Scaling Behavior in the Temporal Context Model

Marc W. Howard
Department of Psychology
Syracuse University

The Temporal Context Model (TCM) postulates a distributed representation of temporal context that provides the cue for episodic recall tasks. TCM, coupled with the Luce Choice Rule for determining probability of recall, a conjunction referred to as TCMFR, is able to explain the existence of the long-term recency effect, as well as predicting the persistence of associative effects even with the inclusion of a delay between items. Here, quantitative predictions of TCMFR such as the magnitude of the delay interval is increased in continuous-distractor free recall are developed. The magnitude of the recency effect is operationally defined as the ratio of the probability of first recall (PFR) of the last list item to the PFR of the next-to-last item. Properties of associative effects are operationalized by using analogous measures derived from conditional response probability (CRP) curves. TCMFR predicts a decrease in recency with increasing delay. The rate of this decay and the qualitative pattern of change with increasing delay depend on the rate of contextual drift. For a range of values of the rate of contextual drift, TCMFR also predicts a transient increase in the recency effect as the length of the delay increases from zero. The model predicts that contiguity effects in free recall should follow a similar pattern, but that associative asymmetry, ubiquitously observed in free recall, should decay monotonically with increases in the delay interval.

In continuous-distractor free recall, presentation of each list item is followed by a distractor-filled interval. The length of this interval during list presentation is referred to as the inter-presentation interval (IPI). In addition, the last item in the list is followed by another distractor-filled interval prior to presentation of the memory test. This interval is referred to as the retention interval (RI). If the IPI and the RI are the same duration, this has the effect of "stretching out" the list in time. Early theorists studying the long-term recency effect hypothesized that the magnitude of the recency effect depends on the ratio of these two intervals (Bjork & Whitten, 1974; Crowder, 1976). This *ratio rule* hypothesis separately predicted data on the scale of minutes, hours and days (Glenberg et al., 1980; Glenberg, Bradley, Kraus, & Renzaglia, 1983).

The ratio rule hypothesis predicts that the magnitude of the recency effect should be scale-invariant; multiplying both the RI and IPI by a time t should have no effect on the magnitude of the recency effect. The Glenberg studies showed that manipulating IPI and RI separately at a particular scale resulted in changes in the recency effect that obeyed the ratio rule. However, more recently Nairne, Neath, Serra, and Byun (1997) have observed systematic deviations from scale-invariance in the serial position curve.

This manuscript was written in response to a question from Michael Kahana, who also provided numerous helpful comments on an earlier draft. Supported by 2-RO1 MH55687. Correspondence to Marc Howard, Syracuse University, Department of Psychology, 430 Huntington Hall, Syracuse, NY 13244-2340, or marc@memory.syr.edu.

The Temporal Context Model

The Temporal Context Model (TCM Howard & Kahana, 2002a) was developed to provide a description of context that could be used to generate recency and associative effects. TCM postulates that the state of the temporal context vector at time step i, $\mathbf{t}_i$ is a function of the prior state of temporal context, as well as some input vector $\mathbf{t}_i^{IN}$:

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

The scalar β is treated as a free parameter constrained such that $0 < \beta < 1$. In TCM, ρ_i is chosen such that $||\mathbf{t}_i|| = 1$. If the length of $\mathbf{t}_i^{IN}$ is always one, as we assume here, then at each time step i, ρ_i is given by

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

Here we will assume that the $\mathbf{t}^{IN}$ vectors caused by non-repeated words are mutually orthonormal vectors on an infinite-dimensional space of reals, so that the dot product in Eq. 2 is always zero. In this situation $\rho_i = \rho$ for all i, where ρ without the subscript is defined as

$$\rho := \sqrt{1 - \beta^2}. \quad (3)$$

Under these circumstances we find that the similarity of any two temporal context vectors falls off as an exponential function of the time between them:

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

We will use Eq. 4 extensively in the present ms to refer to the similarity between contextual states.

Murdock, Smith, and Bai (2001) used an equation similar to Eq. 1 to model contextual drift in applying TODAM2 to data on judgments of frequency and judgments of recency. In that treatment, context changed according to

$$\mathbf{t}_i = \rho \mathbf{t}_{i-1} + \sqrt{1-\rho^2}\zeta_i, \tag{5}$$

where ζ_i is a random vector whose elements are chosen from a normal distribution with standard deviation $1/\sqrt{N}$, where N is the dimensionality of the vector space. This equation is similar to Eq. 1 in that it describes a contextual representation that changes gradually over time. Further, the expectation of the dot product between any two context vectors derived from Eq. 5 falls off exponentially over time in a way similar to Eq. 4. The major difference between TCM and the treatment of Murdock et al. (2001) is the nature of the input vectors. In TCM, the input vector is not random, as in Eq. 5, but is caused by the item being presented. This means that, rather than defining a random walk, as Eq. 5 does, Eq. 1 results in a process in which context drifts in a direction determined by the item being presented. As we shall see, this difference enables TCM to model associative effects between items.

In TCM, the cue strength for an item is derived from the similarity of the current state of temporal context, used as the cue for episodic recall, to the state(s) of context that obtained when the item was presented. This is accomplished by inclusion of an outer product matrix $\mathbf{M}^{TF}$ connecting the $\mathbf{t}$ vector space to the $\mathbf{f}$ vector space. $\mathbf{M}^{TF}$ is updated at time step i using

$$\mathbf{M}^{TF}_i = \mathbf{M}^{TF}_{i-1} + \mathbf{f}_i \mathbf{t}'_i \tag{6}$$

when item pattern $\mathbf{f}_i$ is presented in context $\mathbf{t}_i$.[1] This enables us to calculate a cue strength a_i between a cue state of context $\mathbf{t}$ and each item $\mathbf{f}_i$ as follows. Assuming that item $\mathbf{f}_i$ was presented only once, we have:

$$\begin{aligned} a_i &:= \mathbf{f}'_i \mathbf{M}^{TF} \mathbf{t} && (7) \\ &= \sum_k \mathbf{f}'_i \mathbf{f}_k \mathbf{t}'_k \mathbf{t} \\ &= \sum_k \delta_{ik} \mathbf{t}'_k \mathbf{t} \\ &= \mathbf{t}_i \cdot \mathbf{t}, && (8) \end{aligned}$$

where δ_{ij} is the Kronecker delta, which is 1 if $i = j$ and zero otherwise and we have used the assumption of orthonormality on the $\mathbf{f}$ vectors. From this expression we can see that the cue strength for a particular item is a function of the similarity of the state of context used as a cue to the state of context in which that item was presented.

In calculating the probability of first recall (PFR), the cue is the state of context at the time of test, $\mathbf{t}_T$ and the state of context in which item i was presented is $\mathbf{t}_i$. The cue strength of item i in calculating the PFR is therefore just $\mathbf{t}_i \cdot \mathbf{t}_T$.

How might the activations specified by TCM be used to support recall? In applying TCM to free recall data, Howard and Kahana (2002a) used the Luce Choice Rule to generate probabilities of recall. Because the contextual cues specified by TCM could be applied to a number of tasks using a number of different methods, the conjunction of TCM with the Luce Choice Rule used to describe free recall data will be referred to as TCMFR. Given a set of cue strengths $\{a_i\}$, this rule states that P_i, the probability of recalling item i is given by:

$$P_i = \frac{\exp\left(\frac{2a_i}{\tau}\right)}{\sum_j \exp\left(\frac{2a_j}{\tau}\right)}, \tag{9}$$

where the sum in the denominator extends over all the potentially-recallable items in the list. Notice that τ controls the sensitivity of this process. As τ grows without bound, all items are equally likely to be recalled. As τ goes to zero, the most strongly activated item is recalled with probability one. To initiate free recall, the sum in the denominator of Eq. 9 includes all the list items. In describing associative effects, the sum includes all of the items in the list other than the just-recalled item. A complete description of free recall would require that there be some means to edit out repeated recalls, analogous to that used by SAM (Raaijmakers & Shiffrin, 1980). The present ms, however, relies on relative measures, and so this distinction can be ignored for the present purposes.

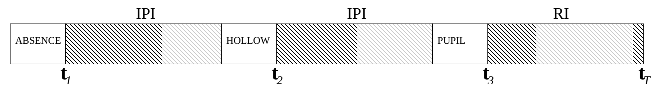

Figure 1. **Schematic of continuous-distractor free recall.** In the illustration, a list of three items is shown. A distractor of duration IPI intervenes between each item in the list. A distractor of duration RI precedes the test at time T. If the value of the IPI is the same value t as the duration of the RI, then the total delay between the time of test and item i is the time associated with the distractors, $t\,(L-i+1)$, plus the time associated with the item presentations, $L-i$, where L is the number of items in the list. Note that the duration that the items are presented is not assumed to change as the delay interval is increased. The figure is not necessarily to scale—items are usually presented for times on the order of seconds (one or two seconds), whereas distractor intervals are often on the scale of tens, or even hundreds of seconds.

Analysis

This section will derive expressions that describe predictions about the scaling behavior of recency and associative effects using TCMFR. Throughout, we will assume, unless an item is explicitly repeated, that an infinite series of non-repeating, mutually-orthonormal input patterns $\mathbf{t}^{IN}_i$ corresponding to the presentation of the words has been presented. In the following subsection, we will derive the pre-

[1] We assume that the $\mathbf{f}_i$'s are mutually orthonormal infinite-dimensional vectors, the same assumptions made for the $\mathbf{t}^{IN}$'s prior to learning.

dictions of TCMFR regarding the scaling behavior of the recency effect. We will do this for a very specific measure of the recency effect derived from the PFR curve. In the next subsection, we will derive predictions of TCMFR for associative measures calculated from conditional response probability (CRP) curves.

These predictions can be derived for specific experimental conditions. In treating the recency effect, we will consider the situation in which each item presentation lasts for one unit of time and each item is followed by a distractor interval equivalent to the contextual drift resulting from presentation of t items. That is, the retention interval (RI) and inter-presentation interval (IPI) are of a duration and type such that context changes as much from the beginning of the distractor interval to the end as it would have if t items had been presented. Because contextual drift in TCM depends on the amount of information being presented, it is not necessary to assume that this is precisely equivalent to the duration of the distractor interval for a specific experimental condition. More specifically, we assume that the $\mathbf{t}^{IN}$ vectors caused by the distractors are orthogonal, but not necessarily orthonormal. Subscripts are reserved for the states of context corresponding to item presentations. In treating associative effects, it is desirable to eliminate the effects of end-of-list context in calculating the CRP. This is consistent with the finding that the CRP changes little, if at all, over output positions in delayed and continuous-distractor free recall (Howard & Kahana, 1999). This is accomplished by assuming that an IPI equivalent to the presentation of t items intervenes between list items, but that an infinitely long RI precedes recall of the first item.

Recency

In describing the recency effect across delay paradigms Howard and Kahana (1999) utilized a measure called the probability of first recall (PFR), a serial position curve that considers only the first item recalled. The PFR was initially introduced by Hogan (1975) and used extensively by Laming (1999).

Here, we will measure recency using R, the ratio of the PFR for the last item in the list to the PFR for the next to last item in the list. This is given by

$$R := \frac{P_L}{P_{L-1}} \tag{10}$$

$$= \exp\frac{2\left(a_L - a_{L-1}\right)}{\tau} \tag{11}$$

where P is given by Eq. 9. The denominators of P_L and P_{L-1} are identical, leading to Eq. 11.

We are interested here in the behavior of R in a continuous-distractor experiment in which the length of the delay interval between items and at the end of the list is given by t, in units of item presentations (see Figure 1). Using Eq. 4, we have for item i in a list of L items

$$a_i = \rho^{t(L-i+1)+L-i}$$

$$= \rho^{(L-i)(1+t)+t} \tag{12}$$

This allows us to write Eq. 11 explicitly as

$$R = \exp\left[\frac{2\rho^t}{\tau}\left(1-\rho^{1+t}\right)\right]. \tag{13}$$

In immediate free recall, $t = 0$ and R simplifies to

$$R_{imm} = \exp\frac{2\left(1-\rho\right)}{\tau}. \tag{14}$$

from this expression it is clear that the value of the recency effect in immediate free recall is larger for smaller values of ρ, if τ is fixed. As the length of the delay increases to infinity, the recency effect decreases:

$$\lim_{t\to\infty} R = 1. \tag{15}$$

Figure 2**a** shows the changes in R as a function of t for several values of ρ. This last expression means that as the length of the delay interval increases, the relative probability of initiating recall with the last item in the list approaches that of the next-to-last item in the list. Given an infinite delay and a fixed duration of item presentation, this is not particularly surprising—the probability of recalling *anything* from the list should go to zero. The implications of this point will be taken up further in the general discussion.

The recency effect R is a function of ρ, τ and the delay interval t: $R(\rho,\tau,t)$. Increasing t from zero has the effect of changing variables, such that

$$R(\rho,\tau,t) = R\left(\rho',\tau',0\right), \tag{16}$$

where ρ' and τ' are defined according to

$$\rho' := \rho^{1+t} \tag{17}$$

$$\tau' := \frac{\tau}{\rho^t}. \tag{18}$$

From this we can see that increasing the delay has the effect of decreasing ρ and increasing τ. Decreasing ρ tends to constrict the scale over which temporal effects contribute (see Eq. 4). Increasing τ tends to equalize the probability of recall for all items in the list (see Eq. 9).

To evaluate the effect of increasing the delay on the recency effect, we will use the recency effect in immediate free recall, R_{imm}, as a baseline against which to measure changes in recency. Using the transformation defined by Eqs. 17 and 18 (or equivalently Eqs. 13 and 14) we find that

$$S := \frac{R(t)}{R_{imm}} \tag{19}$$

$$= \exp\left[\frac{2\left(1-\rho'\right)}{\tau'} - \frac{2\left(1-\rho\right)}{\tau}\right]$$

$$= \exp\left[\frac{2\rho^t\left(1-\rho^{1+t}\right)}{\tau} - \frac{2\left(1-\rho\right)}{\tau}\right]. \tag{20}$$

If TCM, coupled with the Luce Choice Rule were invariant with respect to t, then S would always be one. Clearly it is not; this model is inconsistent with the ratio rule.

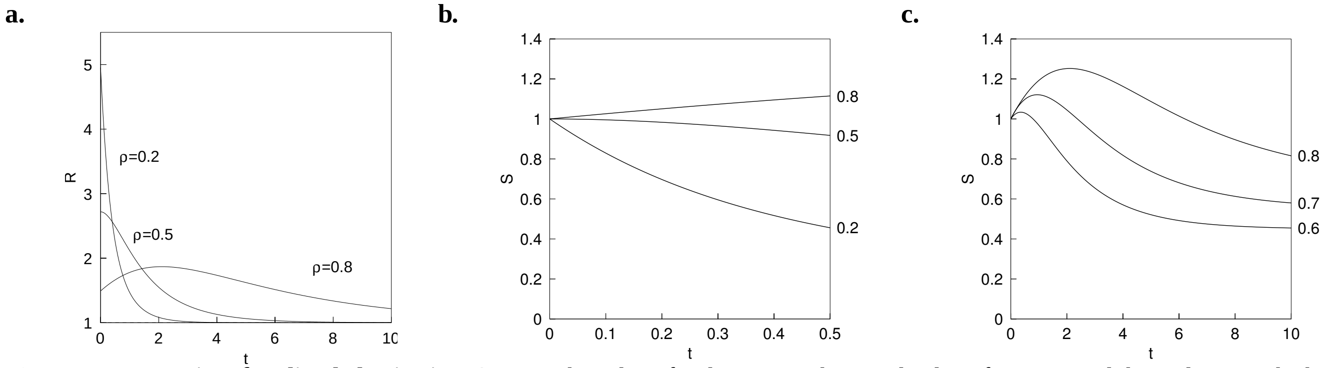

Figure 2. **Properties of scaling behavior in TCM**. **a.** The value of ρ determines the initial value of recency and the scale over which it declines, and the qualitative change in recency with increasing delay. The graphs describe predictions for a continuous-distractor experiment in which the interpresentation interval (IPI) and retention interval (RI) are both set to the value t. R (a measure of the recency effect in the probability of first recall) is shown as a function of t for three different values of ρ. For lower values of ρ, recency in immediate recall ($t = 0$) takes on larger values. For lower values of ρ, R more quickly approaches 1, indicating no recency effect. The qualitative shape of the function also depends on ρ, as illustrated in **b**. In this and the other panels, τ was set to one. **b.** The value $\rho = 1/2$ is critical for the qualitative change in recency with t. When $\rho = 1/2$, S starts out with a slope of zero as t increases from zero. For larger values of ρ, $S(t)$ starts out with a positive slope, for smaller values, a negative slope. **c.** TCM predicts peaks in S that depend only on ρ. For values of $\rho > 1/2$, the ratio S reaches a peak at a value t_{max} that depends on ρ. As ρ tends toward one, t_{max} increases and is associated with larger values of S.

Increases in recency with increasing delay. We will now examine Eq. 20 in more detail to understand the qualitative behavior of recency in TCMFR.

If

$$\rho^t - \rho^{1+2t} - 1 + \rho > 0, \tag{21}$$

then $S > 1$ and the recency effect becomes more pronounced than in immediate free recall. If the derivative with respect to t of the left hand side of the above inequality is positive at $t = 0$, then the recency effect will become more pronounced as the delay is increased. This requires that

$$\log\rho \left(\rho^t - 2\rho^{1+2t}\right) \quad > \quad 0. \tag{22}$$

Evaluating this expression at $t = 0$ gives us the condition for which S increases with t at $t = 0$:

$$\log\rho\,(1 - 2\rho) \quad > \quad 0, \tag{23}$$

which is satisfied for $\frac{1}{2} < \rho < 1$. In other words if $\rho > \frac{1}{2}$, then there is a value of the delay for which $S > 1$. If $\rho > \frac{1}{2}$, the model predicts that the recency effect in the PFR should be greater for some value of t than it is in immediate free recall.

Maximal values of recency. We have just seen that for $\rho > \frac{1}{2}$, there are some values of t for which $S > 1$, indicating an increase in the recency effect over that observed in immediate free recall. For all these values, S always ends up less than one as t increases without bound. It follows that there must be some point at which increasing t stops increasing recency. This happens when $\frac{dS}{dt} = 0$. Now,

$$\frac{dS}{dt} \quad = \quad S\frac{2}{\tau}\left(1 - 2\rho^{1+t}\right)\rho^t \log\rho. \tag{24}$$

This has a zero for

$$\rho^{1+t} = \frac{1}{2}. \tag{25}$$

For all values of $\rho > \frac{1}{2}$, the recency effect begins increasing as t increases from zero. For all values of $\frac{1}{2} < \rho < 1$, there is a non-zero value of t for which S takes its maximal value. This value of t is given by:

$$t_{max} = -\frac{\log 2}{\log\rho} - 1. \tag{26}$$

Associative effects

The preceding section discussed the scaling behavior of the recency effect in TCM. Here we discuss the behavior of associative effects in TCM as the delay between items is increased. In TCM, repetition of an item, whether during list presentation, or after successful recall of an item, results in retrieved temporal context. The input pattern retrieved when the item initially presented at list position i is repeated at a later time r is given by:

$$\mathbf{t}_r^{IN} = \alpha_O \mathbf{t}_i^{IN} + \alpha_N \mathbf{t}_i, \tag{27}$$

where α_O and α_N are non-negative scalars chosen such that $||\mathbf{t}_r^{IN} = 1||$. In practice the ratio of the two components is fixed using a new free parameter

$$\gamma := \alpha_N/\alpha_O. \tag{28}$$

This definition and the condition that $||\mathbf{t}_r^{IN} = 1||$ provide two equations to solve for two unknowns, allowing us to specify values for α_O and α_N at each time step.

The two components, $\mathbf{t}_i^{IN}$ and $\mathbf{t}_i$ together support asymmetric associations between items. Because $\mathbf{t}_i^{IN}$ is a component of $\mathbf{t}_{i+1}$ but not $\mathbf{t}_{i-1}$ (see Eq. 1), $\mathbf{t}_i^{IN}$ provides an effective retrieval cue for items that followed item i, but not those that preceded it. That is $\mathbf{t}_i^{IN} \cdot \mathbf{t}_{i+1} = \rho\beta > 0$, but $\mathbf{t}_i^{IN} \cdot \mathbf{t}_{i-1} = 0$. In contrast, $\mathbf{t}_i$ provides a symmetric retrieval cue for items presented near position i in the list. Prior work provides a more thorough discussion of the importance of these two components of retrieved context in generating associative processes (Howard & Kahana, 2002a; Howard, Wingfield, & Kahana, In revision) and transitive associations that are sensitive to hippocampal damage (Howard, Fotedar, Datey, & Hasselmo, Revised).

For mathematical convenience, we will assume in this section that we are considering a continuous-distractor list in which the length of the IPI is given by t and the length of the RI is infinite. After recall of item i, the current state of context is then given by

$$\mathbf{t}_{T+1} = \rho \mathbf{t}_T + \beta \mathbf{t}_r^{IN}. \quad (29)$$

Because the RI is infinite, $\mathbf{t}_T$ will have no similarity to the contexts associated with the list items. Recalling that the cue strength of an item is the dot product of the item's encoding context to the probe context we can calculate the cue strength of item j using only the context retrieved by item i:

$$\begin{aligned} a_j &= \mathbf{t}_{T+1} \cdot \mathbf{t}_j \\ &= \left(\rho \mathbf{t}_T + \beta \mathbf{t}_r^{IN}\right) \cdot \mathbf{t}_j \\ &= 0 + \alpha_O \beta \mathbf{t}_i^{IN} \cdot \mathbf{t}_j + \alpha_N \beta \mathbf{t}_i \cdot \mathbf{t}_j. \end{aligned} \quad (30)$$

This last expression allows us to write a simple expression for a_j in this experimental situation:

$$a_j = \begin{cases} \alpha_O \beta^2 \rho^{(1+t)(j-i)} + \alpha_N \beta \rho^{(1+t)|j-i|} & , \quad j > i \\ \alpha_N \beta \rho^{(1+t)|j-i|} & , \quad j < i \end{cases}. \quad (31)$$

From this expression, it is clear that there is an associative asymmetry between forward associations (e.g. $j = i+1$) and backward associations (e.g. $j = i-1$).

Contiguity

The contiguity effect refers to the general experimental finding that associations formed between words that were presented in nearby list presentations are stronger than the associations formed between words presented at distant list positions. To measure this tendency, we can define two measures analogous to R, the measure used to describe recency above. The measure C_F will provide an index of the contiguity effect in the forward direction; C_B will provide a comparable index in the backward direction:

$$C_F := \frac{P_{i+1}}{P_{i+2}} \quad (32)$$

$$C_B := \frac{P_{i-1}}{P_{i-2}}. \quad (33)$$

Using these definitions, and the expression Eq. 31 for the item activations, we find that

$$C_F = \exp\left[\frac{2\beta}{\tau}\left(\alpha_O \beta + \alpha_N\right) \rho^{1+t}\left(1 - \rho^{1+t}\right)\right] \quad (34)$$

$$C_B = \exp\left[\frac{2\alpha_N \beta}{\tau} \rho^{1+t}\left(1 - \rho^{1+t}\right)\right]. \quad (35)$$

In both of these expressions, the argument of the exponential has a similar form—a scalar that doesn't depend on t multiplied by $\rho^{1+t}\left(1-\rho^{1+t}\right)$. This form is similar to that of Eq. 13. Following the same logic we used to deduce that R increased with t for certain values of ρ, it is straightforward to reach the same conclusions for C_F and C_B: for values of $\rho > 1/2$, the contiguity effect shows a transient increase as t increases from zero. The maximal value of the contiguity effect occurs at the same value of t_{max} for which the recency effect shows a maximal value (Eq. 26). This illustrates the close correlation between recency and contiguity effects predicted by TCM.

Asymmetry. Equation 31 makes clear that TCM predicts an asymmetry favoring forward associations ($j > i$) over backward associations ($j < i$) in free recall. This asymmetry has been extensively observed in free recall (Howard & Kahana, 1999; Kahana, 1996; Kahana, Howard, Zaromb, & Wingfield, 2002; Klein, Addis, & Kahana, submitted); a similar asymmetry has been observed in serial recall (Addis & Kahana, submitted; Kahana & Caplan, 2002; Klein et al., submitted; Raskin & Cook, 1937).

Here we will define a measurement of asymmetry A as the ratio between the CRP at lag $i+1$ to the CRP at lag $i-1$. Using Eq. 9, this becomes

$$A := \frac{P_{i+1}}{P_{i-1}} \quad (36)$$

$$\begin{aligned} &= \exp\left[\frac{2\left(a_{i+1} - a_{i-1}\right)}{\tau}\right] \\ &= \exp\left(\alpha_O \beta^2 \rho^{1+t}\right). \end{aligned} \quad (37)$$

From Eq. 37 we can clearly see that associative asymmetry disappears as t increases to infinity:

$$\lim_{t \to \infty} \exp\left(\alpha_O \beta^2 \rho^{1+t}\right) = 1. \quad (38)$$

When $A = 1$, this means that the probability of recalling an item one position forward in the list is equivalent to the probability of recalling an item one position backward in the list. It is also clear from Eq. 37 that A decreases monotonically with increases in t. Unlike recency and contiguity, associative asymmetry decreases monotonically.

General Discussion

TCM is a model that has been used to describe the recency effect and associative effects in immediate, delayed and continuous-distractor free recall experiments (Howard & Kahana, 2002a). Here predictions for TCMFR, TCM coupled with the Luce Choice Rule for selecting items in free recall, were derived for the situation in which the length of the delay interval continuously varies from 0, representing immediate free recall, to larger values, representing continuous-distractor free recall with various distractor durations. For reasons of tractability, predictions were made for very specific measures. In treating the recency effect we examined the behavior of the ratio of the PFR of the last item in the list to the PFR of the next-to-last item in the list. In treating associative effects, the ratio of the CRP for lag $+1$ to the CRP for lag $+2$ was taken as a measure of the contiguity effect; the ratio of the CRP for lag $+1$ to the CRP for lag -1 was taken as a measure of associative asymmetry. While the specific nature of these predictions makes it relatively easy for them to be falsified, R is a restricted measure that does not capture every feature of the recency effect. For instance, the finding that R is bigger for continuous-distractor lists followed by a short RI than in immediate free recall (Figure 2) might seem counterintuitive if R is identified with "recency," as measured by, for instance, the probability of recalling the last item in the list. Similarly, R does not capture other features of the recency effect that are different in immediate compared to continuous-distractor free recall (Davelaar, Goshen-Gottstein, Ashkenazi, & Usher, In press; Howard & Kahana, 1999, 2002a). It should also be noted that the PFR can be affected by giving subjects specific recall instructions (e.g. Hogan, 1975), or by rehearsal strategies subjects adopt (see the PFR curves reported by Laming, 1999). The present predictions do not apply to these experimental situations.

TCM, coupled with the Luce Choice Rule for calculating probability of recall, makes several distinctive predictions regarding the scaling behavior of the recency effect. The present analyses have predicted several properties of recency and associative effects:

1. Decrease in recency and contiguity over the long term.
2. Transient increases in recency and contiguity for selected values of ρ.
3. Peaks in recency and contiguity that provide an independent means of estimating ρ from the data.
4. Associative asymmetry declines monotonically with increasing IPI.

These predictions provide strong constraints on TCM, coupled with the Luce Choice Rule for calculating relative probability of recall.

Experimental support

The present ms shows that TCMFR predicts that the recency effect as measured in the PFR should decline, and eventually disappear, as the length of the distractor interval is increased to infinity. On the face of it, this might appear to contradict the finding that recency effects are observed over very long intervals (e.g. Baddeley & Hitch, 1977; Glenberg et al., 1983). However, there is an important difference between the assumptions used in the current treatment and those experiments. Here the study materials and the presentation time were assumed to be unchanged by manipulations in the length of the delay interval, from immediate free recall to continuous-distractor free recall with infinite delay intervals. Studies that have shown recency effects over very long intervals have used stimuli that are more extended in time than the presentation of a single word—rugby matches in the case of Baddeley and Hitch (1977) and stories in the case of Glenberg et al. (1983). Typical immediate free recall experiments present words one at a time for a small number of seconds each. If subjects were presented with a single word daily for a week for, say, 1 s, and then tested at a 24 hour delay, it would be quite remarkable indeed if they remembered any of the words.[2] Extending item presentations in time would have the same effect as decreasing the value of τ (see Eq. 13), which would tend to recover the recency effect. Given the fact that massed item presentations are not nearly as effective as spaced item presentations, it may be necessary to treat extended presentations of an item using a more elaborate scheme than simply treating a longer presentation interval like multiple presentations of the item using Eq. 1.

It should be emphasized that the predictions derived here are for very specific measures of output order in free recall. In particular, the discussion of recency effects depends on the PFR. TCMFR describes a characteristic shape to the PFR—a monotonically decreasing function of recency that is highest for the last item. This pattern of data has been observed in a number of studies of immediate free recall (Hogan, 1975; Howard & Kahana, 1999; Kahana et al., 2002).[3] This finding stands in marked contrast to data from other free recall studies that show a "hump" in the recency portion of the PFR, or a PFR that "plateaus" at the end of the list (Laming, 1999). In those studies, the last item is not the most likely to be recalled first, but is approximately equal in recall probability to the last several items in the list. The resulting PFR shows a plateau at the end of the list. In other cases, subjects are more likely to start recall near, but not at, the end of the list and then recall forward to the end of the list. This pattern results in a hump in the PFR, such that the item a couple of serial positions from the end of the list is most likely to be recalled first. Laming (1999) has shown that non-monotonic PFRs are observed in reanalysis of classic free recall studies (Murdock, 1962; Murdock & Okada, 1970).[4] The reasons for the discrepancy between studies that generate smooth monotonic PFR curves and those that yield PFRs with humps or plateaus are unclear. However, several obvious explanations, such as modality of presentation, list length, and written versus verbal recall do not provide a clear-cut explanation.

TCMFR's account of recency clearly does not hold under conditions that yield PFR curves with humps or plateaus. Such PFR curves presumably reflect a conscious output strategy and/or the operation of rehearsal processes. Although the modality effect in immediate free recall is correlated with a change in output order (Nilsson, Wright, & Murdock, 1975), both auditory and visual presentation modalities give rise to non-monotonic PFRs (Unpublished observation from reanalysis of Murdock & Walker, 1969). In order to develop an explanation of phenomena, such as the modality effect in immediate free recall, that depend on output order in immediate free recall, it would be advisable to first observe these effects

[2] If they did remember the words, it would probably be a consequence of a strategy in which their experience with the items was not limited to 1 s, but rather involved extensive rehearsal.

[3] A number of unpublished studies collected by the author and colleagues has also shown monotonically decreasing PFR curves.

[4] Re-analysis of the data reported in Roberts (1972) by the author of the present ms also shows a non-monotonic PFR for various list lengths and presentation conditions.

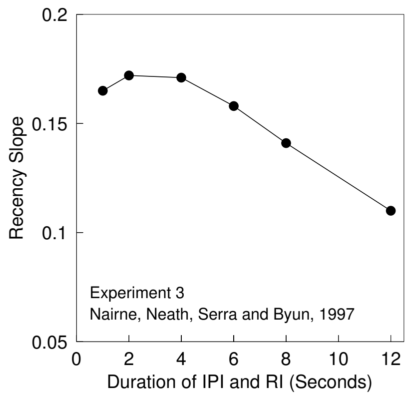

Figure 3. **Prior experimental work provides some support for the scaling behavior predicted by TCM.** Nairne, Neath, Serra, and Byun (1997) measured the recency effect in a continuous-distractor experiment in which the IPI and RI were set to the same value. Subjects were presented with 72 lists of six uppercase consonants. During the IPI and RI, subjects were required to read aloud single digits. At the end of the RI, subjects were presented with all 16 uppercase consonants used to generate the lists. Subjects "recalled" by selecting six letters from among the consonants.

under conditions that give rise to a smooth monotonic PFR function.

The current treatment yields predictions for experimental situations in which the delay intervals (but not the item presentations) are scaled up by the inclusion of a delay of length t. In two experiments, Nairne et al. (1997) found a reliable decrease in recency as t was increased from 1 to 12 s, consistent with the decrease over a wide range of t values predicted for all values of ρ. In their Experiment 3, Nairne et al. (1997) also showed data that suggests a transient increase in recency prior to the decrease (see Figure 3).

There are several caveats that should be kept in mind before accepting the data in Figure 3 as confirmation of the predictions outlined here. First, the Nairne et al. (1997) data was not, strictly speaking, free recall. At study subjects studied lists of consonants. At test subjects were provided with the entire pool of consonants. and were instructed to select the ones that were presented in the study list. This makes it very difficult to interpret order of recall effects, as the physical layout of the choices could make this more of a recognition test with uncontrolled test order than a true free recall test. Second, the data was reported as probability of recall (summed over output positions), rather than PFR, for which predictions were derived here. Third, the dependent measure used by Nairne et al. (1997) to describe the recency effect was the slope, or difference, in the serial position curve over the last two serial positions, rather than the ratio measure we used here. Further, Nairne et al. (1997) did not specifically test for the increase to determine if it is statistically reliable.

In manipulating IPI over a range from 0 to 16 s, Howard and Kahana (1999) found no reliable effect of IPI on associative effects in free recall (see also Howard & Kahana, 2002b, figure 4). Further experimentation is required to properly evaluate the predictions made in the present ms regarding the scaling behavior of recency and associative effects.

Comparison to other models of serial position effects

The changes in the recency effect predicted here could conceivably be generated by any number of other mechanisms that predict exponentially-decaying strength, as in Eq. 4, and a non-linear competitive retrieval rule analogous to Eq. 9. There are a number of potential candidates used to model other tasks in addition to the free recall model studied here. These candidates include so-called random context models developed to describe simple conditioning (e.g. Estes, 1955) and paired-associate learning (e.g. Mensink & Raaijmakers, 1988), and positional models developed to describe serial recall (e.g. Brown, Preece, & Hulme, 2000) or retention of order information (e.g. Lee & Estes, 1977). If these models were applied to free recall with appropriate assumptions it is possible that they would produce similar predictions regarding the scaling behavior of the recency effect to those described here. However, a description of associative effects, especially in continuous-distractor free recall, would be a major challenge for such models.

Nairne et al. (1997) presented modeling results that explained the decrease in the recency effect apparent in their data. The distinctiveness-diffusion positional model presented there showed a monotonic decrease in recency as the length of the delay interval increased. It is unclear whether this pattern is a requirement of that model, or a consequence of the particular parameter choices used in that study. If it is the former, then the observation of a reliable transient increase in recency would provide a means of experimentally distinguishing TCM from the diffusion/distinctiveness explanation of the recency effect proposed by Nairne et al. (1997).

Buffer models of the recency effect (e.g. Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1980) can explain part of the qualitative pattern of results predicted here. It is almost certainly the case that buffer models predict a decrease in the recency effect as the length of the delay is increased (e.g. Howard & Kahana, 1999). Further, buffer models could presumably accommodate a transient increase in recency in the probability of first recall as the length of the delay interval is increased by using a drop-out rule that favors retaining newer items. In a buffer model the probability of recalling any item from short-term memory to initiate recall is a function not only of the probability of that item being in STS, but also of the number of items in STS. This latter number should decrease as the length of the delay increases, potentially resulting in more recency. Similar logic can lead to a transient increase in contiguity effects. It is worth noting that buffer models with appropriate assumptions can describe PFR curves with humps or plateaus (Laming, 1999), whereas these would imply factors external to TCMFR.

TCM and buffer models yield similar predictions with respect to recency and contiguity over relatively short time scales. Given the extraordinary success of buffer models in explaining serial position effects over short time scales (Atkinson & Shiffrin, 1968; Kahana, 1996; Raaijmakers &

Shiffrin, 1980) this similarity is probably a hopeful sign for TCM. However, buffer models do not have the capacity to explain long-term recency effect over appropriately long time scales, whereas this appears within reach of TCM given stimuli that are appropriately extended in time. In particular, Howard and Kahana (1999) showed that the Raaijmakers and Shiffrin (1980) model, a buffer model that has been extensively applied to free recall, did not fit data on recency and contiguity effects over the scale of tens of seconds. Later work (Howard & Kahana, 2002a) showed that TCMFR was able to describe the existence of the long-term recency effect in the PFR and the persistence of associative effects over the scale of tens of seconds.

Conclusions

The present ms provides a series of detailed experimental predictions regarding the scaling behavior of recency and associative effects predicted by TCM, coupled with a Luce Choice Rule for selecting items for recall. This conjunction predicts that the recency effect, as measured by the ratio of the probability of first recall of the last two items in the list, should decrease over a wide range of delay intervals. For particular choices of ρ the model also predicts a transient increase in recency as the delay interval increases from zero. Comparable changes are predicted for contiguity effects. In contrast, associative asymmetry should decay monotonically with increases in the delay interval. These predictions provide strong constraints on the modeling of serial position effects in free recall. Detailed experimentation will be required to confirm or disconfirm these predictions.

References

Addis, K. M., & Kahana, M. J. (submitted). GLIO: Modeling the dynamics of serial learning.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, p. 89-105). New York: Academic Press.

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

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

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127-181.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., & Usher, M. (In press). A context activation model of list memory: Dissociating short-term from long-term recency effects. *Psychological Review*.

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

Glenberg, A. M., Bradley, M. M., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval theory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *12*, 413-418.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., & Gretz, A. L. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 355-369.

Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition*, *3*(2), 197-209.

Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (Revised). The Temporal Context Model in spatial navigation and relational learning: Toward a common explaination of medial temporal lobe function across domains. *http://memory.syr.edu/publications.html*.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *25*, 923-941.

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.

Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*(1), 85-98.

Howard, M. W., Wingfield, A., & Kahana, M. J. (In revision). Dissociations between recency and association in the Temporal Context Model: A description of the mnemonic deficit observed in cognitive aging. *http://memory.syr.edu/publications.html*.

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

Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition*, *30*(6), 841-9.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag-recency effects in free recall. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *28*, 530-540.

Klein, K. A., Addis, K. M., & Kahana, M. J. (submitted). A comparative analysis of serial and free recall.

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, *34*(5/6), 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.

Mensink, G.-J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-55.

Murdock, B., Smith, D., & Bai, J. (2001). Judgments of frequency and recency in a distributed memory model. *Journal of Mathematical Psychology*, *45*(4), 564-602.

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

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

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.

Nilsson, L.-G., Wright, E., & Murdock, B. B. (1975). The effects of visual presentation method on single-trial free recall. *Memory & Cognition*, *3*(4), 427-433.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower

(Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, p. 207-262). New York: Academic Press.

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.