in amnesia: effects of lesion location. *Neuropsychologia*, 38(4), 484-92.
- Vertes, R. P., Albo, Z., & Viana Di Prisco, G. (2001). Theta-rhythmically firing neurons in the anterior thalamus: implications for mnemonic functions of Papez's circuit. *Neuroscience*, 104(3), 619-25.
- Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, 281, 1188-91.
- Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 29(2), 186-210.
- Watkins, M. J., Neath, I., & Sechler, E. S. (1989). Recency effect in recall of a word list when an immediate memory task is performed after each word presentation. *American Journal of Psychology*, 102, 265-270.
- Waugh, N. C., & Norman, D. (1965). Primary memory. *Psychological Review*, 72, 89-104.
- Westmacott, R., & Moscovitch, M. (2001). Names and words without meaning: incidental postmorbid semantic learning in a person with extensive bilateral medial temporal damage. *Neuropsychology*, 15(4), 586-96.
- Whishaw, I. Q., & Maaswinkel, H. (1998). Rats with fimbria-fornix lesions are impaired in path integration: a role for the hippocampus in "sense of direction". *Journal of Neuroscience*, 18(8), 3050-3058.
- Whishaw, I. Q., McKenna, J. E., & Maaswinkel, H. (1997). Hippocampal lesions and path integration. *Current Opinion in Neurobiology*, 7, 228-234.
- Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261, 1055-8.
- Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27(3), 623-33.
- Wu, X., & Levy, W. B. (2001). Simulating symbolic distance effects in the transitive inference problem. *Neurocomputing*, 38-40, 1603-1610.
- Wu, X. B., & Levy, W. B. (1998). A hippocampal-like neural network model solves the transitive inference problem. In J. M. Bower (Ed.), *Computational neuroscience: Trends in research* (p. 567-572). New York: Plenum Press.
- Wyble, B. P., Linstor, C., & Hasselmo, M. E. (2000). Size of CA1-evoked synaptic potentials is related to theta rhythm phase in rat hippocampus. *Journal of Neurophysiology*, 83(4), 2138-44.
- Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 2, 65-74.
- Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauvé, M. J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, 5(11), 1236-41.
- Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, 12, 323-339.
- Young, B. J., Otto, T., Fox, G. D., & Eichenbaum, H. (1997). Memory representation within the parahippocampal region. *Journal of Neuroscience*, 17(13), 5183-5195.

- Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, 16(6), 2112-26.
- Zipser, D. (1985). A computational model of hippocampal place fields. *Behavioral Neuroscience*, 99(5), 1006-1018.
- Zipser, D. (1986). Biologically plausible models of place recognition and goal location. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, p. 423-470). Cambridge, MA: MIT Press.

Appendix A: A worked example illustrating the recency effect

Table 2 shows a worked example that illustrates how contextual drift results in recency in TCM. In this example, items A , B and C were presented at times 1, 2, and 3, respectively. In the example an immediate recall test was presented at time-step 4. In this case, there was no item presented and thus no contextual drift so that $\mathbf{t}_T = \mathbf{t}_3$.²⁰ At the end of the list,

$$\mathbf{M}_T^{TF} = \mathbf{M}_0^{TF} + \mathbf{f}^A \mathbf{t}'_1 + \mathbf{f}^B \mathbf{t}'_2 + \mathbf{f}^C \mathbf{t}'_3, \quad (22)$$

where \mathbf{M}_0^{TF} is the state of the matrix before the list is presented. We will assume for simplicity that there are no terms in \mathbf{M}_0^{TF} involving the item representations of the items in the list.²¹ Let us explicitly calculate the cue strength between \mathbf{t}_T and item B . This cue strength is $\mathbf{f}^{B'} \mathbf{M}_T^{TF} \mathbf{t}_T$. First, let us calculate $\mathbf{f}^{B'} \mathbf{M}_T^{TF}$:

$$\begin{aligned} \mathbf{f}^{B'} \mathbf{M}_T^{TF} &= \mathbf{f}^{B'} (\mathbf{M}_0^{TF} + \mathbf{f}^A \mathbf{t}'_1 + \mathbf{f}^B \mathbf{t}'_2 + \mathbf{f}^C \mathbf{t}'_3) \\ &= \mathbf{f}^{B'} \mathbf{M}_0^{TF} + \mathbf{f}^{B'} \mathbf{f}^A \mathbf{t}'_1 + \mathbf{f}^{B'} \mathbf{f}^B \mathbf{t}'_2 + \mathbf{f}^{B'} \mathbf{f}^C \mathbf{t}'_3 \\ &= 0 + (0) \mathbf{t}'_1 + (1) \mathbf{t}'_2 + (0) \mathbf{t}'_3 \\ &= \mathbf{t}'_2 \end{aligned} \quad (23)$$

The third line follows from our assumption that the item representations are orthonormal. Multiplying \mathbf{M}^{TF} from the left with the item representation $\mathbf{f}^{B'}$ has “picked out” only the terms in \mathbf{M}^{TF} involving \mathbf{f}^B . From this, we see that a_i for item B is just $\mathbf{t}'_2 \mathbf{t}_T$. This illustrates the statement made earlier that the cue strength between an item and a state of context is the similarity of the cue context to the states of context that obtained when the item was presented—in this case \mathbf{t}_B .

Let us explicitly calculate this quantity, using the fact that \mathbf{t}_T in this example is just \mathbf{t}_3 :

$$\begin{aligned} a_B = \mathbf{t}_2 \cdot \mathbf{t}_3 &= \mathbf{t}_2 \cdot (\rho \mathbf{t}_2 + \beta \mathbf{t}_{C_1}^{IN}) \\ &= \rho \mathbf{t}_2 \cdot \mathbf{t}_2 + \beta \mathbf{t}_2 \cdot \mathbf{t}_{C_1}^{IN} \\ &= \rho \times 1 + \beta \times 0 \end{aligned}$$

Where the last line follows from the constraint that $\|\mathbf{t}_i\| = 1$ for all i and the assumption that initially all the input vectors \mathbf{t}_i^{IN} from a random word list are orthonormal. The last column in Table 2 gives the probability that the first item recalled with \mathbf{t}_T as a cue will be A , B or C . These values add up to one, which is consistent with the definition of the probability of first recall (Laming, 1999; Howard & Kahana, 1999).

It is important to note that while TCM has been applied extensively to free recall, it does not contain any of the sampling and recovery rules that would be necessary to produce a complete description of the task (such as those specified by SAM Raaijmakers & Shiffrin, 1980, 1981).

Appendix B: Calculation of α_O and α_N on each trial

When items are presented in contexts that are similar to the input patterns they evoke (i.e. $\mathbf{t}_i^{IN} \cdot \mathbf{t}_{i-1} > 0$), the constraint that $\|\mathbf{t}_i^{IN}\| = 1$ requires that α_O and α_N be different on such trials compared to trials where there is no such similarity. If α_O and α_N were not able to change value from trial to trial to enforce the condition that $\|\mathbf{t}_i^{IN}\| = 1$, then Eq. 9 could enable $\|\mathbf{t}^{IN}\|$ to grow without bound, or decay to zero with repeated item presentations. Each time an item is presented at time step i , the constraint that the length of the input pattern when that item is repeated at time step r is unity, $\|\mathbf{t}_r^{IN}\| = 1$, leads to the equation,

$$\alpha_O^2 = \frac{1}{\gamma^2 + 2\gamma \mathbf{t}_i \cdot \mathbf{t}_i^{IN} + 1}. \quad (24)$$

If there is no similarity between \mathbf{t}_i^{IN} and \mathbf{t}_{i-1} , then $\mathbf{t}_i \cdot \mathbf{t}_i^{IN} = \beta$ from Eq. 6. The value of α_O can be determined from this equation, given γ . When $\gamma = 0$, $\alpha_O = 1$ for all presentations. When $\gamma > 0$, as in the intact case, the value of α_O depends on the similarity of the input pattern \mathbf{t}_i^{IN} to the contextual pattern \mathbf{t}_i . Once a value for α_O is calculated, α_N is then determined from the definition of γ . We assume that the initial inputs evoked by the stimuli are orthonormal, $\mathbf{t}_{A_1}^{IN} \cdot \mathbf{t}_{B_1}^{IN} = \delta_{AB}$, where δ_{AB} is one if $A = B$ and 0 otherwise.

Appendix C: A worked example illustrating associative effects

Table 3 shows a worked example illustrating the associative effects attributed to the two components. In this example, five items, A through E are presented in sequence at time steps 1 through 5. We assume that an infinitely long delay intervenes before the recall test, such that $\mathbf{t}_{T-1} \cdot \mathbf{t}_i = 0$ for all the items in the list. To illustrate the associative effects caused by retrieved context, at the time of test T , we present item C as a cue for recall of the other items in the list. In treating free recall, previous studies (Howard & Kahana, 2002a; Howard et al., In revision; Howard, 2004) have presented a just-recalled item to the network as a cue to retrieve other items to generate a CRP function. Equation 9 tells us that the context retrieved by C when it is presented the second time, as a cue, will be a combination of the state initially evoked, plus the state of context that obtained when C was initially presented:

$$\mathbf{t}_{C_2}^{IN} = \alpha_O \mathbf{t}_{C_1}^{IN} + \alpha_N \mathbf{t}_{C_1}. \quad (25)$$

²⁰ Under some circumstances, it might be desirable to consider that the “recall signal” itself causes some degree of contextual drift.

²¹ We could also assume that there was no overlap between the pre-list contexts and the test context.

Table 2

Worked example showing the values of various quantities necessary to calculate probability of recall for several items. In this example, items A, B and C are presented at times 1, 2, and 3 respectively. The recall test was given at time T. For convenience, τ has been set to 2. The last column shows the probability of recalling each item first if $\rho = 0.7$.

i	\mathbf{f}_i	\mathbf{t}_i	$\Delta\mathbf{M}^{TF}$	a_i	$P(i \mathbf{t}_T)$
0	$\mathbf{0}$	\mathbf{t}_0	$\mathbf{0}$		
1	\mathbf{f}^A	$\rho\mathbf{t}_0 + \beta\mathbf{t}_{A_1}^{IN}$	$\mathbf{f}_A\mathbf{t}'_1$	$\mathbf{t}_1 \cdot \mathbf{t}_3 = \rho^2$	0.26
2	\mathbf{f}^B	$\rho\mathbf{t}_1 + \beta\mathbf{t}_{B_1}^{IN}$	$\mathbf{f}_B\mathbf{t}'_2$	$\mathbf{t}_2 \cdot \mathbf{t}_3 = \rho$	0.32
3	\mathbf{f}^C	$\rho\mathbf{t}_2 + \beta\mathbf{t}_{C_1}^{IN}$	$\mathbf{f}_C\mathbf{t}'_3$	$\mathbf{t}_3 \cdot \mathbf{t}_3 = 1$	0.43
T	$\mathbf{0}$	\mathbf{t}_3	$\mathbf{0}$		

Table 3

Worked example showing the values of various quantities necessary to calculate probability of recall for several items. In this example, items A, B, C, D and E are presented at times 1, 2, and 3 respectively. After a very long delay, item C was presented as a cue for a recall test at time T. For convenience, τ has been set to 2, and γ set to one. The last column shows the probability of recalling each item in response to C as a cue for $\rho = 0.7$.

i	\mathbf{f}_i	\mathbf{t}_i	$\Delta\mathbf{M}^{TF}$	a_i	$P(i \mathbf{t}_T)$
0	$\mathbf{0}$	\mathbf{t}_0	$\mathbf{0}$		
1	\mathbf{f}^A	$\rho\mathbf{t}_0 + \beta\mathbf{t}_{A_1}^{IN}$	$\mathbf{f}_A\mathbf{t}'_1$	$\mathbf{t}_1 \cdot \mathbf{t}_T = \beta\alpha_N\rho^2$	0.22
2	\mathbf{f}^B	$\rho\mathbf{t}_1 + \beta\mathbf{t}_{B_1}^{IN}$	$\mathbf{f}_B\mathbf{t}'_2$	$\mathbf{t}_2 \cdot \mathbf{t}_T = \beta\alpha_N\rho$	0.23
3	\mathbf{f}^C	$\rho\mathbf{t}_2 + \beta\mathbf{t}_{C_1}^{IN}$	$\mathbf{f}_C\mathbf{t}'_3$		
4	\mathbf{f}^D	$\rho\mathbf{t}_3 + \beta\mathbf{t}_{D_1}^{IN}$	$\mathbf{f}_D\mathbf{t}'_3$	$\mathbf{t}_4 \cdot \mathbf{t}_T = \beta\rho(\alpha_N + \beta\alpha_O)$	0.29
5	\mathbf{f}^E	$\rho\mathbf{t}_4 + \beta\mathbf{t}_{E_1}^{IN}$	$\mathbf{f}_E\mathbf{t}'_3$	$\mathbf{t}_5 \cdot \mathbf{t}_T = \beta\rho^2(\alpha_N + \beta\alpha_O)$	0.26
\vdots					
T	\mathbf{f}^C	$\rho\mathbf{t}_{T-1} + \beta\mathbf{t}_{C_2}^{IN}$	$\mathbf{f}^C\mathbf{t}'_T$		

For this example, we assume that $\gamma = 1$, meaning that $\alpha_O = \alpha_N$. We also assume that all of the initial inputs are orthonormal, and orthogonal to \mathbf{t}_0 and \mathbf{t}_{T-1} . Now, the state of context used to cue recall of A, B, D and E is just

$$\begin{aligned}\mathbf{t}_T &= \rho\mathbf{t}_{T-1} + \beta\mathbf{t}_{C_2}^{IN} \\ &= \rho\mathbf{t}_{T-1} + \beta(\alpha_O\mathbf{t}_{C_1}^{IN} + \alpha_N\mathbf{t}_{C_1})\end{aligned}$$

First, let's explicitly calculate the cue strength of \mathbf{t}_T to B:

$$\begin{aligned}a_B &= \mathbf{t}_{B_1} \cdot \mathbf{t}_T \\ &= \mathbf{t}_{B_1} \cdot (\rho\mathbf{t}_{T-1} + \beta\mathbf{t}_{C_2}^{IN}) \\ &= \mathbf{t}_{B_1} \cdot [\rho\mathbf{t}_{T-1} + \beta(\alpha_O\mathbf{t}_{C_1}^{IN} + \alpha_N\mathbf{t}_{C_1})] \\ &= \rho\mathbf{t}_{B_1} \cdot \mathbf{t}_{T-1} + \beta\alpha_O\mathbf{t}_{B_1} \cdot \mathbf{t}_{C_1}^{IN} + \beta\alpha_N\mathbf{t}_{B_1} \cdot \mathbf{t}_{C_1} \\ &= \rho \times 0 + \beta\alpha_O \times 0 + \beta\alpha_N\mathbf{t}_{B_1} \cdot (\rho\mathbf{t}_{B_1} + \beta\mathbf{t}_{C_1}^{IN}) \\ &= \beta\alpha_N\rho\end{aligned}$$

In particular, the last line takes advantage of the fact that $\mathbf{t}_{B_1} \cdot \mathbf{t}_{C_1}^{IN} = 0$. Note that if $\alpha_N = 0$, then $a_B = 0$ —the cue strength of the item immediately preceding the cue goes to zero. Now let's calculate the cue strength of \mathbf{t}_T to A. Picking the derivation up further in than the last one, we find

$$\begin{aligned}a_A &= \mathbf{t}_{A_1} \cdot \mathbf{t}_T \\ &= \mathbf{t}_{A_1} \cdot \beta(\alpha_O\mathbf{t}_{C_1}^{IN} + \alpha_N\mathbf{t}_{C_1})\end{aligned}$$

$$\begin{aligned}&= \beta\alpha_O\mathbf{t}_{B_1} \cdot \mathbf{t}_{C_1}^{IN} + \beta\alpha_N\mathbf{t}_{A_1} \cdot \mathbf{t}_{C_1} \\ &= \beta\alpha_O \times 0 + \beta\alpha_N\mathbf{t}_{A_1} \cdot (\rho\mathbf{t}_{B_1} + \beta\mathbf{t}_{C_1}^{IN}) \\ &= \beta\alpha_N\mathbf{t}_{A_1} \cdot [\rho(\mathbf{t}_{B_1} + \beta\mathbf{t}_{C_1}^{IN}) + \beta\mathbf{t}_{C_1}^{IN}] \\ &= \beta\alpha_N(\rho^2 \times 1 + \beta\rho \times 0 + \beta \times 0)\end{aligned}$$

We see from this that the cue strength of A, the item two before the cue, is also zero if $\alpha_N = 0$. When not zero, it lower than the cue strength for B, because it includes an extra factor of ρ . This illustrates the contiguity effect—items closer to the cue have a higher cue strength (and are thus more likely to be recalled).

In the forward direction, not only does the \mathbf{t}_{C_1} term from $\mathbf{t}_{C_2}^{IN}$ contribute to the cue strength, but so does the $\mathbf{t}_{C_1}^{IN}$ term. This is so because the context from items that followed C_1 include a term with $\mathbf{t}_{C_1}^{IN}$. For instance,

$$\begin{aligned}\mathbf{t}_{D_1} &= \rho\mathbf{t}_{C_1} + \beta\mathbf{t}_{D_1}^{IN} \\ &= \rho(\rho\mathbf{t}_{B_1} + \beta\mathbf{t}_{C_1}^{IN}) + \beta\mathbf{t}_{D_1}^{IN}.\end{aligned}$$

Where the second line follows from the first by expanding \mathbf{t}_{C_1} using Eq. 6. From this we can see that $\mathbf{t}_{D_1} \cdot \mathbf{t}_{C_1}^{IN} = \rho\beta$. As a consequence,

$$\begin{aligned}a_D &= \beta\mathbf{t}_{D_1} \cdot \mathbf{t}_{C_2}^{IN} \\ &= \beta\mathbf{t}_{D_1} \cdot (\alpha_O\mathbf{t}_{C_1}^{IN} + \alpha_N\mathbf{t}_{C_1})\end{aligned}$$

$$\begin{aligned}
&= \beta(\alpha_O \mathbf{t}_{D_1} \cdot \mathbf{t}_{C_1}^{IN} + \alpha_N \mathbf{t}_{D_1} \cdot \mathbf{t}_{C_1}) \\
&= \beta(\alpha_O \rho \beta + \rho \alpha_N) \\
&= \beta \rho (\alpha_O \beta + \alpha_N)
\end{aligned}$$

We can see that this is greater than a_B if $\alpha_O > 0$. This implements associative asymmetry. The cue strength to E is just this expression with an additional factor of ρ :

$$a_E = \rho a_D, \quad (26)$$

showing evidence for contiguity.

Appendix D: Derivation of transitive associations

The following derivation assumes that in a first stage of learning, $A \rightarrow B$ is presented, followed by $B \rightarrow C$ in a second stage of learning. Each pair is presented just once in this derivation. We will refer to the state of context prior to presentation of A in the first stage of learning as \mathbf{t}_1 and the state prior to the presentation of B in the second stage as \mathbf{t}_2 . We assume that the delay between phases of learning is infinitely long so that $\mathbf{t}_1 \cdot \mathbf{t}_2 = 0$. So, item A is presented at time step A_1 , item B is presented at time step B_1 , and then later at time step B_2 , and item C is presented at time step C_2 . We assume that the initial inputs from each item ($\mathbf{t}_{A_1}^{IN}$, $\mathbf{t}_{B_1}^{IN}$ and $\mathbf{t}_{C_2}^{IN}$, but not $\mathbf{t}_{B_2}^{IN}$) are orthonormal (meaning, orthogonal and of unit length), as well as orthogonal to the initial contexts \mathbf{t}_1 and \mathbf{t}_2 . We denote the time of test, when one of the items is repeated as a cue as time step r . We assume that there is an infinite delay prior to test so that $\mathbf{t}_{r-1} \cdot \mathbf{t}_{C_2} = 0$.²²

For the first stage of learning we have:

$$\begin{aligned}
\mathbf{t}_{A_1} &= \rho \mathbf{t}_1 + \beta \mathbf{t}_{A_1}^{IN} \\
\mathbf{t}_{B_1} &= \rho \mathbf{t}_{A_1} + \beta \mathbf{t}_{B_1}^{IN}.
\end{aligned}$$

During the second stage of learning we have:

$$\begin{aligned}
\mathbf{t}_{B_2} &= \rho \mathbf{t}_2 + \beta \mathbf{t}_{B_2}^{IN} \\
\mathbf{t}_{C_2} &= \rho \mathbf{t}_{B_2} + \beta \mathbf{t}_{C_2}^{IN}.
\end{aligned}$$

The state of context at time of test, \mathbf{t}_r , serves as the retrieval cue. This will include the input from the cue item (for instance, $\mathbf{t}_{A_r}^{IN}$ if item A is the cue), as well as the state of context \mathbf{t}_{r-1} prior to presentation of the cue. This component could be responsible for a recency effect (Murdock, 1963b, 1963c, 1963a), but we have assumed that there is an infinite delay so that \mathbf{t}_{r-1} is not an effective retrieval cue for any of the stimuli. In this case, the cue strength is solely determined by the input pattern (e.g. $\mathbf{t}_{A_r}^{IN}$) retrieved by the cue item. If item A is presented as a cue at time-step A_r , then the cue strength of item B is

$$\beta \mathbf{t}_{A_r}^{IN} \cdot (\mathbf{t}_{B_1} + \mathbf{t}_{B_2}). \quad (27)$$

Similarly, the cue strength from A to C after learning is just

$$\beta \mathbf{t}_{A_r}^{IN} \cdot \mathbf{t}_{C_2}. \quad (28)$$

The cue pattern $\mathbf{t}_{A_r}^{IN}$ will be a function of $\mathbf{t}_{A_1}^{IN}$ and \mathbf{t}_{A_1} , according to Eq. 9. To determine the value for each of these cue strengths, we just need to expand the \mathbf{t} 's far enough using Eqs. 6 and 9 so that their relationship to $\mathbf{t}_{A_1}^{IN}$ and \mathbf{t}_{A_1} is made clear.

First we will show that when A is presented as a cue, the cue strength to B is non-zero even when α_N is zero. The cue strength from A to B is proportional to

$$\begin{aligned}
\mathbf{t}_{A_r}^{IN} \cdot (\mathbf{t}_{B_1} + \mathbf{t}_{B_2}) &= (\alpha_O \mathbf{t}_{A_1}^{IN} + \alpha_N \mathbf{t}_{A_1}) \bullet \\
&\quad (\rho \mathbf{t}_{A_1} + \beta \mathbf{t}_{B_1}^{IN} + \rho \mathbf{t}_2 + \beta \mathbf{t}_{B_2}^{IN}) \\
&= (\alpha_O \mathbf{t}_{A_1}^{IN} + \alpha_N \mathbf{t}_{A_1}) \bullet \\
&\quad [\rho \mathbf{t}_{A_1} + \beta \mathbf{t}_{B_1}^{IN} + \rho \mathbf{t}_2 + \beta (\alpha_O \mathbf{t}_{B_1}^{IN} + \alpha_N \mathbf{t}_{B_1})] \\
&= \alpha_O \rho \mathbf{t}_{A_1}^{IN} \cdot \mathbf{t}_{A_1} + \alpha_N \rho \mathbf{t}_{A_1} \cdot \mathbf{t}_{A_1} + \alpha_N^2 \beta \mathbf{t}_{A_1} \cdot \mathbf{t}_{B_1} \\
&= \rho (\alpha_O \beta + \alpha_N + \alpha_N^2 \beta).
\end{aligned}$$

This value is non-zero even when $\alpha_N = 0$. Similarly, the cue strength from B to C is proportional to:

$$\begin{aligned}
\mathbf{t}_{B_r}^{IN} \cdot \mathbf{t}_{C_2} &= (\alpha_O \mathbf{t}_{B_2}^{IN} + \alpha_N \mathbf{t}_{B_2}) \cdot \mathbf{t}_{C_2} \\
&= \alpha_O \rho \beta + \alpha_N \rho
\end{aligned}$$

Again, this is non-zero even if $\alpha_N = 0$. Learning of the premise pairs can proceed even in the absence of new item-to-context learning.

In contrast, a non-zero transitive association between A and C depends completely on the existence of new item-to-context learning. The cue strength from A to C is given by:

$$\begin{aligned}
\mathbf{t}_{A_r}^{IN} \cdot \mathbf{t}_{C_2} &= (\alpha_O \mathbf{t}_{A_1}^{IN} + \alpha_N \mathbf{t}_{A_1}) \bullet (\rho \mathbf{t}_{B_2} + \beta \mathbf{t}_{C_2}^{IN}) \\
&= (\alpha_O \mathbf{t}_{A_1}^{IN} + \alpha_N \mathbf{t}_{A_1}) \bullet [\rho (\rho \mathbf{t}_2 + \beta \mathbf{t}_{B_2}^{IN}) + \beta \mathbf{t}_{C_2}^{IN}] \\
&= (\alpha_O \mathbf{t}_{A_1}^{IN} + \alpha_N \mathbf{t}_{A_1}) \bullet \\
&\quad \{ \rho [\rho \mathbf{t}_2 + \beta (\alpha_O \mathbf{t}_{B_1}^{IN} + \alpha_N \mathbf{t}_{B_1})] + \beta \mathbf{t}_{C_2}^{IN} \} \\
&= \alpha_O \alpha_N \rho \beta (\mathbf{t}_{A_1}^{IN} \cdot \mathbf{t}_{B_1}) + \alpha_N^2 \rho \beta (\mathbf{t}_{A_1} \cdot \mathbf{t}_{B_1}) \\
&= \alpha_O \alpha_N \rho^2 \beta^2 + \alpha_N^2 \rho^2 \beta \\
&= \alpha_N \rho^2 \beta (\alpha_O \beta + \alpha_N)
\end{aligned}$$

The last line is Eq. 17 from the main text. Clearly, this cue strength goes to zero if $\alpha_N = 0$, demonstrating that transitive associations depend on new item-to-context learning.

Within this framework, transitive associations develop because of the context retrieved during the second presentation of B . When B becomes bound to contextual elements from A , these elements form part of the contextual representation associated with C , leading ultimately to the transitive association. To make this explicit, when B is presented the second

²² These simplifying assumptions enable us to avoid changes in α_O and α_N that would occur as a consequence of the assumption that $\|\mathbf{t}_i^{IN}\|$. If there is similarity between the input \mathbf{t}_i^{IN} and the prior context \mathbf{t}_i , then α_O and α_N must be adjusted. See the discussion of normalization in the description of the simulation for more information.

time as part of $B \rightarrow C$,

$$\mathbf{t}_{B_2}^{IN} = \alpha_O \mathbf{t}_{B_1}^{IN} + \alpha_N \mathbf{t}_{B_1}. \quad (29)$$

This second term, \mathbf{t}_{B_1} overlaps considerably with the contextual elements retrieved by A :

$$\mathbf{t}_{B_1} = \rho \mathbf{t}_{A_1} + \beta \mathbf{t}_{B_1}^{IN} \quad (30)$$

The contextual state associated with C , \mathbf{t}_{C_2} , includes $\mathbf{t}_{B_2}^{IN}$. When $\alpha_N > 0$, the context retrieved by B on its second presentation includes \mathbf{t}_{A_1} and $\mathbf{t}_{A_1}^{IN}$. In the presence of new item-to-context learning, contextual elements originating from A are associated to C . In the absence of item-to-context learning (*i.e.* when $\alpha_N = 0$), then only a stimulus-specific representation from B contributes to C 's context. In this case there will be no $A \rightarrow C$ association.