Journal of Mathematical Psychology
doi:10.1006/jmps.2001.1388, available online at http://www.idealibrary.com on IDEAL®

A Distributed Representation of Temporal Context

Marc W. Howard and Michael J. Kahana

Brandeis University

The principles of recency and contiguity are two cornerstones of the theoretical and empirical analysis of human memory. Recency has been alternatively explained by mechanisms of decay, displacement, and retroactive interference. Another account of recency is based on the idea of variable context (Estes, 1955; Mensink & Raaijmakers, 1989). Such notions are typically cast in terms of a randomly fluctuating population of elements reflective of subtle changes in the environment or in the subjects' mental state. This random context view has recently been incorporated into distributed and neural network memory models (Murdock, 1997; Murdock, Smith, & Bai, 2001). Here we propose an alternative model. Rather than being driven by random fluctuations, this formulation, the *temporal context model* (TCM), uses retrieval of prior contextual states to drive contextual drift. In TCM, retrieved context is an inherently asymmetric retrieval cue. This allows the model to provide a principled explanation of the widespread advantage for forward recalls in free and serial recall. Modeling data from single-trial free recall, we demonstrate that TCM can simultaneously explain recency and contiguity effects across time scales.
Key Words: free recall; temporal context; episodic memory.

1. THE SCALE-INVARIANCE OF RECENCY AND CONTIGUITY

We seek to explain two basic principles of human episodic memory—the principle of recency and the principle of contiguity. The principle of recency refers to the ubiquitous finding that memory performance declines with the passage of time or the presence of intervening items. The principle of contiguity refers to the equally general finding that recall of an item is facilitated by the presentation or recall of another item that occurred close in time to the target item. Both of these principles relate memory performance to temporal factors—the time since an item was

The author acknowledge support from National Institute of Health Research Grants MH55687 and AG15685. Marc Howard is now at the Department of Psychology, Boston University, 64 Cunningham Street, Boston, MA 02215.

Address correspondence and reprint requests to Michael J. Kahana, Volen Center for Complex Systems, Brandeis University, MS 013, Waltham, MA 02254. E-mail: kahana@brandeis.edu.

0022-2496/01 $35.00

presented in the case of recency, and the time between item presentations in the case of contiguity.

1.1. Recency, Contiguity, and Asymmetric Association in Free Recall

In free recall, subjects are presented with a list of words. Their task is to recall the words in any order. Data derived from this task have been used to constrain models of short- and long-term memory (e.g., Atkinson & Shiffrin, 1968; Glanzer, 1972), models of episodically formed associations in long-term memory (Raaijmakers & Shiffrin, 1980, 1981b), and models of the semantic organization of long-term memory (e.g., Tulving, 1968).

In free recall, the principle of recency may be seen in the serial position curve (e.g., Murdock, 1962). The last several list items are generally remembered much better than earlier list items. This end-of-list recency effect can also be seen in subjects' tendency to initiate recall from the end of the list. Howard and Kahana (1999) have shown that the probability of first recall, a serial position curve for the very first item subjects recall, is a particularly sensitive measure of the end-of-list recency effect.

After an item is recalled from a given serial position, the item recalled next tends to come from a nearby serial position (Kahana, 1996). This tendency, which we refer to as the *lag-recency effect*, illustrates the principle of contiguity. Although informally suspected since the earliest studies of free recall, the lag-recency effect was first detailed by Kahana (1996). It is measured by the conditional response probability (CRP; Howard & Kahana, 1999; Kahana 1996) plotted as a function of the lag (or separation) of the studied items. The CRP function has a characteristic shape. It is peaked in the middle around lag zero, indicating that recalls to nearby serial positions are more likely than recalls to remote serial positions. It is also asymmetric, with a greater probability of making recalls in the forward, rather than backward direction (see Fig. 1).

Since Ebbinghaus, theorists have assumed that forward associations are stronger than backward associations (e.g., Ebbinghaus, 1885/1913; Gillund & Shiffrin, 1984; Johnson, 1991). Asymmetry is clearly a very robust feature of free recall (see Fig. 1). Although retrieval of paired associates is often symmetric (Asch & Ebenholtz, 1962; Kahana, 1999; Rizzuto & Kahana, 2001), retrieval in free and serial recall shows marked asymmetries. In serial recall, evidence for associative asymmetry comes from free association to probes from serial lists (McGeoch, 1936; Raskin & Cook, 1937), probed recall of serial lists (Kahana & Caplan, in press), and dissociations between forward and backward recall (Li & Lewandowsky, 1993, 1995). Despite the overwhelming evidence for asymmetric association in free and serial recall, and its importance in models of these tasks, there has been essentially no work on mechanisms causing associative asymmetry.

Another striking feature of the serial position curve in free recall is the primacy effect—an increase in the probability of recalling the first few list items (relative to the items in the middle of the list). Numerous studies have linked the primacy effect with rehearsal: Items from the beginning of the list receive a larger number of rehearsals than do items from other list positions (Rundus, 1971; Tan & Ward,

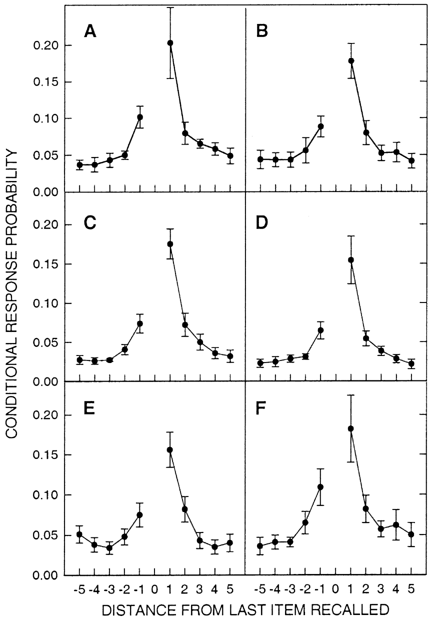

FIG. 1. The lag-recency effect in free recall. Shown are conditional response probability (CRP) curves from several large free recall studies (Murdock, 1962; Murdock & Metcalfe, 1978; Murdock & Okada, 1970). Each curve measures the probability of making a recall to serial position $i+\text{lag}$ immediately following recall of serial position i. All of the CRP curves show the same characteristic properties: an advantage for nearby recalls and an asymmetry favoring forward recalls. The advantage for nearby recalls is shown by the peak in the center of the curves. The asymmetry is evident by comparing points in the forward direction (e.g., $+1$ with analogous points in the backward direction (e.g., -1). Reproduced from Kahana (1996). A. CRP from the LL$=$20, 2 s/item condition of Murdock (1962). B. CRP from the LL$=$20, 1 s/item condition of Murdock (1962). C. CRP from the LL$=$30, 1 s/item condition of Murdock (1962). D. CRP from the LL$=$40, 1 s/item condition of Murdock (1962). E. CRP collapsed over conditions of Murdock and Okada (1970). F. CRP from Murdock and Metcalfe (1978).

2000). When rehearsal is disrupted, the primacy effect virtually disappears (Glenberg et al., 1980; Howard & Kahana, 1999; Marshall & Werder, 1972; Rundus, 1980; Watkins, Neath, & Sechler, 1989). When the serial position curve is plotted as a function of rehearsal position, rather than the presentation position of the items (as in the functional serial position curve of Brodie & Murdock, 1977; Tan & Ward, 2000), the result is a pure recency effect.

In addition to the large primacy effect attributable to rehearsal, there is also a residual one-position primacy effect in the probability of first recall that does not seem to be dependent on rehearsal (Howard & Kahana, 1999; Laming, 1999). A one-position primacy effect is also sometimes found in item recognation experiments where verbal rehearsal is unlikely (e.g., Monsell, 1978; Neath, 1993; Neath & Crowder, 1996; Wright, Santiago, Sands, Kendrick, & Cook, 1985).

1.2. Approximate Scale Invariance

Both recency and contiguity effects are approximately invariant across time scales. The long-term recency effect observed in continuous-distractor free recall (Bjork & Whitten 1974) demonstrates the scale invariance of the end-of-list recency effect. In the continuous-distractor technique, a period of distractor activity separates each of the items in the list. The length of this interval is called the interpresentation interval (IPI). A period of distractor activity also intervenes between presentation of the last item in the list and the free recall test. The duration of this period is called the retention interval (RI). The presence or absence of a recency effect depends not so much on the magnitude of the delays, but very much on the relative values of the IPI and RI (but see Nairne, Neath, Serra, & Byun, 1997). Because recall is a function of relative rather than absolute "strength," approximate scale invariance of this type implicates a competitive retrieval mechanism.

Surprisingly, a similar scale invariance also appears to hold for the principle of contiguity. Howard and Kahana (1999) found that the lag-recency effect is observed in continuous-distractor free recall when subjects engage in 16 s of a demanding arithmetic distractor task between list items. Because this same distractor was sufficient to greatly attenuate the end-of-list recency effect, it is unlikely that the lag-recency effect is a consequence of direct interitem associations formed as a result of co-occupancy in short-term store (STS).[1] Because the lag-recency effect probably does not depend on direct interitem association, we hypothesize that it is the result of some type of temporally sensitive construct. The shape of the CRP curve (an index of the magnitude of the lag-recency effect) does not measurably change with a large increase in the temporal separation of items: the lag-recency effect, like the end-of-list recency effect, is approximately scale invariant. Notably, the asymmetry seen in the lag-recency effect in standard free recall persists in continuous-distractor free recall.

1.3. Toward a Unified Account of Recency and Contiguity

The key to developing a unified account of recency and contiguity in free recall is to find a single construct that measures to different intervals. According to classic strength theory, each item's strength is incremented when the item is studied, and then it decays over time. At test, items compete for recall, with the stronger items being favored. Such a model produces end-of-list recency because recent items have high strength. By extension, it can then be said that strength "measures" the time interval between study and test. To adapt this strength framework to account for the lag-recency effect, we must assume that retrieval of an item strengthens items that come from nearby list positions. In this enhanced model, strength is measuring two intervals: the one between study and test, and the one between the presentation of two different studied items. In effect, successful recall of an item allows our strength construct to "jump back in time," making remote items that were studied close in time to the just-recalled item appear recent.

Like classic strength models, random context models (e.g., Anderson & Bower, 1972; Estes, 1955; Mensinck & Raaijmakers, 1988, 1989; Murdock, 1997; Murdock

[1] In the SAM model of free recall (Raaijmakers & Shiffrin, 1980, 1981b), co-occupancy in STS produces a lag-recency effect that is well matched to the data when there is no interitem distractor (Kahana, 1996).

et al., 2001) are also sensitive to the time between study and test. Estes's (1955) stimulus sampling theory, for example, defines context as a set of binary elements, some of which are "active," while others are not. Context fluctuates from moment to moment, with active elements turning off with probability p and inactive elements turning on with probability q. Studying an item creates an association between the item and the current state of context. At test, cueing with the current state of context will tend to activate recent items, because the number of overlapping active elements decreases as the time interval increases. In this way, random context models can account for the recency effect observed in immediate free recall.

Howard and Kahana (1999) added a contextual retrieval process to a simple random context model. They assumed that when an item is recalled, the population of elements is reset to the state it was in when the item was presented. This retrieved contextual state is then used as the cue for subsequent recalls; it is a better cue for items close in the list to the just-recalled item. If random context is retrieved, it serves as a temporally sensitive construct that measures the distance from the just-recalled item—that is, it jumps back in time. When coupled with a competitive retrieval mechanism, this retrieved random context model explains recency and contiguity across time scales (Howard & Kahana, 1999).

By assuming that context is retrieved, Howard and Kahana (1999) showed that a *retrieved* random context model could account for the basic pattern of recency and contiguity effects across time scales. Although it provided a decent fit to the data, this solution was superficial. How is it, upon recall of item i, that all of the contextual elements were able to be reset to their values at the time of presentation of item i? The lack of a process by which this can be accomplished is a weakness—a lack of constraint. This is troubling in that sometimes we require the contextual state to drift and at other times to jump back in time. How does the system know when to jump back in time and when to drift? Imagine a sequence of items A B C D E F. Suppose that after presentation of item F, we retrieve the context from item B (because it is recalled or presented as a cue). Because F is remote from B, and we have assumed perfect retrieval of B's context, the contextual cue no longer treats F as recent. Intuitively, this seems like an undesirable property.

In addition, some important features of the data required an *ad hoc* solution. The lag-recency effect (as seen in the shape of the CRP curves) shows a reliable asymmetry. Forward recalls are more likely than backward recalls (Howard & Kahana, 1999; Kahana, 1996). We solved this problem by simply adding a free parameter to differentially weight forward and backward associations. Given that asymmetry in free recall is an extremely robust result, and that asymmetry in memory would seem to be a very useful property, a principled account of asymmetry is required.

2. CONTEXT IN MODELS OF MEMORY

In this section we start by describing how contextual cues can be implemented in distributed memory models (DMMs). We then introduce a general framework in which to describe and contrast both traditional random context models (Estes, 1955; Mensink & Raaijmakers, 1988; Murdock, 1997) and the temporal context model presented here.

2.1. Contextual Cues in Distributed Memory Models

DMMs assume that perceptual recognition systems first break down incoming information into meaningful units. Each unit is then represented by a set of abstract feature values; mathematically, this set describes a vector in a high dimensional feature space. These vectors are taken to correspond to items in semantic memory. Let us denote the space on which these item vectors reside as F and specific vectors as $\mathbf{f}$. The state on F at time i is denoted as $\mathbf{f}_i$. We refer to individual items with Greek superscripts and use italic subscripts to refer to a particular occurrence of an item. For instance, if we had a list of words, ABSENCE, HOLLOW, PUPIL, ABSENCE, and the item representation for ABSENCE was $\mathbf{f}^{\gamma}$ then $\mathbf{f}^{\gamma} = \mathbf{f}_1 = \mathbf{f}_4$. For simplicity, we will assume orthonormal item representations throughout.

To accomplish episodic tasks, the brain must distinguish between nominally identical items that are encountered in different places and at different times. In distributed memory models (e.g., Chappell & Humphreys, 1994), a distributed representation of list context has been used to perform this discrimination.[2] List context, in these models, is a random vector chosen anew for each list (see also Anderson & Bower, 1972). This leads to a high degree of similarity between context vectors corresponding to the same list and a much lower similarity between context vectors corresponding to different lists. By linking item and context representations in memory, one could accomplish two useful functions. First, one could determine whether a specific item occurred in a specific list (episodic recognition). Second, one can use a state of context to cue item representations for recall (episodic recall).

In order to use context as a cue for semantic memory, we need a way to connect the context representations to the item representations. Let us denote the state of context at time i as $\mathbf{t}_i$. This vector describes a pattern of activity across another space T. To connect T with F, we will use a matrix $\mathbf{M}^{TF}$. This matrix represents the strength of the connection from each element in T to each element in F. We will assume that this matrix is formed from a set of outer product terms

$$\mathbf{M}^{TF} = \sum_{i=1}^{L} \mathbf{f}_i \mathbf{t}'_i, \tag{1}$$

where $\mathbf{t}'$ is the transpose of the vector $\mathbf{t}$ and the sum is over all of the items in the current list. This matrix implements the association between a state of context on T with a particular item, a state on F. For simplicity, we assume that $\mathbf{M}^{TF}$ is reset at the beginning of each list (cf. Murdock & Kahana, 1993). Later we will clarify the rationale for this simplification and discuss some alternative solutions.

A state on T, $\mathbf{t}_j$, will provide an input, $\mathbf{f}^{IN} \equiv \mathbf{M}^{TF}\mathbf{t}_j$ to F via $\mathbf{M}^{TF}$, the matrix of connections from F to T. We define an activation, a_i, representing the similarity of the input, $\mathbf{F}^{IN}$, to a given item, $\mathbf{f}_i$.

[2] Although not a distributed memory model, the SAM model makes extensive use of fixed list context in recall and recognition (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980, 1981a).

$$
\begin{aligned}
a_i &\equiv \mathbf{f}^{IN} \cdot \mathbf{f}_i \\
&= \mathbf{M}^{TF} \mathbf{t}_j \cdot \mathbf{f}_i \\
&= \sum_{k=1}^{L} (\mathbf{t}_k \cdot \mathbf{t}_j)\, \mathbf{f}_k \cdot \mathbf{f}_i \\
&= \sum_{\mathbf{f}_k = \mathbf{f}_i} \mathbf{t}_k \cdot \mathbf{t}_j .
\end{aligned} \tag{2}
$$

The last line follows from our assumption of orthonormality of the $\mathbf{f}_i$s. The effectiveness of some contextual state $\mathbf{t}_j$ for an item $\mathbf{f}_i$ will be a function of the similarity of $\mathbf{t}_j$ to the context at the times $\mathbf{f}_i$ was presented. If $\mathbf{f}_i$ was only presented once in the experiment, as we shall assume in the applications here, this is just $\mathbf{t}_i \cdot \mathbf{t}_j$.

2.2. A Framework for Context Models

The previous section demonstrated that a contextual state can be used to provide a cue to activate items in semantic memory. The effectiveness of a context cue for an item depends on its similarity to the context in which the item was encoded. This section describes a theoretical framework for constructing contextual representations that change gradually over time.

Assume that the state of context at time i, denoted $\mathbf{t}_i$, is a vector that follows the process

$$\mathbf{t}_i = \rho\, \mathbf{t}_{i-1} + \mathbf{t}_i^{IN}, \tag{3}$$

where $\mathbf{t}_i^{IN}$ is the input at time i, and $0 \leqslant \rho \leqslant 1$ determines the rate of contextual drift. We refer to this process as the drift equation. The new state of context, $\mathbf{t}_i$, is derived from the previous state, $\mathbf{t}_{i-1}$, and the input, $\mathbf{t}_i^{IN}$. Because the new state is derived from the previous state, and the previous state in turn was derived from its predecessor, context changes gradually.

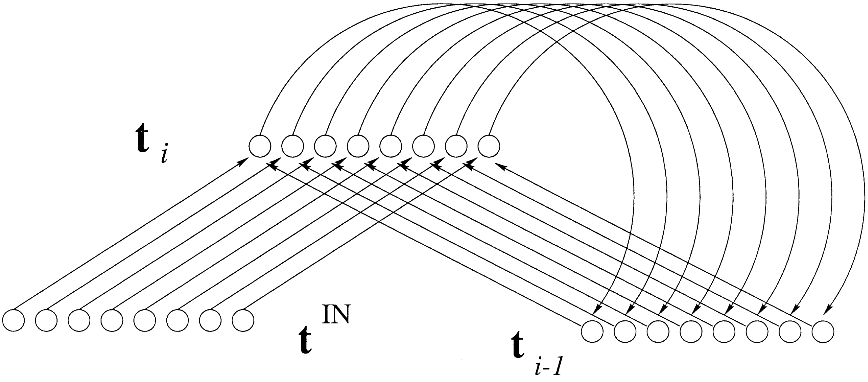

FIG. 2. Recurrency gives rise to slowly varying context, as in Eq. (3.) Equation (3) states that at each time step the current state of context $\mathbf{t}_i$ is constructed from the input vector $\mathbf{t}_i^{IN}$ and the previous state of context $\mathbf{t}_{i-1}$. One way to implement this would be through a recurrent network, as illustrated above (Elman, 1990).

Figure 2 illustrates one possible way to implement the drift equation (Eq. 3).[3] The essential feature is some form of recurrency, so that the previous state can contribute to the new state. Although Fig. 2 shows a separate set of units for the prior state and the current state, this recurrency could occur *via* delay lines within a single layer (Jordan, 1986).

2.2.1. Random Context Models

When $\mathbf{t}^{IN}$ is noise, $\mathbf{t}$ describes a random context model (e.g., Estes, 1955; Mensink & Raaijmakers, 1989; Murdock, 1997), as illustrated in Fig. 3. Murdock (1997) adopts Eq. (3) with $\mathbf{t}_i^{IN} = \sqrt{1-\rho^2}\,\mathbf{n}_i$, where $\mathbf{n}$ is a random vector and ρ is the rate of contextual drift. The drift equation, Eq. (3) with $\mathbf{t}^{IN}$ given by the last expression, is an implementation of randomly drifting context, analogous to that of Estes (1955) and Mensink and Raaijmakers (1988). This random context model is summarized in Fig. 3.

Random context models have an important property: the overlap between the state of context at two different times, i and j, is a decreasing function of the time between i and j. We term this property the *proximity relationship*.

For the Murdock (1997) formulation of random context, we can derive the proximity relationship as follows. For any time $i > j$, $\mathbf{t}_i$ can be written as

$$\mathbf{t}_i = \rho^{i-j}\mathbf{t}_j + \sqrt{1-\rho^2} \sum_{k=j+1}^{i} \rho^{i-k}\mathbf{n}_k.$$

If the random vectors, $\mathbf{n}_i$, are independent and of unit length, $E[\|\mathbf{n}_i\|] = 1$, and the elements are symmetric with zero mean and finite variance, then $E[\mathbf{n}_i \cdot \mathbf{n}_j] = 0$, $i \neq j$, and

$$E[\mathbf{t}_i \cdot \mathbf{t}_j] = \rho^{|i-j|}.$$

This means that the overlap between the state of context at two different times, i and j, is a decreasing function of the time between i and j. The variance of $\mathbf{t}_i \cdot \mathbf{t}_j$ approaches zero as the dimensionality of $\mathbf{t}$ increases without bound (because of the finiteness of the element variance and the normality of the $\mathbf{n}$ vectors), and we obtain what we will refer to as the proximity relationship:

$$\mathbf{t}_i \cdot \mathbf{t}_j = \rho^{|i-j|}. \tag{4}$$

By accumulating noise, $\mathbf{t}$ functions as a temporally sensitive construct, with the value of ρ determining the rate of contextual drift.

[3] This is essentially the architecture used in the simple recurrent network of Elman (1990) and applied extensively to implicit serial learning tasks (Cleeremans, 1993; Cleeremans & McCleeland, 1991). The difference is in the specification of the weight matrices between layers. The simple recurrent networks include weight matrices modifiable by back-propagation between $\mathbf{t}^{IN}$ and $\mathbf{t}_i$, between $\mathbf{t}_{i-1}$ and $\mathbf{t}_i$, and between $\mathbf{t}_i$ and an output layer. We have not yet discussed the connections between layers, but will shortly describe Hebbian weight matrices between $\mathbf{t}_i$ and both input and output layers. In contrast to these models, we will not consider modifiable connections between $\mathbf{t}_{i-1}$ and $\mathbf{t}_i$.

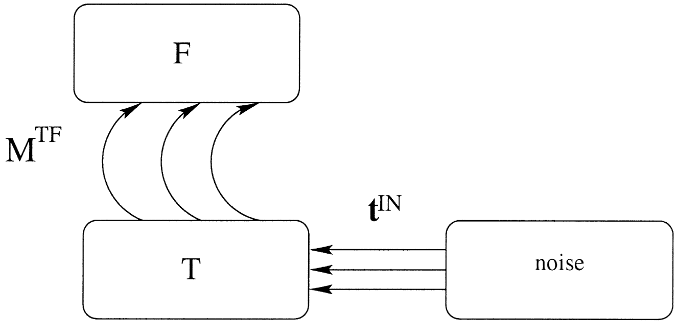

FIG. 3. Random context model. The box labeled T takes its input, $\mathbf{t}^{IN}$, from a source of random noise, and combines it with the previous state of $\mathbf{t}$ according to the drift equation (see Fig. 2). Because the noise at one time is uncorrelated with the noise at another, $\mathbf{t}$ functions as a temporally sensitive construct obeying the proximity relationship, Eq. (4). The matrix $\mathbf{M}^{TF}$ (Eq. 1) allows the state of context at a given time to be associated with the item on F presented at that time.

By substituting the proximity relationship into the definition of a_i (Eq. 2), we can easily calculate the activation for an item in semantic memory, $\mathbf{f}_i$, induced by a contextual state $\mathbf{t}_j$:

$$a_i = \rho^{|i-j|}.$$

Notice that this expression is symmetric with respect to interchange of the indices i and j.

In this random context model, cueing with $\mathbf{t}$ will differentially activate recent memories. This provides a natural account of the recency effect. By adding a competitive retrieval mechanism, this model can also explain the scale-invariance of recency (the ratio of activation for recent and remote items will remain constant).

To explain contiguity effects we need something more. We need a second associative operation linking items to context and thus allowing for retrieval of previously experienced contextual states. This idea of retrieved context serves as the basis for the temporal context model (TCM) described in the next section.

3. THE TEMPORAL CONTEXT MODEL

TCM characterizes the processes of contextual coding, storage, and retrieval. In doing so, it seeks to explain recency and contiguity effects across time scales, and it offers a mechanistic account for the process of contextual drift.

In TCM, retrieved context, rather than random noise, drives contextual drift. When a randomly chosen word list is assembled, the context retrieved by each item is uncorrelated and the model mimics randomly drifting context models. When a word from the list is recalled, it retrieves not only its prior (previously uncorrelated) context, but also the context from the current list.

To retrieve context, we introduce an item-to-context associative matrix. This outer product matrix, $\mathbf{M}^{FT}$, connects the F layer to the T layer, as shown in Fig. 4. In TCM, the input to the evolution equation, Eq. (6), is obtained by presenting the current stimulus item to the matrix $\mathbf{M}^{FT}$:

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

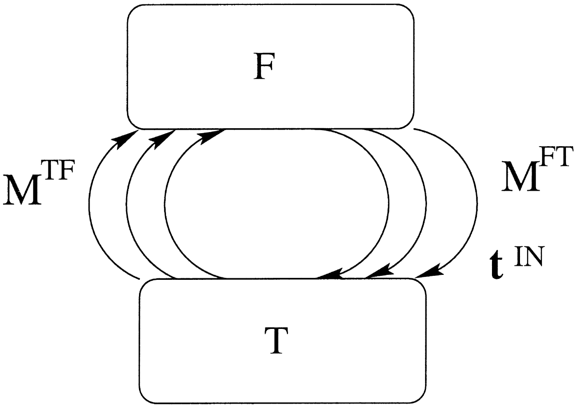

FIG. 4. Temporal context model. Rather than integrating random noise, TCM continuously retrieves context associated with its inputs. T operates by a process similar to that shown in Fig. 3, but now the inputs $\mathbf{t}^{IN}$ are a superposition of prior contextual states retrieved by the item presented on F via the matrix $\mathbf{M}^{FT}$. This results in a tight coupling between the actual sequence of inputs and the state of temporal context.

The matrix $\mathbf{M}^{FT}$ will contain a number of terms of the form $\mathbf{t}_j\mathbf{f}'_j$. When $\mathbf{M}^{FT}$ operates on an item $\mathbf{f}^v$, the result will be a superposition of contextual states $\mathbf{t}_j$ such that $\mathbf{f}_j = \mathbf{f}^v$.

In TCM, we rewrite the process for $\mathbf{t}$ as

$$\mathbf{t}_i = \rho_i \mathbf{t}_{i-1} + \beta \mathbf{t}_i^{IN}. \tag{6}$$

We refer to this as the *evolution equation*. Here β is a free parameter and $0 < \rho_i \leqslant 1$ is chosen to satisfy the constraint that $\|\mathbf{t}_i\| = 1$.[4] Note that if there is no input, $\mathbf{t}_i^{IN} = \mathbf{0}$ and $\mathbf{t}$ does not change. To solve for ρ_i, we note that

$$\begin{aligned}\|\mathbf{t}_i\|^2 &= (\rho_i \mathbf{t}_{i-1} + \beta \mathbf{t}_i^{IN}) \bullet (\rho_i \mathbf{t}_{i-1} + \beta \mathbf{t}_i^{IN}) \\ &= \rho_i^2 + 2\beta\rho_i (\mathbf{t}_{i-1} \cdot \mathbf{t}_i^{IN}) + \beta^2.\end{aligned}$$

To satisfy the constraint that $\|\mathbf{t}\| = 1$, we just have to solve a quadratic equation for ρ_i. The solution of this quadratic equation constitutes the general form for ρ_i:

$$\rho_i = \sqrt{1 + \beta^2[(\mathbf{t}_{i-1} \cdot \mathbf{t}_i^{IN})^2 - 1]} - \beta(\mathbf{t}_{i-1} \cdot \mathbf{t}_i^{IN}). \tag{7}$$

If item i has not been presented for a long time, then $\mathbf{t}_{i-1} \cdot \mathbf{t}_i^{IN} \simeq 0$ and $\rho_i = \rho \equiv \sqrt{1 - \beta^2}$, the asymptotic value for ρ_i.

3.1. Asymmetric Association

In TCM, preexperimental context is retrieved during study. This context, in turn, drives the evolution equation. At test, we assume that both the preexperimental

[4] This constraint cannot be satisfied if $\|\mathbf{t}_i^{IN}\|$ is sufficiently large. In this case, ρ_i will be zero. In this paper, $\|\mathbf{t}^{IN}\|$ is always one.

context and the studied context are retrieved, and that together they drive the evolution equation. As we show in this section, this assumption enables TCM to account for the asymmetric retrieval observed in both free and serial recall (Howard & Kahana, 1999; Kahana, 1996; Kahana & Caplan, in press; Raskin & Cook, 1937).

During study, TCM retrieves preexperimental context. When an item is repeated, which context do we retrieve, the preexperimental context or the study context? At one extreme, we could simply retrieve the preexperimental context again. In this case, the contextual representation would show no learning: the context retrieved by an item would presumably never change. At the other extreme, we could simply retrieve the most recent contextual state associated with an item. In this case, the contextual representation would have no memory: the long-term statistics of item co-occurrence would be lost from $\mathbf{t}^{IN}$.

The relative contribution of preexperimental and newly learned study context needs to be determined before we can write down a learning rule for the item-to-context matrix. Fortunately, in addition to the rather esoteric concerns listed above, there are strong constraints provided by the data.

3.1.1. *The Effect of* $\mathbf{t}_i^{IN}$ *and* $\mathbf{t}_i$ *on the CRP*

If we present some item to the network at time step i, and repeat it later at time step r, then $\mathbf{t}_r^{IN}$ can be expressed in terms of the preexperimental context $\mathbf{t}_i^{IN}$ and the newly learned $\mathbf{t}_i$, the context evoked during presentation of the list at time step i (see Fig. 5). The choice of the relative strength of preexperimental context, $\mathbf{t}_i^{IN}$, and newly learned experimental context, $\mathbf{t}_i$, has important implications. The similarity of the newly learned context to nearby contextual states $\mathbf{t}_j$ is symmetric around i in the absence of repetitions (this is the proximity relationship, Eq. 4). In contrast, the preexperimental context, $\mathbf{t}_i^{IN}$, only participates in subsequent contextual states. That

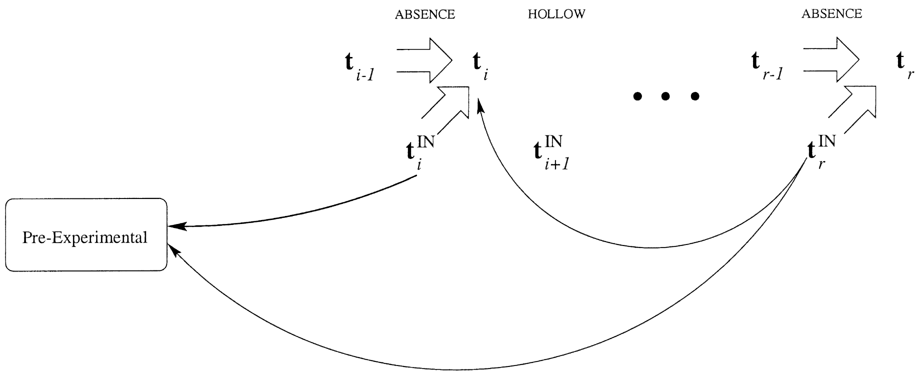

FIG. 5. Let us suppose that an item, ABSENCE in this case, is presented at time i and then repeated at time r. On the first presentation of the item, the retrieved context $\mathbf{t}_i^{IN}$ is from preexperimental experience with ABSENCE. When ABSENCE is repeated at time r, the retrieved context $\mathbf{t}_r^{IN}$ includes both the preexperimental context, $\mathbf{t}_i^{IN}$, and the newly learned context, $\mathbf{t}_i$. Wide arrows indicate the operation of the path-integrator. Curved lines indicate the functioning of retrieved context.

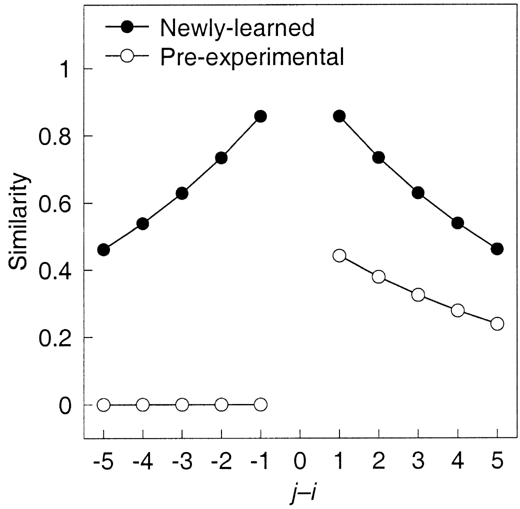

FIG. 6. In TCM, context retrieved by repetition of an item serves as an asymmetric retrieval cue for neighboring contextual states. Let us suppose that an item is presented at time i and then repeated at time r. The context retrieved at time r, $\mathbf{t}_r^{IN}$, will be a combination of the newly learned context $\mathbf{t}_i$ and the preexperimental context $\mathbf{t}_i^{IN}$. The solid circles plot $\mathbf{t}_i \cdot \mathbf{t}_j$ as a function of the interitem lag, $j-i$. The open circles plot $\mathbf{t}_i^{IN} \cdot \mathbf{t}_j$. As is clear from the figure, newly learned context provides a symmetric retrieval cue, whereas pre-experimental context only helps forward recalls.

is, $\mathbf{t}_{i+1}$ includes a $\mathbf{t}_i^{IN}$ term, but $\mathbf{t}_{i-1}$ does not. As a result, the similarity of $\mathbf{t}_i^{IN}$ to nearby states $\mathbf{t}_j$ is asymmetric. Assuming a list without repetitions:

$$\mathbf{t}_i^{IN} \cdot \mathbf{t}_j = \begin{cases} \beta(1-\beta)^{(j-i)/2} & j \geqslant i \\ 0 & j < i. \end{cases}$$

These two contributions to retrieved context are plotted in Fig. 6.

3.1.2. Specifying the Relative Contribution of $\mathbf{t}_i^{IN}$ *and* $\mathbf{t}_i$

Contiguity effects are asymmetric (Howard & Kahana, 1999; Kahana, 1996), favoring forward recalls over backward recalls, as does $\mathbf{t}_i^{IN}$. However, contiguity effects also show an advantage for nearby versus remote recalls in the backward direction, whereas $\mathbf{t}_i^{IN}$ provides no basis whatsoever for recalling items that preceded i. This advantage for nearby versus remote backward recalls must be supported by $\mathbf{t}_i$. The contribution of both of these associative components can be varied by allowing random noise to enter into the process. To gain control over the relative strength of $\mathbf{t}_i^{IN}$ and $\mathbf{t}_i$, we can write

$$\mathbf{t}_r^{IN} = A_i \mathbf{t}_i^{IN} + B_i \mathbf{t}_i. \tag{8}$$

We choose values of A and B so that the length of the input at time r, $\|\mathbf{t}_r^{IN}\|$, is unity. We introduce a free parameter, $\gamma = A_i / B_i$, to determine A_i and B_i at each presentation. Combined with the constraint that $\|\mathbf{t}_r^{IN}\| = 1$, we can solve for A_i and B_i. To solve for A_i and B_i, we note that

$$\begin{aligned} \|\mathbf{t}_r^{IN}\| &= 1 \\ &= A_i^2 + 2A_i B_i \mathbf{t}_i^{IN} \cdot \mathbf{t}_i + B_i^2. \end{aligned}$$

Substituting $A = \gamma B$ into this expression, we find

$$B_i = \frac{1}{\gamma^2 + 2\gamma(\mathbf{t}^{IN} \cdot \mathbf{t}_i) + 1}. \tag{9}$$

Everything on the right-hand side (RHS) of Eq. (9) is known. Using the definition of γ, we can trivially calculate A_i. This calculation does not require any knowledge about the delay $r-i$, but simply the relationship between $\mathbf{t}_i$ and $\mathbf{t}_i^{IN}$. We can now rewrite Eq. (8) explicitly as

$$\mathbf{t}_r^{IN} = \frac{\gamma \mathbf{t}_i^{IN} + \mathbf{t}_i}{\gamma^2 + 2\gamma(\mathbf{t}^{IN} \cdot \mathbf{t}_i) + 1}. \tag{10}$$

3.1.3. A Learning Rule for Contextual Retrieval

The goal of this section is to derive a learning rule for item-to-context associations, $\mathbf{M}^{FT}$, that will implement the desired relationship between the magnitudes of preexperimental and newly learned associations, $\mathbf{t}_i^{IN}$ and $\mathbf{t}_i$, in the retrieved context $\mathbf{t}_r$ at long delays.

Consider a particular item, $\mathbf{f}^v$, presented at time i and then repeated at some later time r ($\mathbf{f}_i = \mathbf{f}_r = \mathbf{f}^v$). Equation (8) states that the context it retrieves changes from presentation i to presentation r. That is, $\mathbf{t}_r^{IN}$ is not equal to $\mathbf{t}_i^{IN}$, but has changed as a result of the inclusion of $\mathbf{t}_i$. Because, by definition, $\mathbf{t}_r^{IN} = \mathbf{M}_r^{FT}\mathbf{f}_r$, and by assumption $\mathbf{f}_r = \mathbf{f}_i$, the change from $\mathbf{t}_i^{IN}$ to $\mathbf{t}_r^{IN}$ implies some change in $\mathbf{M}^{FT}$. Here we make explicit the changes in $\mathbf{M}^{FT}$ that will enable us to use it to implement Eq. (8).

Equation (8) states that

$$\mathbf{t}_r^{IN} = \mathbf{M}_r^{FT}\mathbf{f}_i = A_i \mathbf{t}_i^{IN} + B_i \mathbf{t}_i.$$

Now, $\mathbf{t}_i^{IN} = \mathbf{M}_i^{FT}\mathbf{f}_i$. Substituting this into the equality above gives us

$$\mathbf{M}_r^{FT}\mathbf{f}_i = A_i \mathbf{M}_i^{FT}\mathbf{f}_i + B_i \mathbf{t}_i. \tag{11}$$

We can see from this expression how $\mathbf{M}_r^{FT}$ must differ from $\mathbf{M}_i^{FT}$ for this to be true. The $B_i\mathbf{t}_i$ term on the RHS tells us that $\mathbf{M}_r^{FT}$ must contain a $B_i\mathbf{t}_i\mathbf{f}_i'$ term that $\mathbf{M}_i^{FT}$ does not. The first term on the RHS tells us that terms in $\mathbf{M}_r^{FT}$ contain all of the $\mathbf{t}_j\mathbf{f}_i'$ terms that $\mathbf{M}_i^{FT}$ did, but they have decayed by a factor of A_i.

To implement Eq. (11) for all i, and for this expression to not depend on the lag $r-i$, we must have a way to affect $\mathbf{M}^{FT}\mathbf{f}_i$ while leaving $\mathbf{M}^{FT}\mathbf{f}_j$ unaffected. To do this let us define a decomposition of a matrix $\mathbf{M}$ into two matrices parallel and complementary to some vector $\mathbf{v}$:

$$\mathbf{M} = \mathbf{M}\mathbf{P}_v + \mathbf{M}\tilde{\mathbf{P}}_v.$$

We define $\mathbf{P}_v$, the projection operator with respect to $\mathbf{v}$, as

$$\mathbf{P}_v \equiv \frac{\mathbf{v}\mathbf{v}'}{\|\mathbf{v}\|^2},$$

where the prime again denotes the transpose. Let us define the projection operator complement with respect to a vector $\mathbf{v}$ as

$$\tilde{\mathbf{P}}_{\mathbf{v}} \equiv \mathbf{1} - \mathbf{P}_{\mathbf{v}}.$$

Notice that $\mathbf{M}\mathbf{P}_{\mathbf{v}}\mathbf{v} = \mathbf{M}\mathbf{v}$, whereas $\mathbf{M}\tilde{\mathbf{P}}_{\mathbf{v}}\mathbf{v} = \mathbf{0}$.

Using this notation, we can now write down an equation that will implement Eq. (8) as a learning rule for $\mathbf{M}^{FT}$.

$$\mathbf{M}^{FT}_{i+1} = \mathbf{M}^{FT}_i \tilde{\mathbf{P}}_{\mathbf{f}_i} + A_i \mathbf{M}^{FT}_i \mathbf{P}_{\mathbf{f}_i} + B_i \mathbf{t}_i \mathbf{f}'_i, \tag{12}$$

where A_i and B_i can be easily calculated as described in the previous section. The first two terms on the RHS of Eq. (12) refer to the persistence of preexisting connections between F and T. The last two terms indicate that there is the storage of a new Hebbian association from $\mathbf{f}_i$ to $\mathbf{t}_i$.

There is nothing about Eq. (12) that would make it difficult to implement in a neural network. We can perhaps gain some intuition into this by considering what would happen if the $\mathbf{f}$s and $\mathbf{t}$s were sparse binary vectors.[5] $\mathbf{M}^{FT}_i \tilde{\mathbf{P}}_{\mathbf{f}_i}$ refers to the set of entries of $\mathbf{M}^{FT}_i$ which connect to elements on F which are zero in $\mathbf{f}_i$. These elements are unaffected. $\mathbf{M}^{FT}_i \mathbf{P}_{\mathbf{f}_i}$ refers to the set of entries in $\mathbf{M}^{FT}_i$ that connect to elements on F which are nonzero in $\mathbf{f}_i$. These entries decay by a factor of A_i, which is necessarily less than one. If they connect a nonzero element of $\mathbf{f}_i$ to a nonzero element of $\mathbf{t}_i$, they are also incremented by a factor of B_i.

3.2. Summary of TCM

Table 1 summarizes all of the components of TCM. We begin by distinguishing item representations and contextual representations. Item representations are vectors $\mathbf{f}$ on a space F. Contextual representations are vectors $\mathbf{t}$ on a space T.

TABLE 1

Summary of the Temporal Context Model

Name	Expression	Eq. number
The Context-to-Item Matrix[a]	$\mathbf{M}^{TF} = \sum_i \mathbf{f}_i \mathbf{t}'_i$	(1)
The TCM Input Equation	$\mathbf{t}^{IN}_i = \mathbf{M}^{FT}_i \mathbf{f}_i$	(5)
The TCM Evolution Equation[b]	$\mathbf{t}_i = \rho_i \mathbf{t}_{i-1} + \beta \mathbf{t}^{IN}_i$	(6)
The Item-to-Context Matrix[c]	$\mathbf{M}^{FT}_{i+1} = \mathbf{M}^{FT}_i \tilde{\mathbf{P}}_{\mathbf{f}_i} + A_i \mathbf{M}^{FT}_i \mathbf{P}_{\mathbf{f}_i} + B_i \mathbf{t}_i \mathbf{f}'_i$	(12)

[a] For simplicity, $\mathbf{M}^{TF}$ is assumed to be reset at the beginning of each experiment.
[b] ρ_i is given by Eq. 7.
[c] B_i is given by Eq. (9). A_i is defined as γB_i, where γ is a free parameter. The operators $\mathbf{P}_{\mathbf{f}_i}$ and $\tilde{\mathbf{P}}_{\mathbf{f}_i}$ are defined in the text.

[5] This assumption violates prior assumptions about the orthonormality of the $\mathbf{f}$s as well as the evolution Eq. (6). This example is purely hypothetical.

Context, $\mathbf{t}$, functions as a cue for recall of an item by means of an outer-product matrix, $\mathbf{M}^{TF}$, associating the states on F with states on T. The TCM input equation, (5), specifies that the input to T, $\mathbf{t}_i^{IN}$, is caused by the presentation of the current item, $\mathbf{f}_i$, to an associative matrix, $\mathbf{M}_i^{FT}$, associating states on F to states on T. The evolution equation, Eq. (6), describes the evolution of temporal context $\mathbf{t}_i$ as a function of the previous state of temporal context $\mathbf{t}_{i-1}$ and the context retrieved by item $\mathbf{f}_i$, $\mathbf{t}_i^{IN}$. In the evolution equation, β is a free parameter determining the magnitude of a retrieved context's contribution to $\mathbf{t}_i$, whereas ρ_i is chosen such that $\|\mathbf{t}_i\| = 1$. The item-to-context matrix, $\mathbf{M}_{i+1}^{FT}$, is updated by decaying the part of the matrix parallel to $\mathbf{f}_i$ by a factor A_i and storing an outer product term between $\mathbf{t}_i$ and $\mathbf{f}_i$ weighted by a factor B_i (Eq. 12). These coefficients are chosen such that $\|\mathbf{t}_r^{IN}\| = 1$ for a repeated presentation of $\mathbf{f}_i$. Their relative strength is controlled by a free parameter $\gamma = A_i / B_i$.

Consider the sequence of events that are postulated to occur when an item $\mathbf{f}_i$ is presented. First $\mathbf{f}_i$ retrieves context, $\mathbf{t}_i^{IN}$. Then, this retrieved context is used to update $\mathbf{t}_{i-1}$, resulting in the new state $\mathbf{t}_i$. Then, the matrices $\mathbf{M}_i^{FT}$ and $\mathbf{M}_i^{TF}$ are updated. These three stages, contextual retrieval, contextual integration, and updating the connections, must take place in this precise order after each item presentation for the model to work. This process presumably takes place on the scale of tenths of seconds.

In TCM, during study of a randomly assembled list of words, $\mathbf{t}^{IN}$ is preexperimental context retrieved by the words in the list. When an item is repeated, by virtue of a repetition in the list, or by its successful recall, it again retrieves preexperimental context, but in addition, retrieves the newly learned experimental context in which it was presented. That is, an item initially presented at serial position i and then later at serial position r retrieves $\mathbf{t}_r^{IN}$, a weighted sum of $\mathbf{t}_i^{IN}$ and $\mathbf{t}_i$. Because the preexperimental context, $\mathbf{t}_i^{IN}$, is only similar to subsequent contextual states, $\mathbf{t}_r^{IN}$ is more effective as a cue for items that followed the initial presentation than for items that preceded the initial presentation. Because the list context $\mathbf{t}_i$ is a good cue for items that were presented near serial position i in the list, $\mathbf{t}_r^{IN}$ is able to support a backward association for serial positions preceding i.

4. APPLICATION OF TCM TO RECALL DATA

We adopt a relatively simple, nonlinear competitive rule for mapping a set of activations, a_i, onto probability of recall. This section begins by describing this rule and then employs this rule to evaluate TCM's ability to fit a rich set of data on retrieval dynamics in free recall. In particular, we assess TCM's account of recency across time scales by examining immediate, delayed, and continuous-distractor free recall. We also assess TCM's account of contiguity across time scales by examining delayed and continuous-distractor free recall.[6]

[6] In immediate free recall, the CRP is not stable across output positions (Howard & Kahana, 1999; Kahana, 1996), making it a less-than-ideal environment for asking basic questions about contiguity effects. On the other hand, the fact that the CRP changes shape in immediate, but neither delayed nor continuous-distractor free recall is an interesting phenomenon in its own right. The implications of this phenomenon are discussed in the general discussion.

Our rule for mapping the activations onto probability of recall is a variant on several well-known models of stimulus classification and category learning (Nosofsky, 1986, 1987; Shepard, 1987). Given a contextual cue, a specific state $\mathbf{t}$, many items in semantic memory, $\mathbf{f}_i$, are activated, each by some amount a_i. Given some contextual cue $\mathbf{t}$, the vector $\mathbf{f}^{IN}$ is defined as $\mathbf{f}^{IN} \equiv \mathbf{M}^{TF}\mathbf{t}$ and the activation of item $\mathbf{f}_i$ is defined as $a_i \equiv \mathbf{f}_i \cdot \mathbf{f}^{IN}$. This input to F is ambiguous—it is more or less similar to many items but identical to none. A reasonable assumption is that subjects are more likely to respond with items $\mathbf{f}_i$ that are close to $\mathbf{f}^{IN}$. Using the (Euclidean) distance from $\mathbf{f}^{IN}$ to a given item $\mathbf{f}_i$ we can write

$$P(\mathbf{f}_i \mid \mathbf{f}^{IN}) = \frac{\exp\left(-\frac{1}{\tau}\|\mathbf{f}^{IN} - \mathbf{f}_i\|^2\right)}{\sum_j \exp\left(-\frac{1}{\tau}\|\mathbf{f}^{IN} - \mathbf{f}_j\|^2\right)}. \tag{13}$$

This equation introduces a new free parameter τ, which controls the sensitivity of P_i to differences in the distances. Assuming, as we have all along, that the item representations $\mathbf{f}_i$ are orthonormal, Eq. (13) simplifies to

$$P(\mathbf{f}_i \mid \mathbf{f}^{IN}) = \frac{\exp\left(\frac{2a_i}{\tau}\right)}{\sum_j \exp\left(\frac{2a_j}{\tau}\right)}. \tag{14}$$

A key question involves over what range to extend the sum in Eq. (14). On the one hand, subjects tend to recall items from the current list predominantly. We could restrict the sum to the current list by assuming that $\mathbf{M}^{TF}$ is cleared at the beginning of each list. This is a common assumption in distributed memory models (Murdock & Kahana, 1993). Alternatively, we could assume that subjects use a fixed list context as a retrieval cue (e.g., Raaijmakers & Shiffrin, 1980), where the contexts for different lists are orthogonal. Both of these approaches are poor approximations, at best, to what subjects really do. Subjects sometimes recall items from prior lists, as intrusions, and tend to do so more from recent lists than from distant ones. Subjects can also directly recall items from prior lists (Shiffrin, 1970a). In this paper we allow the sum in Eq. (14) to extend only over the current list, while acknowledging the limitations of this approach.

4.1. Free Recall

The serial position curve in free recall is largely characterized by two processes. The probability of first recall measures where subjects begin recall—a serial position curve for the first item subjects produce. The probability of first recall provides an index of the end-of-list recency effect. After the first recall, subjects tend to make transitions to nearby serial positions. This tendency, called the lag-recency effect, is measured with the CRP as a function of lag (the difference in serial positions of the word recalled at output position $i+1$ and output position i).

We hypothesize that in free recall subjects use the current state of $\mathbf{t}$ to probe memory and that $\mathbf{t}$ evolves according to TCM. End-of-list recency (as measured by the probability of first recall) is a result of cueing with the state of context at the time of test. The lag-recency effect (as measured by the CRP) is a consequence of context retrieved by recalled items. We evaluate this hypothesis by applying the model to data on the end-of-list and lag-recency effects in free recall. Although they were fit simultaneously, these two sets of results will be presented separately in the following sections. Before starting on the end-of-list recency effect, we note some considerations common to all the modeling in this paper.

4.1.1. *Summary of Model Parameters*

TCM has two free parameters: β, which controls the proportion of retrieved context used in updating the current state of context, and γ, which controls the degree of asymmetry seen in the CRP. For the purposes of the current simulations, γ was fixed at 1, implying an equal weighting of newly learned and preexperimental context. The mapping onto probability of recall contributes an additional free parameter, τ. In addition, we allowed d, the effective length of the distractor intervals in delayed and continuous-distractor free recall, to vary freely.

The freedom to make d different from the actual amount of clock time constituting the delay interval is a consequence of the causality of TCM and the constraint that $\mathbf{t}_i$ remain normalized. That is, because ρ_i depends on the similarity of $\mathbf{t}_i^{IN}$ to $\mathbf{t}_{i-1}$, the rate of change of context is a causal consequence of the items presented.[7] We assume that one of the properties of arithmetic distractor problems is that they are relatively similar to each other. As a consequence, they will retrieve similar contextual states and cause less drift than an equivalent duration of list items.

The other variables, ρ_i, A_i, and B_i, are neither free parameters nor fixed parameters. According to TCM, their value is determined at each item presentation to ensure satisfaction of normalization constraints.

4.1.2. *Experimental Data*

To assess TCM's ability to describe recency effects in free recall across time scales, we fit TCM to data from the immediate condition of Experiment 1 and the delayed and continuous-distractor conditions of Experiment 2 of Howard and Kahana (1999). The end-of-list recency effect was measured using the probability of first recall. The lag-recency effect was measured using the CRP as a function of lag at the first output position.

[7] This property also predicts that recency should not be affected by an unfilled retention interval. When no item is presented, $\mathbf{f}_i = \mathbf{0}$, then $\mathbf{t}_i^{IN} = \mathbf{0}$ and $\rho_i = 1$. This prediction is consistent with findings from both free recall and short-term paired associate learning (Baddeley & Hitch, 1977; Murdock, 1963). Furthermore, the amount of information given during a delay period, rather than the time *per se*, should be the key factor affecting the recency effect, as was found by Glanzer, Gianutsos, and Dubin (1969).

Howard and Kahana (1999) used a semantic orienting task in an effort to minimize rehearsal in two free recall experiments.[8] In both experiments, subjects studied lists of 12 items, randomly sampled from the Toronto Noun Pool (Friendly, Franklin, Hoffman, & Rubin, 1982), presented at a fast rate (1 s in Experiment 1, 1.2 s in Experiment 2). The orienting task was a judgment of concreteness to which subjects responded with a button press. Sixteen seconds of self-paced true–false arithmetic distractor intervened between the study of the last item and the free recall test in all the conditions of Experiment 2. In the continuous-distractor data for the fits here, an additional 16 s of arithmetic distractor intervened between presentation of each list item. Subjects were paid a bonus based on their performance on the orienting and distractor tasks.

4.1.3. Recency: The End-of-List Recency Effect in Free Recall

To make model predictions for the probability of first recall we calculate a set of activations for the list items given the state of context at the time of test. The long-term recency effect is predicted because of the joint use of a temporally sensitive mechanism that is sensitive to changes over a long time scale and a competitive rule for mapping activations onto probability of recall.

Modeling. We used the simplex method (Nelder & Mead, 1965) minimizing a variance-adjusted root-mean-squared deviation to generate parameter values. We allowed β and τ to vary. The parameter γ, which controls the magnitude of the asymmetry in the CRP, was set to one, implying an equal contribution of $\mathbf{t}_i^{IN}$ and $\mathbf{t}_i$ in Eq. (8). The effective length of the distractor intervals, d, in delayed and continuous-distractor free recall was allowed to vary.

We used 3000 simulated trials for each set of parameter values. Model predictions were derived by running the experimental analysis software on the simulated data.

Results and discussion. Figure 7 shows that TCM nicely captures the basic pattern of end-of-list recency across time scales. The best-fitting value of β was 0.402. The best-fitting value of τ was 0.247. The best-fitting value of d was 7.24. Because the model has no mechanism to generate primacy, it fails to capture the small one-position primacy effect in the data. For the immediate condition, $\chi^2(8) = 26.9$. A large portion (8.4) of this deviation was accounted for by the first serial position. For the delayed condition, $\chi^2(8) = 57.7$. Again, the most deviant point (18.3) was the first serial position. For the continuous-distractor condition, $\chi^2(8) = 37.3$.

Although the fits were not numerically spectacular, they were acceptable given the simplicity of the model. The prediction of the long-term recency effect is a zero parameter prediction of the model. That is, for every choice of parameters $0 < \beta < 1$, $0 < \tau < \infty$, there is a greater recency effect for continuous-distractor compared to delayed free recall.

[8] The raw data from these (and other) experiments can be downloaded from http://fechner.ccs.brandeis.edu/Experiments/archive.html.

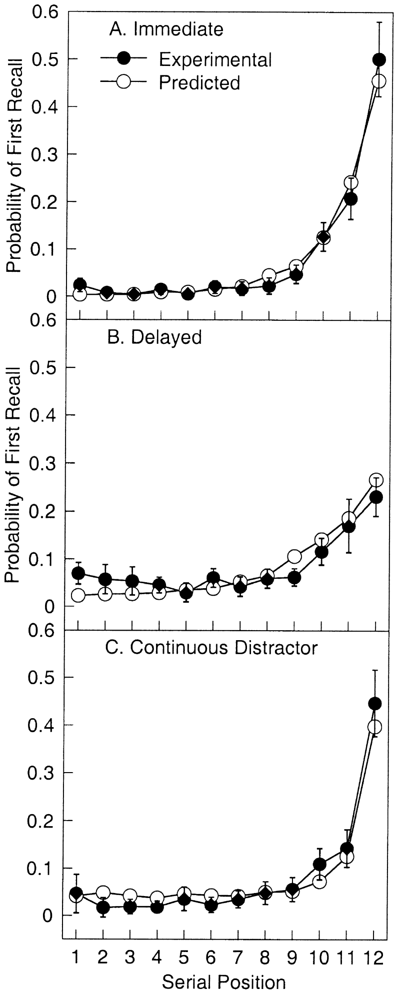

FIG. 7. The end-of-list recency effect in free recall. This figure shows TCM's fit to the probability of first recall in immediate, delayed, and continuous-distractor free recall. The data are from Howard and Kahana (1999). Because **t** is a temporally sensitive construct, and the retrieval rule is competitive and nonlinear, TCM predicts the pattern of recency effects across conditions. Throughout this paper, error bars on the data are 95% confidence intervals calculated with the method of Loftus and Masson (1994) for within-subjects designs.

4.1.4. Contiguity: The Lag-Recency Effect

TCM predicts the lag-recency effect because contextual drift is driven by contextual retrieval; *i.e.*, because $\mathbf{t}_i^{IN}$ is a causal consequence of the presentation of item *i*, repetition of that item at a later time causes some reinstatement of the context present at time *i*, thus mediating an associative effect. Because this retrieved component is similar to the context associated with items from nearby list positions, a lag-recency effect is predicted. The relative strength of the activation of nearby items will be approximately equal in delayed and continuous-distractor free recall

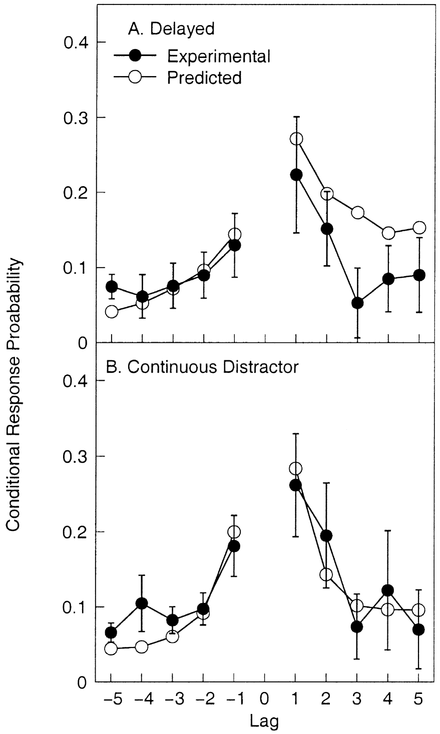

FIG. 8. The long-term lag-recency effect. Experimental points are the CRP from the first output position from the delayed and longest-IPI conditions of Howard and Kahana (1999). The model uses the same parameter values used to generate Fig. 7. The error bars on the data are 95% confidence intervals calculated with the Loftus and Masson (1994) method for within-subjects variables.

because the relative spacing of the items within the list is the same. The mapping onto probability of recall supplies a competitive retrieval rule, so the long-term lag-recency effect is also predicted.

Modeling. We fit TCM to the CRP from the first output position of the delayed and the longest continuous-distractor condition of Experiment 2 of Howard and Kahana (1999). These fits were done simultaneously with those shown in Fig. 7.

Results and discussion. The results are shown in Fig. 8. The fits capture the existence of a lag-recency effect and the asymmetry in delayed and continuous-distractor free recall. The forward direction of the CRP in the delayed condition was the least well fit by the model, $\chi^2(5) = 103$. The forward direction of the CRP for continuous-distractor free recall was best fit by the model $\chi^2(5) = 5.46$. Although not numerically spectacular, the fits illustrate that TCM simultaneously predicts asymmetry in the lag-recency effect, the persistence of the lag-recency effect with inclusion of an interitem delay, and the persistence of the asymmetry with the inclusion of the delay.

5. GENERAL DISCUSSION

At one level, verbal theories of episodic memory performance have defined episodic memory as the type of memory for an item appearing in a specific context, with a sometimes vague and flexible definition of context. More quantitative models of episodic memory performance have relied heavily on associations between items and list context (Chappell & Humphreys, 1994; Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980). To introduce long-term forgetting, several models (Mensink & Raaijmakers, 1988; Shiffrin & Steyvers, 1997) have used randomly varying context. Despite its importance in episodic memory, the definition and mechanisms of context have rarely been studied. This paper introduces a well-specified and falsifiable model of temporal context. This model describes recency and contiguity effects across time scales in a concise and elegant way and provides a completely novel explanation for the associative asymmetry so commonly observed in free and serial recall.

5.1. TCM: Model Summary

TCM consists of five basic components:

1. A space F corresponding to semantic memory. Specific states on F, $\mathbf{f}_i$ correspond to item representations.
2. A space T. Specific states on T, $\mathbf{t}_i$ correspond to specific states of context.
3. A matrix $\mathbf{M}^{TF}$ connecting T to F, enabling a state of context to serve as a cue for recall of items. This matrix stores simple Hebbian outer-product terms. We assume in this paper that $\mathbf{M}^{TF}$ is reset at the beginning of each experiment.
4. A matrix $\mathbf{M}^{FT}$ connecting F to T. This matrix enables presentation of an item $\mathbf{f}_i$ to retrieve prior states of context that were obtained when that item was presented. This matrix uses item-specific unlearning to ensure that the magnitude of the retrieved context vector does not grow without bound as the item is repeated many times.
5. The evolution equation that allows $\mathbf{t}_i$ to change gradually. In TCM, the input to the path integrator is retrieved context. When an item is repeated, it retrieves both preexperimental context, favoring forward recalls, and newly learned context, which serves as a symmetric cue.

We applied TCM, coupled with a simple rule for mapping item activations onto probability of recall, to data on recency and contiguity effects in free recall. In this application we assumed that the current state on T is always the cue for recall of an item on F. (1) Because $\mathbf{t}_i$ drifts gradually when a list of random words is presented, TCM predicts a recency effect (Fig. 7). (2) Because our mapping onto probability of recall is competitive, the long-term recency effect is predicted (Fig. 7). (3) Because repetition of an item retrieves context, and context changes gradually, a lag-recency effect is predicted.

TCM describes the asymmetry seen in the lag-recency effect without additional assumptions. Both preexperimental and newly learned contextual states contribute to the contextual cue used in TCM. Preexperimental context is incorporated into subsequent, but not previous, contextual states. The reinstatement of preexperimental context results in an advantage for subsequent items—a forward recall advantage. Newly learned experimental context is equally similar to preceding and subsequent contextual states. The reinstatement of newly learned experimental context provides an advantage for nearby versus remote recalls in the backward direction. The net result of these two factors is an asymmetric, temporally graded lag-recency effect (Fig. 8).

5.2. Alternative Accounts of Recency and Contiguity

TCM describes recency and contiguity across time scales with a single mechanism. This is not the first attempt at such a unified description of recency and contiguity. The SAM model (Raaijmakers & Shiffrin, 1980) explains immediate recency as a consequence of buffer occupancy. Short-term store is also held to be the locus for the formation of new, episodically formed associations. Positional models[9] (Brown et al., 2000; Johnson, 1991; Lee & Estes, 1977, 1981; Nairne et al., 1997; Neath & Crowder, 1990) can also be used to explain both recency and contiguity. For instance, distinctiveness models (e.g., Glenberg & Swanson, 1986; Neath & Crowder, 1990) assume that recency is a consequence of a decision based on a comparison of memory for items along a temporal dimension. Positional models can also be used to drive associative effects. Because nearby items will tend to have similar positional (or temporal) representations, search along a positional (or temporal) dimension could drive associative effects in the absence of any causal relationship between one recall and the next. This section discusses TCM *vis-a-vis* these other approaches.

5.2.1. Short- and Long-Term Stores

In memory research, short-term store (STS) has been closely tied to the empirical reality of rehearsal. There is little doubt that some concept of a STS must be retained. In some applications, e.g., perception of a compound object or simultaneous perception of multiple objects, some notion of short-term memory is absolutely indispensable. Similarly, the central point of Atkinson and Shiffrin (1968)—that voluntary control processes have a great effect on memory—is as true today as it was 30 years ago. By the same token, our analyses argue that recency and contiguity reflect basic memory processes that probably do not depend on short-term store. This of course does not preclude an influence of STS (or something very much like it) on recency or formation of associations in situations where active maintenance of information takes place.

[9] The definition of positional model used here is considerably more broad than that typically used. The models referred to here all share the property that memory search is accomplished *via* associating items to some construct with some inherent structure that is independent of the items. Whether this construct is the pegboard often used as an analogy in traditional positional models, a temporal dimension, as used in the distinctiveness models, or the state of a set of coupled oscillators, all these models can give rise to associative effects without any actual causality between one recall and another. This broad definition serves to contrast these models with traditional notions of association, as well as the approach to association offered by TCM. In this context, TCM can be seen as a hybrid of traditional and positional approaches to associative effects—associations are mediated by a construct that can have a temporal metric, but this construct, and its mediating effects, are a causal consequence of the specific items and the act of recalling them.

Many of the properties of STS are captured by the context vector $\mathbf{t}_i$. Like STS, $\mathbf{t}_i$ contains prior items—in this case, the $\mathbf{t}^{IN}_{j<i}$. Like STS, $\mathbf{t}_i$ is temporally sensitive—more recent items are more likely to remain in STS and more strongly activated by $\mathbf{t}_i$. Like STS, $\mathbf{t}_i$ changes as a result of the input of new information, rather than the passage of time *per se*. Also like STS, $\mathbf{t}_i$ exhibits sequential dependencies. The effect of a repeated item on STS depends on whether that item is currently in the buffer. This is more or less likely depending on the recency of the item (and the number of unique stimuli to be remembered in an experiment; Atkinson & Shiffrin, 1968). Similarly, the effect of a repeated item on $\mathbf{t}_i$ depends on how recently it was presented.

One could argue that a real strength of TCM is that $\mathbf{t}_i$ mimics STS over the short term. Insofar as this is true, STS is *like* a limiting case of the behavior of $\mathbf{t}_i$. The important difference has to do with the behavior of $\mathbf{t}_i$ over the long term. Rather than being replaced from a buffer in an all-or-none fashion, contextual states just contribute less and less to the contextual cue. Although the probability of an item being in STS may fall off over time exponentially over any duration one might like to consider,[10] just like $\mathbf{t}_i \cdot \mathbf{t}_j$, buffer occupancy cannot be used to generate recency effects over the long term. The reason is that on the occasions on which the item is not in STS, there is no benefit whatsoever for the item. In contrast, $\mathbf{t}_i \cdot \mathbf{t}_j$, although it may also have a small value, is reliable from trial to trial. As a result, it can be used to support recency effects over long periods of time.

5.2.2. Positional Coding: Why Not a Clock?

Another concept that has been used to explain contiguity effects is positional coding (see footnote 9). Although most frequently associated with serial recall, positional codes have also been used to explain free recall. The distinctiveness models (Nairne et al., 1997; Neath & Crowder, 1990), proposed to explain the long-term recency effect in free recall, can be seen as using a (relative) positional code. Although these models are mute as to the existence of contiguity effects in free recall, contiguity effects can be modeled with a positional code. Indeed, positional codes have been used to explain confusion gradients (Brown et al., 2000; Burgess & Hitch, 1999; Estes, 1972; Henson, 1998; Lee & Estes, 1977, 1981). Search through sequential positions could be used to mimic associative effects. The critical distinction between TCM and models of this class is the direct causality between item presentation and associative effects in TCM and the lack of direct causality between these two in positional models.

Recently, positional codes have been reformulated as a kind of replicable random context model. For instance, in the Burgess and Hitch (1992, 1999) model, contiguity effects are driven by a context-timing signal. This signal, a vector, changes slowly from time-step to time-step, much like the random context models (Estes, 1955; Mensink & Raaijmakers, 1988). The context-timing signal differs from

[10] The probability that an item remains in the buffer is exponential if incoming items are equally likely to displace any of the items in STS. The exact form depends on the rule used for displacing items from the buffer.

random context in that it is assumed to pass through the same sequence of states at test that it did during study, much like a clock that has been wound back to a starting point. As a consequence, items are recalled in approximately the sequence in which they were studied. Items recalled out of sequence are likely to be recalled near to their correct location, leading to error gradients. The OSCAR model (Brown et al., 2000) describes a context-timing signal constructed from the states of oscillators with different frequencies.

Although TCM and the clock-context models share some features and some terminology, they are very different in both conception and implementation and can be distinguished experimentally, as we will now discuss.

Retrieved context. In TCM, we make extensive use of the term "retrieved context." By this term we refer to a well-specified process of contextual retrieval mediated by an item-to-context matrix. The clock-context models use closely related terms such as "contextual reinstatement." In the clock-context models this means "resetting of the context vector to the state it occupied at the beginning of list presentation." The mechanism for such reinstatement is left unspecified. However, after the context is reset to a given state, it is assumed to evolve through the same set of states that it did during study. Although OSCAR includes an item-to-context matrix, this could not be used for contextual reinstatement, as the context vector is derived from a set of oscillators, which must be reset.

Context effects. The clock-context models are careful to emphasize that the evolution of the context-timing signal is independent of the items being presented, or recalled.[11] In contrast, in TCM context is driven by particular items retrieving context. These differing conceptions of context lead to different predictions.

TCM predicts context effects (e.g., Falkenberg, 1972), whereas a model that relies solely on a clock-signal cannot. That is, if we precede a TBR stimulus with some event, say a specific distractor task, then memory for that stimulus will be better after a delay if we repeat that distractor task.

If a clock-context signal is taken to be the sole cue for recall, then any data that imply a causal connection between one recall and another cannot be accounted for by such a model. This has been taken by many authors as a strength of these models, in that it allows them to account for experiments in which serial recall of a list that includes phonologically confusable elements does not result in increased errors for the intervening items (Henson, Norris, Page, & Baddeley, 1996). This experiment used written recall, so it is not entirely clear that the order in which the responses appear from top to bottom on the page corresponds exactly to the order in which they were actually recalled. Even granting the conclusion of that particular experiment, it remains to be seen whether the result generalizes to sources of error other than phonological similarity (e.g., semantic similarity, episodically induced associations). Given that models that assume no causal relationship between one

[11] Except of course for the assumptions that grouped presentation recruits oscillators with a specific frequency (leading to hierarchical effects) and that the beginning of a list somehow marks that contextual state as reinstatable.

recall and the next in serial recall have not been applied to results that certainly seem to suggest there is such a relationship (e.g., Chance & Kahana, 1997), the burden of proof should remain squarely on those who wish to claim that chaining models (defined as those that maintain a causal relationship between one recall and the next) are dead. Furthermore, leaving aside the issue of serial recall, there is a great deal of evidence from free recall that suggests that there is a causal relationship between one recall and the next.

5.3. TCM Is Not a Free Recall Model

TCM is a model that prescribes a set of rules for how a distributed episodic representation should change from moment to moment. The way that this representation changes enables it to describe recency effects and associative effects in a way that can persist across time scales. The availability of data on recency and contiguity across time scales in free recall varies. We used these data to constrain TCM and to test its key assumptions. We did not, however, develop a full-fledged model of free recall. This is an ambitious project that will involve tackling a number of other important aspects of memory function not yet addressed by TCM. The purpose of this section is to discuss the additional factors needed to build a complete model of free recall that uses TCM as its associative engine.

5.3.1. Immediate Recall

As shown in Fig. 7, TCM does a good job at describing the recency effect in immediate free recall as manifest in the probability of first recall. However, it fails to quantitatively describe a phenomenon observed in immediate recall at subsequent output positions: in immediate, but in neither delayed nor continuous-distractor free recall, the CRP changes shape with output position. Specifically, at early output positions, the CRP is steeply peaked in immediate free recall. As output continues, the CRP becomes less pronounced, asymptoting at a level consistent with that seen in delayed recall (Howard & Kahana, 1999; Kahana, 1996).

Kahana (1996) showed that the change in the CRP with output position is perfectly consistent with retrieval from short-term store to initiate immediate recall. Although the effect of output position on the CRP in immediate free recall does not necessarily require short-term store, we think it does requires a model of semantic retrieval significantly more complex than the simple activation-choice rule used here.

The traditional view of immediate recency would be that multiple item representations are active in semantic memory at the time of test. Because these representations all tend to come from the end of the list, there is an enhanced lag-recency effect while the active representations are recalled. If the recency effect in immediate free recall was indeed always fueled by active item representations in short-term memory, it would seriously undermine the view of immediate recency presented in this paper: that the recency effects in immediate and delayed recall share a common source. No one would argue that there are not major differences between immediate recall and delayed recall. In addition to the change of the CRP with output position, there are also big differences in the latencies observed in continuous-distractor versus immediate recall: latencies in continuous-distractor free recall are much more comparable to latencies in delayed than immediate free recall.[12]

These incontrovertable differences between immediate recency and long-term recency are not necessarily inconsistent with the idea that the recency effect in immediate and continuous-distractor free recall are a consequence of cueing with end-of-list context. Within TCM there is indeed a difference between the contextual cue in immediate and delayed recall that could be utilized to drive the effect on the CRP in immediate recall. Although the relative activation of the items in the list is comparable in immediate and continuous-distractor free recall, the absolute activation of the items in immediate recall is much higher than in delayed recall. Perhaps retrieval from semantic memory could exploit this difference in magnitude to drive the effect of output position. This would require multiple retrievals taking place in semantic memory in response to the initial (contextual) recall cue in immediate, but not delayed recall. Potentially, the difference in the absolute levels of activation could also manifest as faster recall latencies in immediate recall.

There is a rather subtle distinction between this position and the position a proponent of STS in immediate recency might take. The traditional view of immediate recency is that there are multiple active states in semantic memory at the time of test that drive the recency effect. There is little question that this is possible under at least some circumstances. We are arguing that, at least in some circumstances in which active maintenance of information is precluded, semantic memory is "empty" at the time of test, but multiple item representations are quickly activated by a strong contextual cue.

5.3.2. Avoiding Repetitions

In free recall, subjects rarely repeat words that they have already recalled (Laming, personal communication). Furthermore, what repetitions there are tend to take place after several other recalls have intervened since the initial recall (Laming, personal communication). There are two main classes of mechanisms that have been introduced to account for these data. One mechanism is response suppression. The other is a recognition or editing process that follows item selection or generation.

Response suppression. Response suppression has been popular in models of serial recall (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998; Lewandowsky & Murdock, 1989; Page & Norris, 1998). The basic idea of these models is that after being activated, items are temporarily inhibited. In the simplest variant of response suppression (Brown et al., 2000; Henson, 1998; Lewandowsky & Murdock, 1989), words that have already been recalled are simply removed from the pool of recallable words. More sophisticated versions (Burgess & Hitch, 1999; Grossberg, 1978) explicitly model a transient inhibition that they identify with neuronal fatigue.

Generate–recognize models. According to generate–recognize models, subjects generate candidate responses which are then submitted to a recognition process. If subjects generate an item which is not familiar or does not match the list context,

[12] M. W. Howard and M. J. Kahana, unpublished observation.

the response is censored. In Shiffrin's models (Raaijmakers & Shiffrin, 1980, 1981b; Shiffrin, 1970b), items that are already recalled may still be sampled and recovered. A special assumption is added to prevent these resampled items from being produced. Although the recognition decision is not explicitly modeled in SAM, this editing of recovered items places SAM squarely in the tradition of generate–recognize models (Anderson & Bower, 1972; Bahrick, 1970; Kintsch, 1970). In generate–recognize models, the more items that have been recalled, the lower the probability of sampling a new item. As a consequence the expected number of samples between successful recalls goes up with output position. Accordingly, resampling is likely a factor in the growth of interreponse times with output position (Murdock & Okada, 1970; Rohrer & Wixted, 1994).

Both generate–recognize and response–suppression models account for the absence of repetitions as an effect downstream from the results of an initial inquiry of memory. In the generate–recognize models, the omission of repetitions is the result of a conscious editing that takes place after items have been generated. In the response–suppression models, activations of output codes are modulated by an inhibition local to the level of the output codes. The choice of output mechanism used to inhibit repetition, and for that matter the choice of model for retrieval from semantic memory, is largely orthogonal to the choice of TCM. If one prefers the generate–recognize model, then TCM, including the simple mapping of activations onto probability of recall can be seen as a means for generating potential recalls. If one prefers the response–suppression approach, then TCM can be used to generate a set of activations, which can in turn be modulated by response suppression.

5.4. Conclusions

We developed a distributed model of temporal context that we call TCM, for temporal context model. Building on previous formulations of random context (Estes, 1955; Mensink & Raaijmakers, 1988; Murdock, 1997), TCM uses retrieved preexperimental contextual states to drive contextual drift. This assumption leads to an integrated account of contextual retrieval. As a consequence, repeating an item retrieves an asymmetric retrieval cue. This, in turn, provides a straightforward account of the ubiquitous asymmetry observed in memory retrieval (Howard & Kahana, 1999; Kahana, 1996; Kahana & Caplan, in press).

Because TCM uses retrieved context to drive associative effects, it does not require any direct interitem connections. Finally, we have shown that retrieved context and contextual drift, coupled with a competitive retrieval rule, enable TCM to predict the scale-invariance in both recency and contiguity effects in free recall.

REFERENCES

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, **79**, 97–123.

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, **106**, 135–163.

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, pp. 89–105). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1977). Recency reexamined. In S. Dornic, (Ed.), *Attention and performance VI* (pp. 647–667). Hillsdale, NJ: Erlbaum.

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review*, **77**, 215–222.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, **6**, 173–189.

Brodie, D. A., & Murdock, B. B. (1977). Effects of presentation time on nominal and functional serial position curves in free recall. *Journal of Verbal Learning and Verbal Behavior*, **16**, 185–200.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, **107**, 127–181.

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**, 551–581.

Chance, F. S., & Kahana, M. J. (1997). Testing the role of associative interference and compound cues in sequence memory. In J. Bower (Ed.), *Computational neuroscience, Trends in research* (pp. 599–603). New York: Plenum Press.

Chappell, M., & Humphreys, M. (1994). An autoassociative neural network for sparse representations: Analysis and application to models of recognition and cued recall. *Psychological Review*, **101**, 103–128.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing.* Cambridge, MA: MIT Press.

Cleeremans, A., & McCleeland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, **120**, 235–253.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, **14**, 179–211.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, **62**, 145–154.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.

Falkenberg, P. R. (1972). Recall improves in short-term memory the more recall context resembles learning context. *Journal of Experimental Psychology*, **93**, 39–47.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods and Instrumentation*, **14**, 375–399.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1–67.

Glanzer, M. (1972). Storage mechanisms in recall. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 5, pp. 129–193). New York: Academic Press.

Glanzer, M., Gianutsos, R., & Dubin, S. (1969). The removal of items from short-term storage. *Journal of Verbal Learning and Verbal Behavior*, **8**, 435–447.

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.

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **12**, 3–15.

Grossberg, S. (1978). Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, **17**, 199–219.

Henson, R., Norris, D., Page, M., & Baddeley, A. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology A*, **49**, 80–115.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, **36**, 73–137.

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.

Johnson, G. J. (1991). A distinctiveness model of serial learning. *Psychological Review*, **98**, 204–217.

Jordan, M. I. (1986). An introduction to linear algebra in parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Foundations* (Vol. 1, pp. 365–422). Cambridge, MA: MIT Press.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, **24**, 103–109.

Kahana, M. J. (1999). Associative symmetry and memory theory. *Manuscript submitted for publication.*

Kahana, M. J., & Caplan, J. B. (in press). Associative asymmetry in probed recall of serial lists. *Memory & Cognition.*

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press.

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, **34**, 419–426.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, **16**, 395–418.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 149–169.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, **96**, 25–57.

Li, S., & Lewandowsky, S. (1993). Intralist distractors and recall direction: Constraints on models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 895–908.

Li, S., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 837–847.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, **1**, 476–490.

Marshall, P. H., & Werder, P. R. (1972). The effects of the elimination of rehearsal on primacy and recency. *Journal of Verbal Learning and Verbal Behavior*, **11**, 649–653.

McGeoch, J. A. (1936). The direction and extent of intraserial associations at recall. *American Journal of Psychology*, **48**, 221–245.

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.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, **10**, 465–501.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, **64**, 482–488.

Murdock, B. B. (1963). Short-term memory and paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, **2**, 320–328.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, **1997**, 839–862.

Murdock, B. B., & Kahana, M. J. (1993). Analysis of the list strength effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **19**, 689–697.

Murdock, B. B., & Metcalfe, J. (1978). Controlled rehearsal in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, **17**, 309–324.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, **86**, 263–267.

Murdock, B. B., Smith, D., & Bai, J. Judgements of frequency and recency in a distributed memory model. *Journal of Mathematical Psychology*, **45**, 564–602.

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.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory and Cognition*, **21**, 689–698.

Neath, I., & Crowder, R. (1990). Schedules of presentation and temporal distinctiveness in human memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**, 316–327.

Neath, I., & Crowder, R. G. (1996). Distinctiveness and very short-term serial position effects. *Memory*, **4**, 225–242.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308–313.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39–57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **13**, 87–108.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, **105**, 761–781.

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, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981a). Order effects in recall. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 403–415). Hillsdale, NJ: Erlbaum.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981b). 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.

Rizzuto, D. S., & Kahana, M. (2001). An autoassociative model of paired-associate learning. *Neural Computation*, **13**, 1835–1852.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, **22**, 511–524.

Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, **89**, 63–77.

Rundus, D. (1980). Maintenance rehearsal and long-term recency. *Memory & Cognition*, **8**, 226–230.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317–1323.

Shiffrin, R. M. (1970a). Forgetting: Trace erosion or retrieval failure? *Science*, **168**, 1601–1603.

Shiffrin, R. M. (1970b). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 374–447). New York: Academic Press.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin and Review*, **4**, 145.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**, 1589–1627.

Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Englewood Cliffs, NJ: Prentice–Hall.

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.

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. *Science*, **229**, 287–289.

Received July 12, 1999; revised March 14, 2001