A quantitative model of internal time in memory

Marc W. Howard    Karthik H. Shankar
Department of Psychology and Center for Memory and Brain
Boston University

William R. Aue    Amy H. Criss
Department of Psychology
Syracuse University

Revised for *Psych Review*, June 19, 2012, please do not quote without permission

Abstract

We experience ourselves existing in a point at the now of a time-line that extends back into the past and forward into the future. Because it incorporates perceptual information present at different physical times, this internal time-line must be a product of memory, the construction of which is extremely non-trivial. We present a mechanistic quantitative hypothesis for generating a distributed representation of internal time. Internal past time is generated from a scale-invariant representation of the temporal history leading up to the present moment. The temporal history provides a representation of what happened when; the representation becomes less distinct for events progressively further in the past. Internal future time is constructed from trajectories of prediction that sweep through future time points. These trajectories enable us to predict what will happen when in the future. Using this representation of internal past and future time, we construct simple behavioral applications to foundational problems in learning and memory. These include the judgment of recency task over short and long scales, the recency and contiguity effect across scales in episodic recall, interval timing and the time-left procedure, and temporal mapping phenomena in conditioning. We review a growing body of neural data suggesting something like this representation is implemented in the physical brain. Such a representation of internal time could form a cornerstone of a satisfactory theory of a broad variety of memory phenomena.

The authors gratefully acknowledge support from AFOSR award FA9550-10-1-0149, NSF award BCS-1058937 and NSF CAREER award 0951612. We thank Chris MacDonald, Howard Eichenbaum, Avital Adler and Hagai Bergman for sharing the data in Figure 21. We gratefully acknowledge helpful conversations with Bill Hoyer, Fabio Idrobo, Ralph Miller, and Aude Oliva. Address correspondence to Marc Howard, marc777@bu.edu.
We experience ourselves as residing in the present moment, moving as it were from an ever-receding past towards an ever-approaching future. This experience, and a wealth of behavioral and neural data, implies that the brain has a representation of our recent experience and predicted future events. One could understand many memory functions as operations performed on a representation of internal time. For instance, episodic memory, referred to as “mental time travel” by verbal theorists (e.g., Tulving, 1985a) could be understood as recovery of a prior state of the time-line so that it seems as if we are re-experiencing that prior moment. Similarly, a growing body of theoretical work holds that learning and conditioning must be understood as the learning of temporal contingencies between stimuli (e.g., Gallistel & Gibbon, 2000; Balsam & Gallistel, 2009; Matzel, Held, & Miller, 1988; Savastano & Miller, 1998); temporal contingencies could be learned by associating representations of internal time to experienced stimuli. Much of what we think of as working or short-term memory could be understood as detailed examination of the current state of the time-line and strategic attention to stimuli in the present (e.g., the N-back task McElree, 2001; Verhaeghen, Cerella, & Basak, 2004). One can also imagine accounts of a variety of other behavioral paradigms, including interval timing, delayed discounting, and statistical learning being built up from operations on a representation of internal time.

In this paper we describe a principled, concrete, mechanistic hypothesis for how the brain could maintain a representation of past (see also Shankar & Howard, 2012) and future temporal history. We then sketch a few behavioral applications that represent phenomena across what are currently thought of as distinct fields, including interval timing, episodic memory, short-term memory, and conditioning. Previous behavioral applications have sketched the utility of the representation of internal past time for problems in conditioning, timing, and the recency effect in free recall (Shankar & Howard, 2012). The present applications were chosen on the basis of their tractability in generating quantitative model predictions, the degree to which they require a genuinely temporal representation, the degree to which they sample diverse fields of cognitive psychology and their ability to illustrate important properties of the model, especially internal future time and the flexibility that comes from being able to recover previous states of past temporal history. Finally, we review a growing body of evidence that suggests the brain actually does maintain an ever-moving representation of the past and the future with at least some of the properties we have hypothesized.

The experience of internal time

Philosophers, notably Edmund Husserl (Husserl, 1966) have described the phenomenology of our internal experience of the passage of time. Consider the experience of a sequence of stimuli extended over physical time. To be more concrete, one may imagine a melody composed of a series of notes or a slideshow consisting of a series of visual images. At each moment, the senses give rise to perception of the current stimulus. According to Husserl, just after a perception has been registered, it induces a “retention.” The retention maintains the content of the perceptual experience. If a particular stimulus A is perceived at one point in time, the retention of that event still retains the properties that distinguish stimulus A from stimulus B. However, in retention, the retentions of A no longer has the property of being a direct perception. A retention has the character of being tied
Figure 1. Schematic of Husserl’s conception of the phenomenology of time. Physical time runs left to right. Internal time runs vertically. a. As a sequence of stimuli A B C are presented, each stimulus appears first as a perception at the appropriate moment of physical time. After the stimulus appears as a perception, it becomes a retention endowed with both the content of the perception that gave rise to it, but also a temporal character. As physical time moves on, the temporal character of each retention “tags” it as more and more distant from the present moment. When C is perceived, memory includes a retention B’, which has the content of stimulus B with a tag placing it one step in the past and A” which has the content of stimulus A with a tag placing it two steps in the past. b. If the sequence is well-learned such that the observer can predict the rest of the sequence, the observer can also utilize protention. With protention, internal time can have both a past (below the horizontal line) and a future (above the horizontal line). Anticipations of stimuli also have a temporal character. For instance, immediately after A is presented in physical time, B’ is a protention with the content of stimulus B and a temporal tag one unit in the future; C” is a protention with the content of stimulus C with a temporal tag two units of time into the future.

to a perception some amount of time in the past. As physical time continues to advance, the retention caused by stimulus A acquires the character of being progressively further in the past (diagonal line extending down from A in Fig. 1a). While the clarity or vividness of the retention diminishes as the occurrence of A recedes further and further into the past, the retention nonetheless retains some degree of “A-ness” even as it becomes less distinct.

At any particular moment in physical time, the process of retention implies an internal time-line that carries information about the history of perceptions leading up to that moment. Consider the line extending vertically down from the perception of stimulus C in Figure 1a. This implies that at that moment of time, the internal history contains the information that stimulus B was presented one unit of time in the past and stimulus A was presented two units of time in the past.

Representing retention. Before discussing Husserl’s description of internal future time, let us consider how to represent this time-line for the past. Husserl’s description of internal time seems to require two things. First, because the content of retentions does not change as they become less recent, this suggests that the representation must be able to distinguish stimuli from one another at all stages of retention. Second, because the temporal content of a retention changes over time, the representation must include information about how
Figure 2. Schematic for representing temporal history. a. Set of nodes to discriminate perception of three stimuli, A, B, and C. Here stimulus A is being perceived. b. Set of nodes to discriminate perception of A presented at three different stages of retention. Here stimulus A was experienced one time step in the past. c. Set of nodes to distinguish both stimulus identity and time of presentation. Here the nodes are active during perception of stimulus C at the last stage of the sequence A B C.

long ago the perception was experienced.

One simple way to build a temporal representation of recent history would be to have different processing elements, think of them as neurons in the brain or more abstract processing nodes, that respond to a conjunctive combination of stimulus and time. To build an idea of what this representation would look like, let us start by first considering what a representation of just stimulus or just time would entail.

Suppose that there are three stimuli that need to be distinguished, A, B, and C. Suppose further we have a set of nodes that respond differently to the stimuli. We could then examine the activity of these nodes at any moment and determine which, if any, of the three stimuli is being perceived at that moment. Our ability to distinguish one stimulus from another would depend on the number and selectivity of the nodes we have devoted to representing the stimuli. To make things simple, suppose that we have three nodes, each of which is active only when one of the three stimuli is being perceived. By examining the three nodes we can determine which, if any, of the three stimuli are being perceived (Figure 2a).

We can use an analogous scheme to distinguish states of temporal history. For the moment, let us consider the case in which there is only one stimulus, but we want to be able to distinguish how long ago it was presented. In order to do so, let us have a set of nodes, each of which is active when the stimulus has been experienced a particular time in the past. Suppose we have three nodes. One node is active when the stimulus is being perceived, the second node is active when the stimulus was presented approximately 1 s ago and the third node when the stimulus was presented approximately 2 s ago. Suppose for simplicity that the “tuning curve” of the nodes is such that there is no overlap in the activity of the three nodes. As the stimulus is presented and in the time following its presentation, the three nodes become active in sequence as the retention becomes progressively further in the past. At any moment in time, we can examine the activity of the three nodes and distinguish, say, whether the stimulus was presented 1 s ago or 2 s ago (Fig. 2b). Just as with distinguishing one stimulus from another, the accuracy with which we can distinguish different histories
from one another depends on how many nodes we have and how selective they are in their activity.

If retentions maintain their content as they recede across remembered time, this suggests that we should be able to distinguish stimuli at all stages of retention. To represent both stimulus and time, we need nodes that respond to a conjunction of stimulus and time—that respond when a particular stimulus was experienced a certain time in the past. Let us imagine these as organized into a sheet with each row corresponding to a different degree of retention, with the most recent moment in the top row, and each column corresponding to one of the three stimuli (Fig. 2c). When stimulus $a$ is presented it activates the node in the top of the $a$ column. As time goes on, activity “drips” down. When $b$ is presented some time later, activity starts in the top of the $b$ column and follows the activity in the $a$ column. By examining the set of all nodes at any one time we can gather information about the temporal history of stimuli leading up to the present moment.

**Protention of events in the temporal future.** Husserl argued that our experience of the passage of time is not limited to retention of the past but also an anticipated future that also contains temporal information. Suppose that a listener is hearing a familiar melody. As the melody unfolds, the listener can anticipate future tones. Husserl argued that these predicted future events, which he referred to as protentions, also have a temporal character. There are two claims here. First, that there is a prediction. Second, that the prediction contains temporal information. It should not require a detailed argument to convince the reader that we can create a prediction, at least of the immediate future. However, appreciating the point that predictions for events that are not immediately impending and that these predictions have a temporal character may require a bit more thought.

At any moment, there are a number of future events that might be expected to happen at various times in the near future. The meaning of the claim that protentions have a temporal character can be made more explicit by analogy with retention. We argued that as an event recedes further into the past, the retention does not simply become weaker. Rather, as a retention moves back in time, it conveys different information about the time at which the perception took place. Similarly, the claim is that protentions about events expected to happen further in the future are not merely weaker than events that are predicted to occur sooner in time, but carry information about how far in the future it will take place. Suppose that after perceiving stimulus $a$ in Figure 1b, protentions corresponding to the anticipated future presentations of $b$ and $c$ become available. According to Husserl, after the presentation of $a$, the protention for $b$ has the character of being one step in the future whereas the protention corresponding to $c$ has the character of being further in the future. As physical time unfolds, these protentions become less remote, moving continuously towards the present in register with one another.

Let us consider a real world example that may be familiar to many readers. Consider the case in which a landline phone rings while the listener is engaged in some task (e.g., washing dishes) that prevents him or her from immediately moving to answer the phone. The listener knows from experience the time between rings of the phone and that the call goes to an answering machine after the third ring. In the moment after the first ring, the listener has an expectation not merely that the phone will ring again, but an expectation of how far in the future that the next ring will take place. Similarly, there is not merely
an expectation that the answering machine will pick up; the prediction carries information about how far in the future that it will pick up. Note that the listener can take actions appropriate to the magnitude of the time interval in the moments immediately after the first ring. For instance, one might compare the amount of time before the machine picks up to the time it would take to finish rinsing the current dish, dry one’s hands and walk to the location of the phone. This judgment about future events could be used to inform a decision about whether to ask someone else in the house to answer the phone. This comparison would not be possible if there was no representation of the time at which these future events are likely to take place.

In the same way that a set of retentions can be strung together to form a temporal history extending back in time, it should also be possible to form a trajectory of anticipated future temporal states by starting at the present and examining the predicted events for successively more remote future times. Completing the picture of internal time, a vertical line running from bottom to top in Figure 1b traces an internal time-line that extends from retentions in internal past time, through the currently perceived stimulus (the horizontal physical time-line) and extending to protentions in internal future time.

**Behavioral evidence for a temporal representation of the recent past**

The foregoing argues that our experience seems to imply a rich representation of internal time that could be used to perform a variety of tasks. If one knew with certainty that such a representation existed, one could use it as a starting point to account for a broad array of findings from cognitive psychology. Thus far, however, all we have discussed is philosophy and introspection. Is there hard psychological evidence for such a temporally organized representation? No one finding will be sufficient in isolation to convince a skeptical reader. However, here we describe findings from one task that is particularly compelling for a temporally organized representation of the recent past. Because consideration of the details of behavioral performance strongly argue against alternative explanations, we will describe these results in some detail.

In Hacker’s (1980) short-term judgment of recency (JOR) experiment, subjects saw a list of consonants presented one at a time. The list was presented quite rapidly, at a rate of several items per second. Immediately after presentation of the list, subjects were given a pair of consonants and asked to judge which of them was presented most recently in the series. When subjects correctly selected the more recent probe item, their response time was a monotonic function of the recency of the more recent probe such that more distant items had longer response times. There was little or no effect of the recency of the less-recent probe item on reaction time. In contrast when the subjects incorrectly chose the less-recent probe, response time was a monotonic function of the recency of that item, but was only minimally affected by the recency of the more recent probe.

These results suggest that subjects examine their memory by starting at the present moment and scanning serially to earlier and earlier times to find the probe stimuli. Indeed this interpretation was presented by Hacker (1980) and favored by subsequent authors considering the paradigm (Hockley, 1984; McElree & Dosher, 1993; Muter, 1979). It is important to note that the scanning interpretation is not only supported by mean response times. Several studies included a careful consideration of response time distributions (Hockley, 1984; Hacker, 1980). For instance, Hockley (1984) found that the entire reaction
time distribution shifted as the more recent probe was moved further into the past. McElree and Dosher (1993) explicitly observed just this pattern using the response-signal procedure. The time after presentation of the probe at which subjects were able to discriminate above chance changed systematically with the recency of the more recent probe, as if subjects were scanning backward through their memory. Because both the RT distributions and response-signal findings reveal a period where no useful information about the time of occurrence of the probe is available, and because that period increases as the probe becomes less recent, they argue strongly in favor of a backward scanning model in which subjects examine their memory for each temporal interval successively.

The important point is that a backward scanning model is only possible if subjects have some form of organized temporal representation through which to scan. A scanning model implicitly postulates a temporal representation of recent history that can be sequentially accessed. Remarkably, when the subject is asked to choose the probe that came earlier in the list, rather than more recently in the list, correct response times no longer depend on the recency of the more-recent item but now depend (inversely) on the recency of the less-recent items (Chan, Ross, Earle, & Caplan, 2009). These results are just as one would expect if the subject began at the beginning of the list and serially scanned forward toward the present. The simple approach to representing temporal history we sketched above could support scanning if we assume that the subject is able to selectively examine the set of nodes representing different periods of time. These findings are extremely difficult to reconcile with an account in which there is not, in some sense, an internal time line that subjects can sequentially examine.

**Scale-invariance in internal time**

The results from the Hacker (1980) short-term JOR task suggest strongly that we have access to a temporal representation of the recent past. Given that, it seems unlikely that a rich source of information about the past would be utilized for just one isolated short-term memory task. This would require that the representation of internal time provide information further in the past than the one or two seconds required for the Hacker (1980) task. This naturally raises the question of how far back into the past should be retained. Experimental data from timing (Lewis & Miall, 2009), conditioning (Balsam & Gallistel, 2009), and episodic memory (Glenberg et al., 1980; Howard, Youker, & Venkatadass, 2008; Unsworth, 2008) all suggest that memory should persist, with similar properties, over a wide range of scales up to perhaps a few thousand seconds. In naturalistic timing and conditioning tasks, it is extremely difficult to know *a priori* what temporal scale will be relevant in a natural environment. Moreover natural signals contain potentially useful information over a very wide range of temporal scales (e.g., Voss & Clarke, 1975).

It seems that the representation of internal time should not end abruptly after some small duration of time has elapsed. On the other hand, it is certainly the case that less recent events are remembered less well than more recent events. These two considerations suggest a solution. As a perceived stimulus becomes less and less recent, its temporal representation should become less distinct. This means that the discriminability between points in the temporal history becomes smaller as the stimuli that caused the perceptions become less and less recent. We propose that our representation of history should degrade smoothly and continuously over a wide range of delays from about a hundred milliseconds to a few
thousand seconds. To enable this degradation of the representation to change smoothly over a wide range of time scales, the degradation should be mathematically scale-invariant.

Mathematical scale-invariance, the property that memory does not have a characteristic scale, has been argued to be a central principle of cognitive psychology (Chater & Brown, 2008) and is a key feature of many behavioral models of timing, conditioning and episodic memory (Brown, Preece, & Hulme, 2000; Brown, Neath, & Chater, 2007; Gallistel & Gibbon, 2000; Gibbon, 1977; Miall, 1989). There are numerous empirical findings that suggest a scale-invariant representation of history. For instance, the distribution of errors in an interval reproduction task scales up with the duration to be reproduced if counting is prevented (Rakitin et al., 1998, see Lejeune & Wearden, 2006; Wearden & Lejeune, 2008 for reviews). To the extent that the forgetting curve is described by a power-law function (e.g., Ebbinghaus, 1885/1913), it is mathematically scale-invariant. In conditioning experiments, the number of trials to acquire an autoshaping response only depends on the ratio of the stimulus-response delay interval to the inter-trial interval, and does not explicitly depend on the delay time scale explicitly (Gallistel & Gibbon, 2000). Although mathematical scale-invariance has not been established, both the recency effect and contiguity effect in episodic memory persist over a wide variety of scales, from hundreds of milliseconds to several hundred seconds (Glenberg et al., 1980; Howard, Youker, & Venkatadass, 2008). Note that a failure to find mathematical scale-invariance in a particular behavioral experiment does not allow us to conclude that the underlying memory representation is not scale-invariant. It is possible that there is a scale introduced by the experiment, or perhaps the subject is able to engage in some strategic behavior (e.g., counting) that sets a temporal scale. It is also possible that the retrieval process or some operation that takes place on the representation introduces a scale. In contrast, any finding of behavioral scale-invariance is very difficult to account for unless all of the components that produce behavior, including the memory representation, are scale-invariant.¹

There are two deviations from the simplistic representation of history described above (Fig. 2c) that enable our approach to be scale-invariant. First, the “tuning curve” of the nodes representing the history spread out in time with the delay they represent. That is, as in Figure 2c, different nodes respond optimally to a particular stimulus presented some time in the past. Oversimplifying for clarity, if one node shows peak responding when its preferred stimulus is 1 s in the past and shows a temporal spread of 0.2 s, then the node shows peak responding at 10 s will have a spread of 2 s. In other words, the nodes representing each time scale have the same relative spread in their tuning curve. Second, the spacing of the nodes is not constant. Rather the spacing is chosen such that the ratio between adjacent nodes is constant. Suppose that one node has a peak response at 1 s and the next node going down the column has a peak response at 1.1 s. Then if we look at the node adjacent to the 10 s node, we would find it has a peak response at 11 s. This property implies that the preferred delay for each node goes up logarithmically with the number of the node. These two properties will enable the representation of internal time to maintain information about a wide range of time scales. The accuracy of the history is weaker for events further in the past, but the degradation is smooth and continuous across a variety

¹It is logically possible that one could observe scale-invariant behavior when several components have scales, but those scales happen to cancel each other out. This kind of account would require an extremely unlikely convergence and is in any event deeply inelegant.
of time scales.

A scale-invariant representation of internal past and future time

Here we lay out a specific quantitative hypothesis for how to construct a scale-invariant representation of internal time. The model starts with a scale-invariant representation of temporal history (Shankar & Howard, 2012) that corresponds to retention of the past. This history, like the toy model sketched above, constructs a set of nodes that respond when a stimulus has been perceived a certain time in the past. This representation provides a “smeared” description of the history leading up to the present moment. The smear becomes more broad as the stimulus recedes into the past. We exploit properties of the representation of internal past time to construct trajectories of predictions going forward from the present moment. These trajectories constitute a representation of internal future time describing a set of protentions at times successively further in the future.

Before going into the mathematical details, we provide a brief overview of the components of this mathematical framework. We take the actual stimulus in physical time, \( f(\tau) \) as given. \( f(\tau) \) drives a set of leaky integrators, \( t(s) \). \( t(s) \) is used to construct an imperfect internal representation of the history of stimuli across internal time, \( T^{s}(\tau) \). At any one moment \( f \) is a vector where the dimensions of \( f \) correspond to the set of nodes representing the stimuli; physical time is referred to as \( \tau \). Thus \( f(\tau) \) refers to the actual stimulus in physical time. As in the simple example above we use a localist representation in \( f \) such that each distinct stimulus corresponds to a distinct stimulus dimension. This is done for ease of exposition and is not a necessary assumption.

At each moment, we construct a representation of the history leading up to the present moment, \( T^{s}(\tau) \), where \( \tau \) refers to internal past time. In the case with one stimulus dimension per stimulus, \( T^{s}(\tau) \) can be thought of as a matrix, or sheet of nodes, with each column of \( T^{s}(\tau) \) referencing a stimulus and each row referring to a value of internal past time. That is, each node of \( T \) corresponds to both a specific stimulus dimension and a specific value of \( \tau \). In this sense, \( T^{s}(\tau) \) is similar to the simple example in Figure 2. However, \( T^{s}(\tau) \) contains a “smeared” representation of when each stimulus was presented; this representation becomes increasingly smeared as the presentation of the stimulus recedes further into the past. Each node in \( T^{s}(\tau) \) reaches its maximal activation at a specific time after a stimulus presentation.

\( T^{s}(\tau) \) is constructed from an intermediate representation, \( t(s) \). The stimulus function \( f(\tau) \) drives a set of leaky integrator nodes in \( t(s) \) with a variety of rate constants \( s \). In the same way that each node in \( T^{s}(\tau) \) corresponds to a stimulus dimension and a time scale \( \tau \), each node in \( t(s) \) corresponds to a stimulus dimension and a rate constant \( s \). It turns out that \( t(s) \) is the Laplace transform of \( f(\tau) \); \( T^{s}(\tau) \) is an approximation to the inverse Laplace transform applied to \( t(s) \), thus recovering an approximation to the stimulus history \( f(\tau) \). An integer \( k \) controls the accuracy of the reconstruction. When \( k \) is larger, \( T^{s}(\tau) \) is less smeared and thus a closer approximation to \( f(\tau) \).

The representation of history, \( T^{s}(\tau) \) is used to support associative memory via an operator \( M \). \( M \) associates actual stimuli with the internal representation of history such that probing \( M \) with \( T^{s}(\tau) \) results in a prediction of what stimulus is likely to occur in the immediate future. Similarly, the translation operator, \( R \), can be used to generate a
Table 1: Summary of notation. Each row refers to a particular component. The second column gives the order of tensor of the component along with a label describing the meaning of the dimensions in parentheses. Scalars are tensors of order zero, vectors are tensors of order 1 and so on. Here $d$ refers to the stimulus dimensions.

<table>
<thead>
<tr>
<th>Order (dims)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>0</td>
</tr>
<tr>
<td>$f(\tau)$</td>
<td>1 ($d$)</td>
</tr>
<tr>
<td>$s$</td>
<td>0</td>
</tr>
<tr>
<td>$t(s)$</td>
<td>2 ($s \times d$)</td>
</tr>
<tr>
<td>$L^{-1}_k$</td>
<td>2 ($s \times \tau$)</td>
</tr>
<tr>
<td>$\tau^*$</td>
<td>0</td>
</tr>
<tr>
<td>$T(\tau^*)$</td>
<td>2 ($\tau^* \times d$)</td>
</tr>
<tr>
<td>$M$</td>
<td>3 ($d \times d \times \tau^*$)</td>
</tr>
<tr>
<td>$R$</td>
<td>2 ($s \times s$)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0</td>
</tr>
<tr>
<td>$p(\delta)$</td>
<td>1 ($d$)</td>
</tr>
</tbody>
</table>

...prediction of the timing of future stimuli, that is, the future state of $T(\tau^*)$. The translation operator $R$ is applied to the leaky integrator nodes $t(s)$. The temporal history constructed from the translated state of $t(s)$ approximates a state of $T(\tau^*)$ translated into the future. This predicted future state of $T(\tau^*)$ can be used to probe $M$ and predict the stimulus that is likely to occur at that future time. In the same way that the current state of $T(\tau^*)$ can be used to predict the stimulus that will be presented at the next moment, the estimate of the future state of $T(\tau^*)$ can be used to predict the stimulus that will be presented at the corresponding point in the future. By systematically varying the amount of translation we apply to $T(\tau^*)$ we can trace out trajectories of predicted future states. Table 1 provides a reference for the components of the mathematical framework for representing internal time.

*Maintaining a temporal representation of the past.* Let us start by assuming that there is a vector-valued stimulus function $f(\tau)$ that describes the perceptions available at
Figure 3. Schematic for visualizing $t(s)$ and $T(\tau)$. Here the stimulus dimensions (columns) have been ordered with the sequence in which the stimuli were presented as in Figure 2. That is, if the sequence A B C ... P was presented, the first column corresponds to nodes that respond to stimulus A, the second column corresponds to nodes that respond to stimulus B and so on. The stimulus function $f(\tau)$ would correspond to stimulus A being presented 16 steps in the past, B being presented 15 steps in the past and so on. Dark values correspond to higher levels of activation. Different rows correspond to different values of $s$. Each column of $t$ responds to the presentation of its stimulus with an exponential function across different values of $s$. At each moment, $t(s)$ is the Laplace transform of the stimulus function up to that point. The operator $L_k^{-1}$ approximates the inverse Laplace transform at each moment, taking $t$ into $T$. Stimuli presented in the recent past have a concise representation; stimuli further in the past have overlapping representations. When a stimulus is presented, it enters the column with a sharp peak that gradually “drips down” to values of $\tau$ representing the more distant past. As the peak moves down the column, it also spreads out. For visual clarity, values in different rows of $T$ have been scaled to their maximum value.

Each moment in physical time leading up to the present moment. Given $f(\tau)$, our initial goal is to store, i.e. remember, the stimulus function. Consider the differential equation

$$\frac{dt}{d\tau} = -st + f(\tau).$$

(1)

Here $t$ is a vector with dimensions corresponding to the dimensions of $f$. Equation 1 describes a leaky integration of $f(\tau)$ with rate parameter $s$. Suppose that we start with $t = 0$. Equation 1 says that $t$ will remain zero until a non-zero value of $f(\tau)$ causes the right-hand-side to become positive. Note that if $s$ is large, $t$ changes rapidly; if $s$ is close to zero $t$ changes very slowly.

The solution of Eq. 1 is just

$$t(\tau, s) = \int_{-\infty}^{\tau} e^{-s(\tau-\tau')} f(\tau') \, d\tau'$$

(2)

From this we can see several important properties of $t$. First, after a non-zero value of $f(\tau')$, $t$ decays exponentially over time with rate parameter $s$. Second, $t(\tau, s)$, with all values of
s, is the Laplace transform, with real s, of f(τ). That means that, assuming f(τ) is well-behaved, there is a one-to-one mapping between the function f(τ) over the interval (−∞, τ] and t(τ, s). Put another way, t(τ, s) contains all of the information about the stimulus function from beginningless time up to the present moment τ. It is worth noting that t(τ, s) contains information about the temporal history despite the fact that Eq. 1 is local in time. That is, in order to compute Eq. 1, one need only have access to the currently available value of the corresponding component of f(τ) and the current value of the corresponding component of t.

We can visualize t(s) as a sheet of values. Each column of t(s) corresponds to one of the stimulus dimensions of f(τ). Each row of t(s) corresponds to nodes with one particular value of s; different rows of t(s) correspond to different values of s. All rows of t(s) decay exponentially in the time following presentation of their corresponding stimulus dimension, but each row decays with a different rate. We refer to a specific element of t(s), corresponding to one specific stimulus dimension (i.e., one component of f) and one specific value of s as a node of t(s) (Figure 3).

*Extracting a representation of the past by approximating the inversion of the Laplace transform.* The insight that t(s) is the Laplace transform of f(τ) has another extremely important implication. This tells us that it is possible, in principle at least, to reconstruct the entire stimulus history up to the present moment by applying the inverse Laplace transform to t(s). If we had perfect resolution of t(s) (i.e., every possible value of s) and the computational ability to perform the inversion perfectly, we could perfectly recover all information about the stimulus function back to beginningless time. For modeling human memory, it would be preferable if the reconstruction were imperfect. In particular, the reconstruction should be less precise for events further in the past. Moreover, to enable a description of scale-invariant temporal behavior, the lack of precision should increase with delay in a scale-invariant manner.

Post (1930) introduced an elegant approximation to the inverse Laplace transform that provides just the properties we want. The inversion operator, which we refer to as L_k^{-1}, includes an integer parameter k that controls the accuracy of the inversion. Post (1930) proved that in the limit as k → ∞, L_k^{-1}t(s) perfectly reconstructs f(τ′ < τ). For finite k, the reconstruction is not perfect but is “smeared.” Smaller values of k result in a more smeared reconstruction of f(τ) than do higher values of k. In this paper, we fix k = 4 throughout for all of the behavioral applications. The quality of the behavioral demonstrations do not depend critically on this value of k. We refer to the approximation of the inverse Laplace transform as T(τ):

\[
T(τ) = L_k^{-1} \ t(s) = \frac{(-1)^k}{k!} s^{k+1} \ t^{(k)}(s)
\]

Equation 3 introduces a variable τ (pronounced “tau star”) that corresponds to a representation of internal past time. The values of τ are in one-to-one correspondence with the values of the rate constants s, with \( *τ = \frac{-k}{s} \). The nodes of \( T(τ) \) are in one-to-one correspondence with the nodes of t(s) (Fig. 3). Equation 4 explicitly writes out the form of L_k^{-1}.
notation $t^{(k)}(s)$ refers to the $k$th derivative of $t(s)$ with respect to $s$. The mathematical properties of $T$ and $L^{-1}_k$ are described in considerable detail in Shankar and Howard (2012). Here, we simply need to note that $L^{-1}_k$ is a linear operator that can be visualized as a square matrix operating on $t(s)$ to yield $T(\tau)$ (see Shankar & Howard, 2012, for details).

Whereas the nodes in $t(s)$ respond to a non-zero value of $f(\tau)$ immediately and then decay exponentially, in $T(\tau)$, the nodes respond to a non-zero value of their stimulus dimension after a particular latency controlled by their value of $\tau$ (Fig. 4a). Taking a column of $T(\tau)$ at a particular moment gives the history of a that stimulus dimension expressed as a pattern of activity across the set of nodes representing different values of $\tau$ (Figure 4b). Similarly, taking a row of $T(\tau)$ gives a pattern estimating the pattern of activity across stimulus dimensions at one point in past internal time. Shankar and Howard (2012) proved numerous properties of $T(\tau)$. Let us label the current moment as $\tau = 0$ and consider the spread in $T(\tau)$ caused by a stimulus presented at some particular time $\tau_o$ in the past. It turns out this spread is mathematically scale-invariant depending only on the ratio $\frac{\tau_o}{\tau}$. Moreover, the relative activity of a particular node as a function of physical time after presentation of the stimulus also depends only on the ratio $\frac{\tau_o}{\tau}$ (see Shankar & Howard, 2012, for the details of the derivation).

The distribution of $\tau$. Here we discuss two important issues regarding the selection of values of $\tau$. While it is mathematically acceptable for $\tau$ to range from zero to negative infinity, this would be unacceptable from a practical standpoint and implausible from a neural and behavioral standpoint. In addition, the mathematics do not require that adjacent rows of $T(\tau)$ be evenly spaced. The spacing of adjacent values of $\tau$ has a big impact on the behavioral predictions of the model and theoretical considerations motivate a specific choice. We discuss these two issues in turn.

For the mathematical model to be physically implemented, we need to specify a finite value for the shortest delay, $\tau_{\text{min}}$, and the longest delay $\tau_{\text{max}}$ that can be captured with this mechanism. As discussed extensively elsewhere Shankar and Howard (2012), we expect that $\tau_{\text{min}}$ is approximately one or two hundred milliseconds and $\tau_{\text{max}}$ does not extend more than a few thousand seconds in the past. Because these limits are far from the scale of the experiments we consider, we treat $\tau_{\text{min}}$ and $\tau_{\text{max}}$ as if they were zero and infinity for the purposes of doing the calculations.

If $T(\tau)$ is to represent four orders of magnitude of temporal history with any degree of resolution, one might be concerned that the number of values of $\tau$ would be prohibitively high if there is to be any degree of accuracy in the representation. However, the difference between the value of $\tau$ for adjacent nodes does not have to be a constant. In order to discuss the mathematical form of the distribution of $\tau$, let us introduce a variable $N(\tau)$ to represent the number of each node in a column of $T(\tau)$. The node representing $\tau = \tau_{\text{min}}$ would have $N = 1$, and the node representing $\tau = \tau_{\text{max}}$ would have $N = N_{\text{max}}$. In order to specify the

---

2The values of $\tau_{\text{min}}$ and $\tau_{\text{max}}$ are not identifiable on the basis of the behavioral data we are considering. For instance, although changing $\tau_{\text{min}}$ would have an effect on the time taken to scan through the representation to a particular time in the past, this is not separately identifiable from non-decision time.
Figure 4. Construction of a representation of remembered time. **a.** Dynamics of nodes in real time. Activity of a node in \( f \), two nodes in \( t \) and two nodes in \( T \) as a function of physical time. Top: the input in \( f(\tau) \) is non-zero for two periods of time. Middle: Activity of two nodes in \( t \) with different values of \( s \) are shown as a function of physical time. Each node functions like a charging and discharging capacitor with the value of \( s \) controlling the time constant. Note that the activity of the \( t \) nodes increases during the time \( f(\tau) \) is non-zero. Bottom: The activity of two nodes with different values of \( \tau^* \) in \( T \) as a function of physical time. The two nodes respond to \( f(\tau) \) after a characteristic delay. The value of \( \tau^* \) controls the latency at which each node responds to the stimulus. **b.** A representation of temporal history over internal past time. At one moment in physical time the activity across nodes with different values of \( \tau^* \) captures the stimulus history leading up to that moment. Top: \( f(\tau) \) leading up to the present moment. Bottom: Examining the activity of nodes with different values of \( \tau^* \) represents the history. The more recent stimulus presentation causes a relatively sharp peak; the less recent stimulus presentation causes a more diffuse peak. After Shankar and Howard (2012).

The choice of how to space the nodes has important implications. In order to preserve scale-invariance we require that \( g(\tau^*) \) be a power law. As an additional constraint, it is desirable if resources are spent wisely—adjacent nodes in \( T(\tau^*) \) should carry the same amount of additional information about the history. Note that constant spacing of nodes does not satisfy this requirement. The top row of figure 5 illustrates constant spacing. In the left panel we note that if the spacing between adjacent nodes is constant, \( g(\tau^*) \) is a constant and \( N(\tau^*) \) is a linear function of \( \tau^* \) (middle). On the right, we see that \( T(N) \) is just \( T(\tau^*) \). Note that adjacent nodes in \( T(N) \) are increasingly redundant with one another as the stimulus recedes into the past—as the stimulus recedes further in the past, the curves become more and more flat across evenly-spaced nodes. The bottom row of Figure 5 illustrates another
Figure 5. Illustration of the importance of $g(\tau)$, the number density of nodes. Top row: $g(\tau) = 1$. Bottom row: $g(\tau) = \tau^{-1}$. Left: the number density of nodes. Middle: Number of each node as a function of $\tau$. Right: $T(N)$ for three values of $\tau$. When $g(\tau) = \tau^{-1}$, $T(N)$ does not change width as a function of $\tau$. Note also that much larger values of $\tau$ can be represented with $g(\tau) = \tau^{-1}$.

possibility. In this case, the spacing between nodes goes up like $\tau$, so $g(\tau)$ goes down like $\tau^{-1}$ (left). This choice leads to a logarithmic function for $N(\tau)$ (middle). Critically, when $g(\tau)$ is chosen in this way, the shape of the distribution of $T(N)$, that is the distribution of $T$ across nodes does not depend on the delay of the stimulus (right, see Appendix A for a derivation). Note also that choosing $g(\tau)$ in this way also enables one to represent a much wider range of histories than constant spacing with the same amount of nodes. Throughout the remainder of this paper, we will fix $g(\tau) \propto \tau^{-1}$.

Extracting statistics of the world using an associative memory. In order to predict future stimuli, we associate the current state of the temporal history $T(\tau, \tau^*)$ to the currently-presented stimulus, $f(\tau)$. At retrieval each stimulus pattern is cued by the current state of history to the extent it resembles the state of the history in which that stimulus was encoded. That is, a stimulus is predicted to the extent that circumstances resemble the circumstances in which it was previously experienced. This association is stored in a tensor $M$. We write, in continuous time:

$$\frac{dM}{d\tau} = |f(\tau)\rangle\langle T(\tau, \tau^*)|.$$

The notation $|\cdot\rangle\langle\cdot|$ denotes the outer product of two tensors. The notation $\langle\cdot|\cdot\rangle$ denotes the inner product of two equi-ranked tensors.
Because \( f \) is a 1-tensor (a vector across stimulus dimensions) and \( T(\tau) \) is a 2-tensor (with columns corresponding to stimulus dimensions and rows corresponding to different values of \( \tau \)), \( M \) is a 3-tensor. To gain some intuition into \( M \), we can think of it as composed as a set of 2-tensors (matrices), one corresponding to each value of \( \tau \). For each value of \( \tau \), we can write

\[
\frac{dM^\tau}{d\tau} = |f\rangle\langle T^\tau|.
\] (7)

Here the notation \( T^\tau \) refers to the vector taken from \( T \) corresponding to a particular value \( \tau \) across stimulus dimensions. Similarly, \( M^\tau \) refers to the matrix within \( M \) that corresponds to a particular value \( \tau \).

A state of the temporal history can be used as a probe to predict what stimulus will be presented at the next time step.

\[
|p\rangle = M |T\rangle,
\] (8)

where \( p \) is referred to as the prediction vector (\( p \) is a one-tensor). Each stimulus is predicted to the extent that the states of the temporal history when it was encoded match the probe state of temporal history. We conceive of \( p \) and \( f \) as vectors in the same space, representing different states of activation across the same nodes. An advantage of treating \( p \) as a pattern across the same nodes as \( f \) is that one can understand memory retrieval as a result of the same deblurring process as perception of a noisy stimulus. A disadvantage is that one must have some means of separating encoding and retrieval to prevent \( p \) from overwriting \( f \) and vice versa. Here we assume that encoding and retrieval are sufficiently well-separated in time that we don’t need to worry about this problem. Mathematically, it is not necessary to make this assumption—in the same way that a row of \( t(s) \) or \( T(\tau) \) can have the same dimensionality as \( f \) but reflect the activity of distinct nodes, \( p \) and \( f \) could also be understood as distinct nodes in register to one another.

In order to calculate the right-hand side of Eq. 8, we need to do an integral over \( \tau \), weighting by \( g(\tau) \) to take into account the density of the nodes of \( T(\tau) \):

\[
|p\rangle = \int_{-\infty}^{0} M^\tau g(\tau) |T^\tau\rangle d\tau.
\] (9)

As mentioned earlier, \( g(\tau) \) is fixed to be proportional to \( \tau^{-1} \) throughout this paper.

**Predicting the future using a translation operator.** The matrix \( M \) enables us to predict the stimuli that will be presented at the next moment using the current state of the temporal history (Eq. 8). However, predicting the immediately impending moment is not the same as protention. In the same way retention requires that we have separate temporal information about events that happened at different moments in the past, protention requires that we have temporal information about how far in the future predicted events are expected to occur. Protention in this sense is possible in the current framework.

If at time \( \tau \) we had a means to obtain the state of the temporal history that will obtain at a future time \( \tau + \delta \), \( T(\tau + \delta, \tau) \), then we could just estimate the future state of

---

\(^4\text{We have suppressed the time dependence on the right hand side for clarity.}\)
\textbf{p} at time \(\tau + \delta\) using \(\mathbf{M}\). The problem of predicting remote future events is replaced with the problem of approximating future states of the temporal history \(\mathbf{T}(\tau + \delta, \hat{\tau})\) given only \(t(\tau, s)\). The dynamics of \(t(s)\) makes this relatively straightforward. Because \(t(s)\) changes deterministically over time in the absence of new input, given \(k(\tau, s)\), we can easily construct the \(t(\tau + \delta, s)\) that we would find \(\delta\) time steps in the future if there was no additional input during that interval. Because \(\mathbf{T}(\tau, \hat{\tau})\) is just constructed from the state \(t(\tau, s)\) available at that moment, we can directly calculate \(\mathbf{T}(\tau + \delta, \hat{\tau})\) from \(t(\tau + \delta, s)\).

To see how this can be accomplished, note that we can write a discrete analog of Eq. 1 for the 2-tensor \(t(s)\):

\[
t_{\tau+1}(s) = \mathbf{R} t_\tau(s) + f_\tau. \tag{10}
\]

It is easy to see that \(\mathbf{R}\) is just a diagonal matrix with \(e^{-s}\) on the row corresponding to each value of \(s\). If no stimuli are presented between time \(\tau\) and time \(\tau + \delta\), we can write:

\[
\mathbf{T}(\tau + \delta, \hat{\tau}) = \mathbf{L}^{\delta}_k t(\tau + \delta, s) \tag{11}
\]

\[
= \mathbf{L}^{\delta}_k \left[ \mathbf{R}^\delta t(\tau, s) \right] \tag{12}
\]

\[
= \left[ \mathbf{L}^{\delta}_k \mathbf{R}^\delta \right] t(\tau, s) \tag{13}
\]

Equation 12 follows from Eq. 11 because of Eq. 10. Applying the operator \(\mathbf{R}\) a total of \(\delta\) times advances \(t(\tau, s)\) forward \(\delta\) steps in time.\(^5\) Equation 13 just follows from Eq. 10 from the associative property. We get the same answer if we operate on \(t(\tau, s)\) with \(\mathbf{R}^\delta\) first, then operate on their product with \(\mathbf{L}^{\delta}_k\) as we would if we first operated on \(\mathbf{R}^\delta\) with \(\mathbf{L}^{\delta}_k\) and then used that product to operate on \(t(\tau, s)\). Equation 13 says that if we temporarily replace \(\mathbf{L}^{\delta}_k\) in Eq. 4 with \(\mathbf{L}^{\delta}_k \mathbf{R}^\delta\), then instead of the current state of history, we can estimate the future state of history \(\delta\) time steps in the future (see Figure 6).

Using the translation operator, at each moment \(\tau\) we can construct a trajectory of predictions going forward from the present moment by sweeping through different values of \(\delta\):

\[
|\mathbf{p}(\tau, \delta)| = \mathbf{M} \mathbf{L}^{\delta}_k \mathbf{R}^\delta |t(\tau, s)| \tag{14}
\]

\[
\overset{5}{\approx} \mathbf{M} |\mathbf{T}(\tau + \delta, \hat{\tau})| \tag{15}
\]

We refer to \(\delta\) as internal future time. Equation 15 is not an exact equality because other stimuli could intervene between time \(\tau\) and time \(\tau + \delta\). Here, \(\mathbf{p}(\delta)\) is a vector-valued function of \(\delta\). All values of \(\delta\) can be accessed, but not necessarily at the same instant.\(^6\)

\(^5\)To see this, note that if \(f_\tau\) is zero, \(t_{\tau+1}(s) = \mathbf{R} t_\tau(s)\). Continuing an additional time step we see that if the input \(f_{\tau+1}\) is also zero, \(t_{\tau+2}(s) = \mathbf{R} t_{\tau+1}(s)\). Substituting the previous expression for \(t_{\tau+1}(s)\) we find \(t_{\tau+2}(s) = \mathbf{R}^2 t_\tau(s)\). When the process is continued for \(\delta\) steps we find \(t_{\tau+\delta}(s) = \mathbf{R}^\delta t_\tau(s)\).

\(^6\)This is unlike \(t(s)\) or \(\mathbf{T}(\hat{\tau})\), for which all values of \(s\) or \(\hat{\tau}\) are accessible at any moment. It is conceivable that one could construct a 2-tensor version of \(\mathbf{p}(\delta)\), in analogy with \(t(s)\) or \(\mathbf{T}(\hat{\tau})\), in which the value of \(\mathbf{p}\) for all values of \(\delta\) are always available. In this case, \(\mathbf{p}\) would contain nodes that correspond to different points in future time. That model is mathematically possible but seems mechanistically implausible. If we wanted to simultaneously generate \(x\) values of \(\delta\) using the approach of altering \(\mathbf{L}^{\delta}_k\), this would require having access to \(x\) separate operators and \(x\) separate representations of \(\mathbf{T}(\hat{\tau})\). This seems like a tremendous storage cost for representations that are highly redundant with one another.
Comparing the model of internal time to empirical data

In the following subsections, we sketch behavioral applications of the mathematical framework for representing internal time described above. We start by building a simple scanning model and applying it to the Hacker (1980) data. We show that the scanning model can also be applied to a variety of JOR data over longer time scales, capturing the central findings, suggesting that there is a continuity between short-term and long-term memory. In the next subsection we demonstrate that this representation of internal time can describe the recency effect and contiguity effects in episodic memory across a range of time scales. This treatment demonstrates that the representation of internal time can perform the role of temporal context in retrieved context models of memory (Howard & Kahana, 2002; Sederberg, Howard, & Kahana, 2008; Polyn, Norman, & Kahana, 2009). In the next subsection, we show that the model of internal future time enables a simple treatment of the “time left” procedure (Gibbon & Church, 1981) in which subjects must make a decision based on rewards that will occur in the future. In the following subsection we show that the representation of internal time can account for temporal mapping phenomena observed in second order conditioning experiments (e.g., Cole, Barnet, & Miller, 1995). The temporal mapping demonstration utilizes both internal future time and the ability to recover prior states of temporal history and provides a close correspondence between this mathematical framework and the temporal coding hypothesis of Miller and colleagues (Matzel et al., 1988; Savastano & Miller, 1998).

In the last subsection we review neural evidence that suggests that the brain could implement this representation of internal time. We show evidence that is consistent with several key features of this representation. We review evidence showing that neurons in the hippocampus and prefrontal cortex exhibit correlated activity over time scales out to
at least a few thousand seconds. We review recent evidence suggesting that the medial
temporal lobe can show a “jump back in time” during performance of an episodic memory
task, a key requirement of the behavioral applications to free recall and temporal mapping.
We review recent evidence for “time cells” in the hippocampus and putamen that appear
to code for time since a stimulus was experienced, a key feature of the representation $T(\tau)$.

In the behavioral applications, our goal is not to describe the details of performance
in these tasks. Rather, our aim is to choose the simplest possible mapping between the
hypothesized representation of internal time and fundamental empirical findings. In some
applications (the contiguity effect in free recall, temporal mapping) we invoke the ability
to recover previous states of the stimulus history $T$, i.e., jumping back in time. In others
(JORs, the time left procedure), we do not invoke this ability. Some of the applications (the
time left procedure, temporal mapping) rely heavily on the representation of internal future
time whereas others (JORs, recency and contiguity in free recall) make no use of internal
future time. Our hypothesis is that the representation of internal time should play a central
role in a variety of tasks. This hypothesis is perfectly consistent with the position that
these tasks may rely on operations and/or additional sources of information appropriate for
performing in response to the particular task demands.

**Judgments of recency across time scales**

The JOR task taps subjects’ estimates of the time in the past at which an event took
place. In a relative JOR task, participants are presented two items and asked to select the
item that occurred more recently. In an absolute JOR task, participants judge the time, or
number of stimulus presentations, that have passed since the most recent presentation of
the test item. We construct a simple behavioral scanning model based on the representation
of internal time. We show that the model not only describes short-term JOR data from the
Hacker (1980) paradigm, but also accounts for key features of JORs observed over longer
time scales.

The scanning model is extremely simple. For each stimulus in the JOR probe, the
values of $T(\tau)$ for that stimulus are scanned sequentially starting from the present and
move subsequently to progressively less recent values of $\tau$. At each step of the scan, the
process terminates with probability $\alpha$ times the magnitude of $T(\tau)$ for that probe stimulus.
In a relative JOR task, the subject chooses the probe whose process terminates first. If
the search goes on long enough with neither probe being selected, the subject guesses.
In absolute JOR, the subject knows exactly the number of the node on which the search
terminated and uses that information as the basis for a response. Recall that node number
goes up logarithmically with $\tau$. The parameter $\alpha$ is the only free parameter central to the
search.

Appendix B provides a more formal description of the scanning model. A couple of
comments are in order to give the reader a stronger intuition before comparing the model
to data. First, the smear in the representation $T(\tau)$, controlled by $k$, makes it easier to
confuse pairs of events further in the past than recent events. That is, the representation of
two events separated in time by a fixed duration overlap less if the two events are close in
time and overlap more as the pair of events recede into the past. Because the probability
of termination depends on the value of $T(\tau)$, more overlap means that there is a greater
probability that the incorrect probe will be chosen. Second, more recent stimuli are more likely to return information even if it there is only one probe competing. This is because when $g(\tau^*) = \tau^{-1}$ the area under $g(\tau)T(\tau)$ for a once-presented item is a decreasing function of the delay. Both of these factors make memory worse for stimuli further in the past. Third, the choice of $g(\tau)$ combined with the assumption that subjects only have access to the number of the node the search is at when it terminates causes the scanning model to show a logarithmic increase in scanning time as a probe becomes less recent. Moreover, the scanning model of absolute JOR generates logarithmic increase in rated recency as a function of actual recency. In a continuous judgement of absolute recency, Hinrichs and Buschke (1968) observed a logarithmic function relating judged recency to actual recency (see also Hinrichs, 1970), suggesting a close correspondence with the model.

Serial scanning in short-term memory. The bottom panels in Figure 7 show predictions of the simple scanning model described above for the Hacker (1980) results. The parameter $\alpha$, which as discussed above controls the probability of the search terminating on a given value of $T(\tau)$, was manipulated to provide the best fit to the experimentally-observed accuracy. We calculated the probability of a correct response by numerically integrating the appropriate expressions (see Appendix B). The spacing between values of $\tau^*$ in the numerical integration was set to .005; the upper limit of integration was set to three times the longest lag. To estimate RT, we calculated the expectation of the number of the node at which the search terminated. Because of the allocation of nodes to different values of $\tau^*$, the expectation of node number is equivalent to taking the expectation of the value of $\log(\tau^*)$ at which the search terminated. There were no parameters manipulated to try and fit the RT data per se, but several choices that had to be made to get out reasonable numbers. We set the guess RT to the log of three times the $\tau^*$ corresponding to the longest list length to roughly correspond to the parameters of the experiment. We also added a constant to the expectation of $\log(\tau^*)$ to represent the lower limit of integration (and/or the non-decision time). This affected all RTs by the same constant amount and had no effect on their relative spacing.

Hacker (1980) proposed a serial scanning model that provided a good fit to his results. Subsequent authors (Hockley, 1984; McElree & Dosher, 1993) have also found that serial

---

Note that log scaling gives rise to Weber-Fechner law behavior for absolute judgements of recency. The constraint that the shape of the distribution across nodes should be invariant with the delay of the stimulus gives rise to this Weber-Fechner law behavior in this model of internal time. That is, the Weber-Fechner law is not assumed, but follows from the physical constraint that the mutual information between adjacent nodes should be scale-invariant.

Here one might have a slight concern that we are treating $\tau_{min}$ as zero. In the Hacker (1980) experiment, there was a total delay of 360 ms after the last stimulus was initially presented, including a mask, before the probe is presented. If $\tau_{min}$ is 125 ms, then the value $T(\tau_{min})$ for the most recently-presented stimulus has only decreased to about .1 of its peak value. Cutting off the remainder of the integral results would result in some correction for the most-recently presented stimulus. However, because the leading edge of $T$ is steep near zero, this region does not contribute much to the integral and the correction is minor. Of course, if one takes into account some additional processing time for the probe before the decision is made (or in fact the duration of the decision itself), even this small value decreases considerably. Even another 170 ms reduces $T(\tau_{min})$ to less than .005 of its peak value. The correction due to non-zero $\tau_{min}$, while non-zero, is probably minor.
Figure 7. Short-term memory scanning. In Experiment 2 of Hacker (1980), participants were shown a rapidly-presented list of consonants. At the end of the list, they were shown two probe stimuli and asked to select the most recent item. The top row shows experimental results for accuracy (top left), reaction time for correct responses (top center), and incorrect response reaction times (top right). Note especially that the recency of the less recent probe has little or no effect on correct RT, but has a dramatic effect on error RT. The bottom panels contain corresponding model predictions using a simple scanning model over internal past time. Memory for each of the probes was queried by scanning successive values of $^*\tau$. The probe that returned useful information first was selected. If neither probe returned useful information when $^*\tau$ reached three times the list length, the model guessed. Reaction time is estimated as the number of cells ($\log^*\tau$) that were scanned during the search. $\alpha = 2.69$. 
scanning models are necessary to describe the rate at which information about the probes becomes available. The present serial scanning model, although built out of a mechanistic representation of internal time, also captures the qualitative aspects of the data quite nicely. It does so for essentially the same reasons as prior scanning models. Despite not being fit to RT, the scanning model captures the basic pattern of results for response times for both correct responses and errors. Correct RTs depend a great deal on the lag of the more recent probe and only minimally on the lag of the less recent probe. In contrast, error RTs depend greatly on the recency of the less recent probe and almost not at all on the recency of the more recent probe. The model deviates slightly from the data in that the correct RT functions are not precisely flat, trending slightly in the opposite direction as the trend in the data. The model produces this behavior for exactly the same reason as the Hacker (1980) model—less recent items are missed at a higher rate than more recent items. As a consequence, the proportion of correct responses attributable to guesses goes up as the recency of the less recent item decreases. Because the guess RT is big relative to the non-guess RT, the result is a slight increase in RT. If one wanted to construct a more detailed model that eliminates this increase in RT for probes that are close together in the list, one could have the searches interact, competing with one another when they would be expected to terminate at about the same point in time.

In terms of describing behavioral performance, this demonstration does not do much to advance our ability to account for results in the short-term relative JOR task beyond previously-developed serial scanning models. Conceptually, however, the link between purely descriptive behavioral models and a hypothesis formulated at a lower level of description is extremely important. Specifying this lower-level representation correctly is essential for establishing connections between tasks that require very different descriptive models at the behavioral level. This is not true for all possible lower-level descriptions. For instance, if the behavioral scanning model was specified as scanning through, say, a push-down stack of slots in short-term memory, it would be extremely difficult to establish a connection with interval timing on the order of a thousand seconds (Lewis & Miall, 2009) or the contiguity effect in free recall across hundreds of seconds (Howard, Youker, & Venkatadass, 2008; Unsworth, 2008). The key point of this demonstration is not that a scanning model can describe the behavioral results from the short-term JOR task, but that this representation of internal time can support a scanning model.

Serial scanning model in long-term memory. In this subsection, we show results from two experiments that illustrate important properties of the scanning model with respect to JORs in long-term memory. First, we show that the simple scanning model captures the basic pattern of accuracy as a function of temporal distance in a relative JOR task (Yntema & Trask, 1963). Second, we show that the scanning model is able to account for the separate effect of multiple presentations of an item, a key feature of the data lends support to multiple-trace models of the JOR task (Hintzman, 2010).

Yntema and Trask (1963) performed a continuous relative JOR task over a broad range of time scales. On each study trial, a pair of words was presented. On each test
Figure 8. Long-term memory scanning. In Experiment 2 of Yntema and Trask (1963), subjects were presented with a continuous stream of pairs of words. On test trial, they were shown two probes and asked to select which one was presented more recently. The plot gives the probability of choosing item A as the more recent item (either correctly or incorrectly) as a function of its recency. The separate lines represent the recency of the other probe item, B. The model predictions are for the same scanning model as in Figure 7 with $\alpha = 9.71$.

trial, a pair of probe words was also presented, with the subject required to judge which of the probes was presented most recently. In a relative JOR, there are two relevant delays associated with the test of the two probes. Let us refer to the difference between the time of presentation of the two probes as the lag and the delay from the more recent probe at the time of test as the retention interval. In the relative JOR task, accuracy is relatively good when the retention interval is small compared to the lag; as the retention interval gets large relative to lag, accuracy decreases (e.g., Wasserman, DeLong, & Larew, 1984; Yntema & Trask, 1963; Lockhart, 1969). The basic findings (Figure 8a) are that, all other things being equal,

1. accuracy decreases with the retention interval to the more recent probe and
2. accuracy increases with the lag between the two probes.

The scanning model captures both of these properties quite well (Figure 8b). Note that the self-terminating scanning model used here would predict latency differences in long-term relative JORs similar to those observed in the Hacker (1980) data. However, it would be straightforward to construct an exhaustive search model that would make very similar predictions with respect to accuracy but make very different predictions about RT.

The Yntema and Trask (1963) findings are perfectly consistent with a strength account and do not independently require a temporal representation. However, the fact that subjects show memory for separate presentations of an item argues strongly against a strength account of JOR performance. Hintzman (2010) presented subjects a continuous absolute JOR task. Words were presented three times; we refer to the three presentations as P1, P2 and P3. On P1, the first presentation of an item, subjects were to judge it as new. For words the subjects judged as old, they were to rate its absolute recency. Let us refer to recency of the probe the second time it is presented as R1 and the recency of the probe (i.e., the difference between P2 and P3) as R2 (see Figure 9, top). At P2, subjects rated R1 for the words they judged as old. Their ratings are as one would expect on the basis of prior studies
Figure 9. Relative independence of separate presentations of an item. Top: In Experiment 1 of Hintzman (2010), words were presented on three occasions, P1, P2, and P3. At each presentation, the subject had to indicate whether the word had been presented earlier and, if so, estimate its recency. The recency of the word when it was repeated at P2 is denoted R1. The recency when the word was repeated the second time at P3 is referred to as R2. R1 and R2 were set to 5, 10, or 30. Bottom, Left: Empirical results for the absolute recency judgment for probes judged old as a function of R1. Note that R1 has little or no effect on the judgments at P3. Bottom, Right: The mean value of $\log \tau$ at which the scan through history returned. $\alpha = 15$.

(e.g., Hinrichs, 1970; Hinrichs & Buschke, 1968), displaying an approximately logarithmic judgment for the words they rated as old (Figure 9a, black line). At P3, there are two prior presentations of the probe, one at recency R2, and the other at recency R2+R1. As R1 and R2 are factorially varied, the question is whether R1 has an effect on the judgment at R2. The nearly horizontal colored lines in Figure 9a show that there is a minimal effect of R1 on the judgment of R2. There is a slight tendency for the judgment at P3 to increase with R1, but this is modest compared to the dramatic effect of R2, which accounts for the separation between the lines.

Because the Hintzman (2010) experiment only reported JOR ratings for probes that were also judged old, we only included results for searches that returned useful temporal information. The false alarm rate to the recognition judgment on the first presentation in the Hintzman (2010) data was only .03 (the hit rate on P2 was .93; on P3 it was .98) so it is reasonable that very few judgments were based on guessing. The relatively high value $\alpha = 15$ for the parameter controlling the instantaneous probability of the search returning means that the probability of a search returning a useful value at any point during the scan is correspondingly high. We assume, but do not model, a linear mapping between $N$, the
output of a particular search, and the subjects’ understanding of physical time, leaving the answer in terms of \( \log \tau \).

The model captures the basic trends in the data. It describes the form of the increase in the rating at P2 with R1. It describes the strong effect of R2 on the recency judgment at P3 and the relatively minimal effect of R1 on the judgment at P3. It even describes the slight increase in the judgment at P3 as R1 increases. In most cases, the query is already returning information due to the most recent presentation of the item. Because the search is self-terminating and there is a strong recency effect, additional information about the earlier presentation does not contribute much to the judgment. However, it does tend to move the judgment slightly towards the past.

If one examines the data closely, the model is missing a subtle feature of the data. Figure 9a shows that the judged recency was actually slightly lower when the item was initially presented with a short recency than if it was not presented. For instance compare the filled square all the way on the left of the data with the filled circle. This seems like a very small effect, but Hintzman (2010) reported it to be statistically reliable. The decrease in the judgment on P3 with small values of R1 cannot be captured by this scanning model. It is consistent with recursive reminding (Hintzman, 2004a, 2010, 2011). On the second presentation of the word, the subject makes a recency judgment. On P3, the subject could remember the stimulus presentation on P2 and/or the judgment that they made to that word on P2. In principle recursive reminding could be accomodated within the current framework by adding elements of the judgment to the stimulus history function and/or forming an association between the stimulus history and the judgment.

**Discussion.** We showed that a one-parameter backward self-terminating scanning model built on top of the distributed representation of internal time captures canonical effects from the JOR task. The model smoothly transitions between putative short-term (Hacker, 1980) to long-term (Yntema & Trask, 1963) results using a scale-invariant memory representation (Burrows & Okada, 1975; Donkin & Nosofsky, 2012). This is not to say that the behavioral model is scale-invariant. In fact, the scanning model has a scale set by \( \alpha \) and exhibits different properties in different regimes. The area under the curve of a column of \( T(\tau) \) for a probe stimulus presented a time \( \tau_0 \) in the past is proportional to \( \tau_0^{-1} \). When the recency of a probe item \( \tau_0 \ll \alpha \), memory scans return useful information almost certainly. In this regime, searches compete with one another. On the other hand, when \( \tau_0 \gg \alpha \), memory searches do not often return information and the searches behave as if they were independent. Although the underlying representation is precisely scale-invariant, the behavioral model exhibits qualitatively different behavior in these two regimes. This is an existence proof demonstrating that different qualitative patterns of behavior across scales do not imply different memory stores for each scale. It is extremely straightforward to introduce a scale to a scale-free memory representation in constructing a behavioral model. In contrast, it is quite difficult to cause scale-free behavior to result from operations on a memory representation that is not scale-free.

Over the last several decades of memory research, most previous accounts of the JOR task in long-term memory can be classified as strength models (but see Murdock, 1974;
Flexser & Bower, 1974; see Brown et al., 2000 for a treatment of short-term JORs using a clock model). In strength models, a scalar value associated with each stimulus is incremented when a stimulus is presented and decays as a function of time and/or intervening stimuli (e.g., Hinrichs, 1970). It should be noted that strength models of JOR can be conceptually quite elaborate. For instance, Bower (1972) described a stimulus sampling model in which the nominal stimulus was recoded into a set of sampled stimulus elements. The active stimulus elements were then conditioned to a list marker. At a JOR test, the subject examines the number of active elements evoked by the probe that are conditioned to the list marker. On average, the number of such elements is a decreasing function of the probes recency; this measure can support performance in JOR tasks (Bower, 1972). However, because the history of a nominal stimulus is summarized as a scalar, this approach boils down to a strength model. Another recent treatment of the long-term JOR task can also be classified as a strength model (Murdock, Smith, & Bai, 2001).

The evidence against simple strength models from JOR tasks and related tasks (e.g., the judgment of frequency task and judgment of serial position) is very strong (e.g., Hintzman & Block, 1971; Flexser & Bower, 1974; Hintzman & Block, 1973; Hintzman, Block, & Summers, 1973). At the same time, JORs are sensitive to strength. All other things being equal, strong probes are rated as more recent than weak probes (e.g., Hintzman, 2002, 2003, 2005; Peterson, Johnson, & Coatney, 1969). One of the tensions inherent in trying to construct a satisfactory model of the JOR task comes from the apparently conflicting requirements that the model must be on the one hand, sensitive to strength (e.g., Klein, Shiffrin, & Criss, 2007) but at the same time, able to capture information about the time of occurrence of an item rather than just retaining a simple scalar value (e.g., Hintzman, 2010). The present framework seems to be sufficiently flexible to account for the effects of strength on JORs. Suppose strengthening an item is equivalent to multiplying the values of $T(\tau)$ by some number greater than one. This is equivalent to turning up the value of the parameter $\alpha$ for that stimulus. This makes it more likely that a strong probe will retrieve information about its time of presentation and that this information will return information earlier in the scan. Under circumstances where discriminability would be expected to be very low, such as study-test designs where the effective delay to the probes is much greater than their difference, one would expect the JOR decision to be completely controlled by strength (Klein et al., 2007).

Scale-invariant recency and contiguity in episodic recall

In the free recall task, subjects are presented with a list of words and asked to recall as many of the words as they can in the order they come to mind. Data from the recency effect in free recall (Glanzer & Cunitz, 1966; Murdock, 1962) were instrumental in the adoption of the distinction between short-term and long-term memory (Atkinson & Shiffrin, 1968; Glanzer, 1972). The immediate recency effect was taken as evidence of the enhanced availability of items that remained active in short-term memory. Inserting a delay at the end of the list (Glanzer & Cunitz, 1966; Postman & Phillips, 1965) sharply attenuates the recency effect, presumably because the end-of-list items are displaced from short-term memory by the distractor. In some sense, the distinction between short-term and long-term memory can be understood as a strong assertion that memory has a characteristic scale defined by the capacity of short-term memory.
However, subsequent evidence on the long-term recency effect (Bjork & Whitten, 1974; Glenberg et al., 1980) showed that the recency effect in free recall persists even in circumstances where the delay between the last item and the test is long enough to eliminate the recency effect in delayed free recall. In addition to recency, free recall also manifests a contiguity effect. When an item is free-recalled, the next item recalled tends to come from a nearby list position (Kahana, 1996; Kahana, Howard, & Polyn, 2008). In analogy to the recency effect, there is also a long-term contiguity effect that is observed even with long delay intervals between the items (Howard & Kahana, 1999; Howard, Youker, & Venkatadass, 2008; Kılıç, Criss, & Howard, In press; Unsworth, 2008).

Temporal distinctiveness models propose an account of the recency effect across scales (Murdock, 1960; Neath & Crowder, 1990; Brown et al., 2007). If memory search takes place along a temporal dimension and is subject to competition among the items, then the relative spacing rather than the absolute spacing between the items should determine the magnitude of the recency effect. The distinctiveness heuristic leads naturally to a prediction of strict scale-invariance; this led to the proposal of a very successful model of serial position effects in free recall and serial recall (Brown et al., 2007).

Like the recency effect, the contiguity effect also seems to be sensitive to the relative rather than absolute spacing of the items. Both recency and contiguity manifest a sensitivity to temporal relationships, so why shouldn’t a model of temporal effects in memory account for both findings? The temporal context model (TCM, Howard & Kahana, 2002; Sederberg et al., 2008; Polyn et al., 2009) can be seen as an attempt to integrate the contiguity effect into the temporal distinctiveness framework. If the recency effect results from looking backwards through time from the present moment, perhaps the contiguity effect results from first jumping back in time (Tulving, 1983), then looking for other list items from this new vantage point. Unlike short-term store, which contains a discrete number of stimuli at any one time, in TCM items do not abruptly fall out of the temporal context vector. This property enables TCM to describe scale-persistent recency and contiguity (Howard & Kahana, 2002; Sederberg et al., 2008). But temporal context in TCM is not mathematically scale-invariant (Howard, 2004).\footnote{Although the evolution of the temporal context vector in TCM obeys a discrete version of Eq. 1, the temporal context vector only maintains a single value of $\rho$. This is equivalent to a single row of $t(s)$ with $s = -\log rho$. Choosing a single $\rho$ fixes a scale. One might say that TCM is similar to short-term memory in that it imposes a scale on memory, but is unlike short-term memory in that the scale is less rigorously enforced (for further discussion see Usher, Davelaar, Haarmann, & Goshen-Gottstein, 2008; Kahana, Sederberg, & Howard, 2008; Howard, Kahana, & Sederberg, 2008).}

In practice it is very hard to empirically establish whether free recall is scale-invariant because there are many aspects of a free recall experiment procedure that could introduce a scale. Nonetheless it appears clear that recency and contiguity persist over scales from at least a few hundred milliseconds (Neath & Crowder, 1996; we have observed contiguity effects with presentation times down to 250 ms but not at a presentation time of 125 ms), up to at least hundreds of seconds (Howard, Youker, & Venkatadass, 2008; Unsworth, 2008) and perhaps even to the scale of months (Moreton & Ward, 2010). Here we focus on describing the recency effect, as measured by the probability of first recall (PFR) and the contiguity effect, as measured by the conditional response probability as a function of lag (lag-CRP) with special attention to capturing their qualitative behavior across time scales.
To calculate the PFR, one simply takes a serial position curve for the first word that the subject recalls in their free recall protocol. To calculate the lag-CRP, one counts the number of times that recall transitions of a certain lag are observed and divides that number by the number of times correct recall transitions of that lag could have been observed given the serial position of the just-recalled word and the previous recalls (Howard & Kahana, 1999; Kahana, Howard, Zaromb, & Wingfield, 2002; Howard, Addis, Jing, & Kahana, 2007). The “lag-CRP” curve shows remarkable consistency across experiments (Kahana, Howard, & Polyn, 2008) and is not simply a confound of serial position effects (see especially Farrell & Lewandowsky, 2008; Howard, Sederberg, & Kahana, 2009) nor correlated encoding processes (Howard, Venkatadass, Norman, & Kahana, 2007; Kılıç et al., In press).

In the previous subsection Hacker’s (1980) serial scanning model served as a proxy between the underlying representation of internal time and a description of observable behavior. Here, TCM plays an analogous role—we describe PFR and lag-CRP curves using just the account offered by TCM, but with $T(\tau)$ playing the role of the temporal context vector. In TCM, the temporal context vector is the exponentially-weighted sum of recent inputs caused by the stimuli; the temporal context vector in TCM is essentially a strength model for recent experience. Here, $T(\tau)$ is a much richer representation of temporal history. In TCM, when an item is remembered, it causes a partial reset of the state of the temporal context vector to the state in which that item was encoded—a “jump back in time.” Here, remembering an item causes partial recovery of the state of $t(s)$ when the item was encoded. Because $T(\tau)$ is generated from the current state of $t(s)$, partial recovery of $t(s)$ results in a partial recovery of the state of internal past time when the item was encoded.

In order to generate potential recalls, we generate a prediction via $M$ (Eq. 8). To describe the PFR, we use the current state of the temporal history at the time of the free recall test as a cue. Although the details of our treatment of recency are slightly different due to fixing $g(\tau) \propto \tau^{-1}$ here, this approach is quite similar to the treatment of recency in Shankar and Howard (2012). In calculating the lag-CRP we assume that the delay and/or the process of averaging over many retrieval attempts renders the state of $T$ prior to retrieval of the recalled item ineffective in generating the contiguity effect (see Farrell & Lewandowsky, 2008; Howard, Sederberg, & Kahana, 2009). The contiguity effect has not been described in previous publications utilizing $T(\tau)$ to describe behavior (Shankar & Howard, 2010, 2012).

The prediction vector (Eq. 8) specifies some degree of activation for all of the stimuli that can be represented in $f$. In free recall, only one of those items is recalled at a time, so we need some mechanism to choose a recall. Let $p_i \equiv \langle f_i | M | T \rangle$ (16) be the match between the $i$th word in the list and the prediction vector. Then, we assume that the probability of recalling item $i$ in a list of $\ell$ items is

$$P_R(i) = \frac{p_i^a}{\sum_{j=1}^{\ell} p_j^a},$$

where $a$ is a free parameter specific to the recall applications. We fix $a = 2$ in all the free recall applications that follow. When calculating the lag-CRP, we add a constant $c$ to the $p_i$
for each item. This provides a way to take into account associative cues that do not depend on the lag between the just-recalled item and potential recalls. This would include semantic effects as well as the effect of the preceding state of $T$ on the lag-CRP. The problem of generating predictions for recency and contiguity effects thus reduces to the problem of calculating the relevant values of $p_i$ describing the degree to which each item is predicted by the state of $T$ at the end of the list and a time after recovery of a list item, respectively.

**Scale-invariant recency in the PFR.** Because list items are encoded in their state of $T$, and because $T$ changes gradually over time, this approach will generate a recency effect. Because $T$ changes in a scale-invariant way and because the retrieval rule Eq. 17 describes a power law, the model predicts recency effects that are precisely scale-invariant. In the continuous distractor procedure, the subject experiences some type of stimulus, either a word or a distractor, at each moment in order to prevent rehearsal. The details of the calculation can be found in Appendix C. Under these circumstances, if a stimulus is presented every $\delta \tau$ seconds, then a word presented $n$ time steps in the past will have activation

$$p_n \simeq n^{-1}(\delta \tau)^{-2}$$  \hfill (18)

when $g(\tau) = \tau^{-1}$. \hfill (12) All that remains in order to apply Eq. 17 to free recall experiments is to plug in the appropriate delays between the list items and the time of test. Let the number of distractor stimuli between two words be $D - 1$, making the total time elapsed between the presentation of two successive words (including the presentation of the first word) just $D$. Let us denote the duration of the distractor interval after the last word in the list and prior to the recall test as $d$. Now the positions of the list items in the sequence of stimuli are $(d + D.n)$, where $n$ is word number starting from the end of list and $D.n$ just refers to $D$ multiplied by $n$. We can now rewrite eq. 17 in the following form

$$PFR(n) = \frac{(d + D.n)^{-a}}{\sum_m (d + D.m)^{-a}},$$  \hfill (19)

where the summation over $m$ goes over all $\ell$ words in the list.

In order to fit this expression to experimental data, we simply set $d$ and $D$ to correspond to the experimental values. In continuous distractor experiments, the delay between items is not repeated at the end of the list. To get the delays in Eq. 19 to correspond to the experimental values, we let $n$ in Eq. 19 run from 0, corresponding to the last word in the list, to $\ell - 1$, corresponding to the first word in the list in continuous distractor experiments. In final free recall experiments, the delay after presentation of an item (consisting primarily of the recall period and the time between lists) is repeated after the last item. To get Eq. 19 to correspond to the delays in the actual experiments, we let $n$ run from 1, corresponding to recall of the last list $\ell$, corresponding to recall of the first list. For all demonstrations $a$, the only free parameter, is fixed at 2. This simple framework is sufficient to account for the basic properties of the recency effect in free recall measured in the PFR.

**Property 1:** A delay at the end of the list decreases the sharpness of recency

The top left panel of Figure 10 shows the PFRs from a free recall experiment (Polyn, et al.,

\footnote{Notice that $(\delta \tau)^{-2}$ will factor out when we plug into Eq. 17.}
unpublished data) where subjects recall lists of 16 items with different end of list delays of 0, 8 and 16 seconds. In the top right panel, Eq. 19 is plotted $D = 1$ and $d = 0, 8$ and 16 for the three conditions. The form of Eq. 19 makes it clear why increasing $d$ should decrease the sharpness of recency. As $d$ increases, the effect of $n$ on the probability of first recall is decreased.

**Property 2: Increasing the gap between the items increases the sharpness of recency.** The bottom left panel of Figure 10 shows the PFRs of a continuous distractor free recall experiment (Howard & Kahana, 1999) where subjects recalled lists of 12 words. The conditions had a fixed retention interval of 16 s and inter-item intervals of 2, 4, 8 or 16 s. In the bottom right panel, Eq. 19 is plotted with $d = 16$ and $D = 2, 4, 8$ and 16. Although the numerical similarity to the data is not particularly strong, the qualitative pattern is correct. Again, the functional form of Eq. 19 makes it clear why increasing $D$ while holding $d$ fixed should increase the sharpness of recency. Note that increasing $D$ with a fixed $d$ has a similar effect as decreasing $d$ with a fixed $D$.

**Property 3: Recency effect persists across multiple time scales.** The scale-invariance of $T$ enables the model to generate recency effects across a broad variety of time scales. If nothing in the performance function (e.g., Eq. 17) disrupts the scale-invariance, models built using $T$ predict precisely scale-invariant recency. The recency effect over time-scales of seconds to dozens of seconds is already demonstrated in Figure 10. Final free recall experiments have demonstrated recency effects over even longer time scales. In final free recall experiments, subjects study and recall a number of lists. At the end of the session, subjects are asked to free recall all the words from all the lists in the order they come to mind. In final free recall experiments, the recency effect is observed spanning across several lists over several hundred seconds (Glenberg et al., 1980; Howard, Youker, & Venkatadass, 2008; Tzeng, 1973).

To describe final free recall, ideally we would sum over the contributions from all items within each list and compare the relative activations of different lists. However for the purpose of a qualitative demonstration, we consider an approximation. Let us assume that every word in a given list is equally recallable at the time of the final free recall test. If we had $\ell$ items within a list with precisely the same activation, $\ell$ would factor out of Eq. 19. To estimate the probability of initiating final free recall with a word from list $n$, we use Eq. 19 with $d$ and $D$ chosen in accordance with the experiment and with $n$ running from 1 to the number of lists. Our approximation is reasonable to the extent that the variability in within-list recency is small compared to the variability due to across-list recency.

Figure 11 shows the recency effect, as measured by the PFR, from immediate and final free recall in Howard, Youker, and Venkatadass (2008). In this experiment, lists of 10 words were followed by an immediate test. The PFR shows a strong recency effect extending over the last several items (top). The PFR across lists is shown in the bottom row of Figure 11. Again, there is a strong recency effect across the last several lists of items. We described these findings using Eq. 19. As for all the model predictions for free recall data, we fixed $a = 2$. For immediate free recall, we took $d = 1$ and $D = 1$; for final free recall we took $d = 250$ and $D = 49$, consistent with the values of the experiment.
Figure 10. The recency effect depends on the duration of filled distractor intervals between stimulus presentations and during the delay prior to recall. Probability of first recall (PFR) from two experiments (Polyn et al., unpublished, top; Howard and Kahana, 1999, bottom). In the Polyn et al., data (top), the duration of the filled distractor interval at the end of the list is manipulated. The recency effect decreases as the delay interval increases. In the Howard and Kahana data (bottom), the duration of the distractor interval between items is manipulated while the delay prior to test is kept constant at 16 s. Increasing the delay between items enhances the recency effect. On the right side, the corresponding experiments are modeled using eq. 19 with $a = 2$. Note the change of scale on the bottom row.
Figure 11. Evidence suggesting scale-invariant recency in free recall. Subjects performed immediate free recall on lists of ten items. After 48 such lists, they recalled all the items from all the list. Within-list (top) and across-list (bottom) probability of first recall (PFR) functions. Data from Howard, Youker & Venkatadass (2008) is shown on the left. Eq. 19 with $a = 2$ and delay parameters set to experimental values is shown on the right.

While scale-invariance of the recency effect cannot directly be assessed with these experimental data, because the model is mathematically scale-invariant, insofar as it provides a good description of the data, this suggests the recency effect across scales in FFR is also scale-invariant. The major obstacle to this conclusion is the fact that the effective time of presentation of the items is not strictly controlled. After presentation of each list, subjects had 30 s to free-recall that list. During this period, some of the words would have been repeated as recalled items. Moreover, it is possible that words that were presented in one list were intruded during recall of later lists. Howard, Youker, and Venkatadass (2008) reported that about 9% of words that were recalled during the final free recall session were reported as intrusions during study of another list. Beyond that, we have no way of knowing how many words might have been generated as potential recalls but not spoken out loud because the subject knew that they were errors (Kahana, Dolan, Sauder, & Wingfield, 2005). Nonetheless, the fact that the model is scale-invariant and that it does a reasonable
job describing this disparate set of experimental results with no free parameters suggests that the recency effect in the probability of first recall is scale-invariant. Nairne, Neath, Serra, and Byun (1997) showed that serial position curves were not scale-invariant but did not report results for the probability of first recall. Further experimentation will be necessary to clearly establish this. Of course the observation of power law forgetting (Wixted & Ebbesen, 1997; Rubin & Wenzel, 1996) in other memory tasks is perfectly consistent with a scale-invariant mechanism.

**Contiguity.** We focus on the transition from one recalled word to the next. We consider only the contribution of the just-recalled word to $T(\tau^*)$ in calculating the probability of recalling each of the candidate words from the list. One of the notable features of the lag-CRP in free recall is the near-ubiquitous asymmetry. Across a variety of studies the lag-CRP corresponding to adjacent forward transitions is higher than the lag-CRP corresponding to adjacent backward transitions (Kahana, Howard, & Polyn, 2008). We account for this asymmetry in a way analogous to the account TCM provides.

Suppose that $f_o$, the stimulus initially presented at time $\tau_o$ is repeated (and remembered) at $\tau_r$. That is, at a particular moment $\tau_r$, we find that $f(\tau_r) = f(\tau_o)$. For the moment, let us consider the implications of the fact that the stimulus function is the same at these two moments. Immediately after repetition of the stimulus, the state of the temporal history $T(\tau_r, \tau^*)$ includes the pattern corresponding to $f_o$ at $\tau^* = 0$. Assuming that the time between stimuli is non-zero, this state of the history does not resemble the state of history for the stimuli that followed the initial presentation of $f_o$. In the period after $\tau_r$, the pattern corresponding to $f_o$ gradually drips to earlier values of $\tau^*$ and smears out as it goes. As the time after the repetition of $f_o$ approaches the time between presentation of the stimuli, the state of $T$ resembles the encoding state of $T$ for the words that followed $f_o$, providing a boost in the activation of those words. There would not be an overlap for words whose initial presentation preceded $f_o$—the pattern $f_o$ was not part of the state of $T$ when the words preceding $f_o$ were presented.

In contrast, suppose that when $f_o$ is repeated, it instantaneously recovers the state of the temporal history when it was initially presented, $T(\tau_o, \tau^*)$. Because the temporal history changes gradually, this recovered state provides a cue both for words that followed $f_o$ as well as those that preceded it. The recovered state is initially a symmetric cue that gradually skews forward in the moments after $\tau_r$.

In order to enable both of these components to contribute to the contiguity effect, we momentarily alter the differential equation controlling $t(s)$. If the item originally presented at time $\tau_o$ is recalled at time $\tau_r$, then we can write

$$\frac{dt(s)}{d\tau} = -st(s) + \gamma t(\tau_o, s) + (1 - \gamma) f(\tau_o) \quad (20)$$

The factor $\gamma$ is less than one and should be interpreted as the strength of reinstatement of the context. When operated on by $L^{-1}_k$, this gives rise to two new components of $T$. One is the repetition of the stimulus itself, weighted by $1 - \gamma$ that enters at $\tau^* = 0$ and “drips” through $T$ just as it did when the item was initially presented (except weighted by $1 - \gamma$). The other is $\gamma L^{-1}_k t(\tau_o)$. This is the state $T(\tau_o)$ that was present just as $f_o$ was originally presented. These two components are roughly analogous to the two components of retrieved
context in TCM. Again in analogy to TCM, \( \gamma \) is a free parameter that controls the mixing of these two components. Appendix C works out the details of the calculation under these circumstances.

There are two additional considerations prior to examining model predictions. First, as should be clear from the preceding discussion, the effectiveness of repeating an item as a cue for subsequent recalls changes in the moments after \( \tau_r \) when the stimulus is repeated. For simplicity, we assume that the state a non-zero time \( r \) after repetition of the word, \( T(\tau_r + r, \tau^*) \), is the cue for retrieval of the next word. Second, as discussed above, there are a great many reasons why one word would cue another in a way that does not depend on their temporal contiguity. Accordingly, we alter Eq. 17, adding a constant \( c \) to the prediction calculated above:

\[
\text{CRP}(n) = \frac{(p_n + c)^a}{\sum \ell (p_{\ell} + c)^a} \quad (21)
\]

In the illustrations of model predictions that follow, we set \( \gamma = 0.9, \tau_r = 0.75 \), corresponding to 750 ms if \( \delta \tau = 1 \) s, and \( c = 2 \times 10^{-5} \). The numerical value of \( c \) should be compared to the magnitude of the \( p_n \)s. As before, \( a = 2 \).

Figure 12 shows the lag-CRPs of the final free recall in the experiment of Unsworth (2008). Unsworth (2008) presented subjects with 10 lists of 10 words each. Each list was presented at a rate of 1 word per second followed by a 16 second distracting math task, subsequently followed by a 45 second recall period for that list. The next list was presented immediately following the end of the previous recall period. We take the time interval between the study of two successive lists as 71 seconds. During the final recall, whenever a successively recalled pair comes from the same study list, it is grouped as a within list transition. The CRP of the within-list transitions is plotted in the top left panel. Whenever the successively recalled pairs were from different lists, they are grouped as across-list transitions, meaning an across-list transition of zero is not possible in this analysis. The lag-CRP of the across-list transitions is plotted in the bottom left panel.

Note that the empirical data are not strictly scale-invariant. If they were, the within- and across-list curves would perfectly align. In addition to the change in the steepness of the curves, there is also a change in the asymmetry apparent in the data. Equation 21 is plotted on the right side of Figure 12. The model captures the qualitative aspects of the data. Both within- and across-list contiguity effects exist in both the forward and backward directions. Within-list transitions are asymmetric favoring forward transitions. Across-list transitions are more nearly symmetric and lower in amplitude. This pattern is largely attributable to the role of \( r \), the time after recall of the preceding word, on retrieval. For the within list transitions, \( r \) is similar to the time interval between list items (1 s) and we see a strong forward asymmetry. When \( r \) is small compared to the study time between items, the lag-CRP will be more nearly symmetric because the temporal history when the stimulus was initially presented will dominate. This property is a basic prediction of the model and reflects a qualitative deviation from scale-invariance. Importantly, this deviation appears to be necessary to account for the data. Although it is very difficult to definitively assess the empirical question of the asymmetry of the lag-CRP in FFR it is worth noting

\[13\] Howard, Youker, and Venkatadass (2008) conducted similar analyses of FFR data and found qualitatively similar results. However, in that experiment there was residual recency in the contiguity effect observed in FFR, so that the simplifying assumptions used here do not apply.
Figure 12. The contiguity effect persists across time scales but is not strictly scale-invariant. After a word is free-recalled, the lag-CRP estimates the probability that the next recalled word comes from neighboring positions. Within- and across-list lag-CRP were measured during final free recall of ten lists of words by focusing on transitions between words from the same list and between words from different lists. The within-list lag-CRP (top) gives lag as the difference between serial positions within the list. The across-list lag-CRP (bottom) gives the list lag for transitions between different lists as the difference in list number. Data from Unsworth (2008) is plotted on the left side. The asymmetry is decreased in the across-list lag-CRP. On the right side, the experiment is modeled using Eq. 19. See text for details.
that another study that measured the contiguity effect in final free recall also failed to measure an asymmetry in the across-list lag-CRP while measuring a reliable asymmetry in the within-list lag-CRP (Howard, Youker, & Venkatadass, 2008).

**Discussion.** We accounted for contiguity by assuming that repetition of an item recovers the previous state of \(t(s)\), which in turn causes the reinstatement of the temporal history \(T(\tau)\). The assumption that recall of \(t\) causes recall of \(T\) is not necessary to describe the PFR and lag-CRP. Mathematically, the model could predict the same results for the situations considered here if \(T\) were recovered directly. However, there are several potential advantages to recovering \(t\) to recover \(T\) rather than just recalling \(T\) directly.

The most important factor is that if \(t(\tau_0, s)\) is recovered, this enables, in principle at least, recovery of the entire internal timeline, including not only internal past time but also internal future time. If internal future time is constructed as hypothesized above, it requires \(t(\tau_0, s)\) as a starting point. If there is a genuine “jump back in time” with information about both past history and future trajectories this could give subjects the ability to retrieve information about the relative time of presentation of stimuli, rather than just a scalar measure of associatedness. For instance, in Kahana and Caplan (2002), subjects were overtrained on a serial list. They were then given items as probes and asked to recall the item that came at a specific position relative to the probe. For instance, if the list included the subsequence

\[\ldots \text{ABSENCE HOLLOW PUPIL} \ldots,\]

a probe might be

\[? \text{HOLLOW},\]

for which the correct recall would be ABSENCE. The task not only requires the subject to recover the other stimuli associated with HOLLOW, but to respond in a way that expresses an understanding of their temporal relationships. Jones and Pashler (2007) showed evidence for explicit awareness of temporal relationships after learning a presumably implicit statistical learning task. Although HOLLOW is associated to both ABSENCE and PUPIL, and one might be more likely to free recall PUPIL than to free recall ABSENCE after recall of HOLLOW, the representation of internal time can nonetheless discriminate the temporal relationships between these stimuli. After recall of HOLLOW and recovery of its prior state of temporal history, ABSENCE is active in the temporal history whereas PUPIL is not.

Similarly, imagine an analog of the free recall task in which one is asked to start recall with an item from a particular point in time. For instance, suppose recall initiation is post-cued and the instruction to the subject is to recall the item presented 1 or 10 or 100 s ago. It seems very likely that, when presented with these different instructions, subjects would tend to recall different items (e.g. Dalezman, 1976). If one just probed with the end-of-list temporal context, it would be very difficult to account for different recalls given the different instructions without surreptitiously recalling the intermediate items. In contrast, if one could respond to the cue by differentially activating a specific value (or range) of \(\tau\) and using only that as a probe, different recalls in the different conditions could be readily accounted for.

Modeling of recency in the PFR and contiguity as measured by the lag-CRP is a requirement for a free recall model, but it does not constitute a complete model of the free
recall task. Using the temporal history $T(\tau)$ as a generalization of temporal context does not sacrifice the ability to describe recency or contiguity effects, showing performance at least comparable to the first presentation of TCM (Howard & Kahana, 2002). Although the devil is often in the details, it ought to be possible to account for similarly extensive set of empirical findings from free recall and related memory tasks (Sederberg et al., 2008; Howard, Jing, Rao, Provyn, & Datey, 2009; Polyn et al., 2009; Sederberg, Gershman, Polyn, & Norman, 2011) using this generalization as an associative engine. The present approach also extends TCM in two important ways. First, it is possible to account for genuinely scale-invariant recency effects. Second, the temporal history has autonomous dynamics that follow retrieval of a word. This induces a potentially rich set of retrieval dynamics in the retrieval cue itself.

After presentation of an item during study, the state of the temporal history changes autonomously over time as activity “drips” through subsequent values of $\tau$. When an item is repeated, or a state of $T(\tau)$ is recovered, the dynamics take $T(\tau)$ through a similar sequence of states. Autonomous retrieval dynamics are an essential component of several important models of serial recall (e.g., Burgess & Hitch, 1999; Brown et al., 2000; Farrell & Lewandowsky, 2002). These models are often contrasted with chaining models of serial recall, in which recalling an item causes a change in the memory state for retrieval of the next item (e.g., Lewandowsky & Murdock, 1989).

The present model offers an interesting hybrid of autonomous dynamics and chaining. Consider the change in the state of temporal history during learning of a serial list. If it were the case that the initial state of temporal history were somehow recovered at test, then it would drift through a set of states even if there were no subsequent recalls. That is, immediately after the initial state is recovered, it would be an optimal cue for recall of the first list item. After some time passes, however, the initial state more closely resembles the encoding state of the second item, even if the first list item is not successfully recovered. The states that are autonomously visited will overlap sequentially with the encoding states of the list items, giving rise to sequential recall with a primacy gradient. If, in addition to the autonomous dynamics, the list items were also recovered, this would supplement recall like in a chaining model.

There are two ways that recovering an item would help recall of subsequent items, both of which can be thought of as chaining. First, the item itself would contribute to the temporal history, providing a better retrieval cue for subsequent items (Kahana, Mollison, & Addis, 2010). Second, if recovered items can also cause recovery of their states of temporal history, then this would also contribute to the contiguity effect. In fact the contiguity effect seen in serial recall resembles to a large extent the contiguity effect seen in free recall (Bhatarah, Ward, & Tan, 2008; Klein, Addis, & Kahana, 2005; Howard, Addis, et al., 2007; Solway, Murdock, & Kahana, 2012 but see Farrell & Lewandowsky, 2004). The hybrid nature of the temporal representation, incorporating aspects of both chaining models and non-chaining models could make it well-suited for describing similarities between free recall and serial recall (Bhatarah, Ward, Smith, & Hayes, 2009; Ward, Tan, & Grenfell-Essam, 2010).
Predicting future temporal trajectories

The mathematical framework provides a model for both retention, our estimate of the stimulus history available to us at the present moment and protention, our momentary prediction of the events that will happen in the future. The behavioral applications up to this point have not made use of the representation of internal future time. This application and the next one rely heavily on the temporal representation of internal future time.

Gibbon and Church (1981) trained pigeons to respond based on their expectation for the delay to future rewards. An optimally-behaving subject in this time-left procedure should choose the stimulus that predicts reward a shorter distance in the future. The correct answer, however, changes over time. At each time considered, the subject must update their prediction about the relative time of future rewards. The basic idea of the model is that subjects “play forward” future events using the translation operator with successively larger values of $\delta$. In this way, they can select the stimulus that gives rise to a future trajectory that results in reward earlier. In order to do this, we need to calculate the time until the peak of a reward function in future time $\delta$. As a starting point for this calculation, we start by considering how the model accounts for a simpler task, temporal reproduction, in which the subject can respond based on the prediction available at the present moment.

Non-veridical timing in temporal reproduction. In a typical temporal reproduction task, the subject is presented with, say, a tone for some duration. After a brief interval, the tone starts again. The subject is instructed to press a button to terminate the tone when its duration matches that of the study duration. One can simply model performance in this task by assuming that during reproduction, the subject attempts to predict the peak corresponding to the end of the original stimulus.

When the interval is learned, the subject experiences the end of the interval $\tau_D$ seconds after some kind of start signal at the beginning of the interval. This relationship could be learned by storing the state of $T(\hat{\tau})$ at the time when the interval ended in $M$ (Eq. 6). The relevant part of this encoded $T(\hat{\tau})$ is the column corresponding to the signal starting the interval. This column contains the pattern caused by the start signal presented $\tau_D$ seconds in the past. When the subject tries to reproduce the interval, the start signal enters $T(\hat{\tau})$ at $\hat{\tau} = 0$ and gradually drips down the column as time passes. We propose that to decide when to end the interval, the subject observes the degree to which the end signal is predicted as the temporal history changes and responds when the prediction reaches a maximum. During reproduction, let’s call the time at which the start signal was presented time zero and write down the function for the prediction of the end signal $\tau$ seconds after repetition of the start signal. This function can be readily calculated (see Shankar & Howard, 2012) and is given by

$$p(\tau) = \frac{1}{\tau_D^2} \frac{y^k}{(1 + y)^{2k+2}},$$

(22)

where $y \equiv \tau/\tau_D$. Differentiating reveals that the peak of this function, which gives the time

---

14This expression differs from an analogous expression in Shankar & Howard (2011) because in that paper we kept $g(\hat{\tau}) \propto 1$ for simplicity.
Figure 13. Nonveridical temporal reproduction. Across a variety of experiments, reproduced durations go up linearly with actual duration but with a slope reliably less than unity. a. Eisler (1975) presented subjects with a tone for some duration. After a brief (300 ms) delay, the tone started again. Subjects were instructed to press a button to end the tone at the moment that would yield the same duration. Mean reproduced duration as a function of experienced duration. Perfect performance would correspond to the thick black line along the diagonal. The thin black line is a linear regression to the data points, with intercept .54 s and slope of .61 ± .02 ($R^2 = .992$). b. Median reproduction times reported by Kowalski (1943). This reproduction task used a light rather than a tone and varied the delay between the standard and the reproduction. The small grey symbols show results from the shortest delay (2.5 s), the black symbols show the results averaged across all delays. The thin black line shows the results of a linear regression, which has an intercept of .13 s and a slope of .87 ± .01 ($R^2 = .9994$). Note that the reproductions are slightly long for short intervals, but are short for longer intervals. c. Lewis & Miall (2009) used visual presentation intervals of a very wide range of durations. Subjects initiated the interval and the reproduction. The thin black line shows the result of a linear regression with intercept 6.6 s (note the very large range of durations in this experiment) and a slope of .61 ± .01 ($R^2 = .994$).

at which the subject responds, is at

$$\tau_R = \frac{k}{k + 2} \tau_D.$$  \hspace{1cm} (23)

There are a couple of things to be observed about this equation. First, the reproduced duration $\tau_R$ is linearly related to the actual duration $\tau_D$. Second, the reproductions are systematically shorter than the actual intervals. For all valid $k$, $\frac{k}{k + 2}$ must be less than 1. For $k = 4$, which we have been using throughout, the coefficient is $\frac{2}{3}$.

It may seem like a serious problem that the model predicts a systematic underestimation of experienced durations in the temporal reproduction task. However, this prediction appears to be consistent with a wide variety of empirical findings from the temporal reproduction task, at least when feedback about subjects’ performance is not possible. Figure 13 shows results from three large experiments in which subjects reproduced intervals of varying durations. The y-axis gives the reproduced duration, $\tau_R$ as a function of the experienced duration, $\tau_D$, on the x-axis. In all three cases, the data are well-described by a linear function with a slope reliably less than one and an intercept close to, but reliably larger than, zero (Eisler, 1975; Kowalski, 1943; Lewis & Miall, 2009). For two of the three studies (Eisler, 1975; Lewis & Miall, 2009), the slope is very close to the value one would expect for
The non-zero intercept is what one would expect from a motor reaction time that is independent of the duration but relatively large for short intervals (for instance, in the Lewis & Miall, 2009 experiment, the shortest duration to be reproduced was 68 ms).

This account of temporal reproduction data suggested by the present modeling work would require more extensive empirical examination before it can be definitively accepted. However, it seems to be the case that when durations are randomized across trials, the result is non-veridical performance with the pattern predicted by the model—a linear function with a slope considerably less than 1 (Carlson & Feinberg, 1968, 1970; Eisler, 1975; Kowalski, 1943; Lewis & Miall, 2009; Ulbrich, Churan, Fink, & Wittmann, 2007). In contrast, experiments that use blocked trials of the same duration seem to show more nearly veridical performance with a slope very close to 1 (Treisman, 1963; Woodrow, 1934; for an explicit comparison of the effects of random vs blocked temporal reproduction see Doehring, 1961). Note that if the durations are blocked, the procedure gives the subject an opportunity for feedback. In a blocked group of, say, 50 trials with the same duration, the subject can predict the termination of the interval during study and use this information to calibrate his or her performance. Indeed, Treisman (1963) described a “lengthening effect” in which reproduced durations increase systematically within a block. This is perfectly consistent with a timing mechanism that times early in the absence of feedback and gradually lengthens. This brief review of data from the temporal reproduction task suggests that a scale-invariant model that consistently times early provides a viable account of temporal reproduction data.

Linear time in the time-left procedure. Rather than predicting an event at the time it is expected to occur, as in the temporal reproduction task, the time-left procedure requires subjects to predict how far in the future an event will occur. Critically for our purposes, the judgment is made long before the event, forcing the subject to rely on predictions of the future.\footnote{Note that the time-left procedure is distinct from what is termed prospective time estimation. Prospective time estimation is used to refer to tasks in which the subject knows that an interval will need to be timed as it is occurring. This term is used in contrast to retrospective time estimation in which the subject is not explicitly made aware of the need to time an interval as it is happening (e.g., Taatgen, van Rijn, & Anderson, 2007). The key point is that prospective memory experiments do not necessarily rely on protention because the judgment or response is made at the moment the interval would end. In the time-left procedure, the judgment is made about an event that is still in the future when the response is made.}

Figure 14 illustrates the basic idea of the time-left procedure used by Gibbon and Church (1981). Pigeons were first trained that pecks to a white key resulted in the availability of food after a long delay, \(C\); pecks to a green key resulted in the availability of reward after a short delay, \(C/2\), that was half of the duration of the delay from the white key. The keys were in different locations within the pigeon box; these locations were consistent throughout the experiment. After initial training, the subjects were trained in time-left trials. At the beginning of each trial, the subject was able to peck at the white key in its regular location or a red key that was illuminated in the same location as the green key.\footnote{It is possible that the difference is due to a cohort effect—the slope in temporal reproduction appears to be affected by working memory capacity (Broadway & Engle, 2011).}
Figure 14. Schematic of the time-left procedure (Gibbon & Church, 1981). At the beginning of a trial, the animal is presented with the long stimulus that will result in reward being available after some delay $C$ relative to the start of the trial. A second key gives the animal the option to switch to the short stimulus that will result in reward a delay $C/2$ after the choice point. After some time $\tau_d$ has passed, the next peck makes the choice between sticking with the long stimulus, finishing the delay $C$, or switching to the short stimulus that will result in reward after a delay $C/2$. The value of $\tau_d$ changes across trials and is not known to the animal on any given trial. Whether it is advantageous to switch depends on the duration $\tau_d$ that has elapsed prior to the choice.

After a delay of $\tau_d$ seconds, the next peck to either of the two keys resulted in one of two outcomes. A peck to the white key darkened the red key. As in initial training, reward was provided for pecks to the white key $C$ seconds after the trial began. If the first peck after $\tau_d$ seconds was to the red key, the red key turned green and the white key was darkened. After an additional $C/2$ seconds elapsed, a subsequent peck to the green key resulted in reinforcement. If the subject had perfect knowledge of the delay intervals, the optimal strategy would be to peck at the red key until $C/2$ into the trial and then switch to the white key (Fig. 14). When the time left until reward of the white key, $C - \tau_d$ is less than the delay associated with the green key, $C/2$, the pigeon should pick the long stimulus (white key). That is, the subject should switch from pecking the red key to pecking the white key when $C - \tau_d = \frac{C}{2}$.

As it turns out, subjects did not quite follow this optimal relationship. Instead, Gibbon and Church (1981) observed that the point of indifference, i.e., the value of $\tau_d$ where the subject is equally likely to peck the red key and the white key, comes a bit earlier than it would if performance were optimal. The indifference point was observed when the time into the trial when the switch is offered, $\tau_d$, was a bit less than $\frac{C}{2}$. In different sessions, Gibbon and Church (1981) manipulated the overall scale of the delays by changing the value of $C$. Critically, the point of indifference went up linearly with the magnitude of $C$. They argued that this rules out models in which the time left is estimated on a logarithmic scale. It will turn out that both properties, the finding that the indifference point $\tau_{1/2}$ is less than
C/2 and the linear increase of \( \tau_{1/2} \) with increases in \( C \), are natural predictions of the model of protention proposed here.

We assume that when presented with the choice of stimuli \( \tau_d \) after presentation of the long stimulus, subjects sample each stimulus (white key and green key) and then “play forward” the future from the present moment by operating on \( T(\tau_d, \delta) \) with \( R^\delta \) with successively larger values of \( \delta \). While examining future time, they get information about when reward is predicted to become available. There is some point in internal future time where the prediction reaches a maximum. The value of \( \delta \) where this future peak is observed for each stimulus is extracted and used to inform the decision about which stimulus is preferable. If the decision is unbiased, the point of indifference should come when the perceived time to stimulus reinforcement for the two choice stimuli is equally far in internal future time. That is, we are interested in the value of the temporal delay, \( \tau_d \) at which the peaks of the two stimulus functions are aligned in internal future time.\(^{17}\) Figure 15 illustrates the location of the peaks in future time at various delays points within the interval.

Now we want to calculate the value of internal future time \( \delta \) at which the prediction of reward peaks for each of the stimuli. Because the model is scale-invariant, we can keep everything in units of \( C \). We will refer to the stimulus presented at the beginning of the trial, and trained with the reinforcement interval \( C \) as the long stimulus. The green key, which was trained at a delay to reinforcement of \( C/2 \) will be referred to as the short stimulus. The relevant predictions can be easily computed using Eq. 22. First, the peak of the short stimulus will be observed at

\[
\delta_{p,s} = \frac{k}{k+2} \frac{C}{2} \quad (24)
\]

In contrast, because the prediction function for the long stimulus has been advancing for \( \tau_d \) seconds (see Figure 14), its peak in imagined future time will be at

\[
\delta_{p,l} = \frac{k}{k+2} C - \tau_d \quad (25)
\]

The point of indifference, \( \tau_{1/2} \), is just the value of \( \tau_d \) where \( \delta_{p,s} = \delta_{p,l} \), i.e.,

\[
\tau_{1/2} = \frac{k}{k+2} \frac{C}{2}. \quad (26)
\]

The model predicts that the indifference point should be a linear function of the short interval \( C/2 \) with a slope less than one. The slope is controlled by \( k \). When \( k = 4 \), the slope should be \( \frac{2}{3} \). Larger values of \( k \) should yield slopes closer to one. Figure 16b shows this function plotted over the range of values of \( \frac{C}{2} \) used in the Gibbon and Church (1981) experiment for a variety of values of \( k \).

\(^{17}\)One can ask about the rate of scanning through different values of \( \delta \). There is no reason to assume that subjects scan linearly through future time. The rate of scanning for \( \delta \) is somewhat analogous to \( g(\hat{\tau}) \). The change in the density of states of \( \delta \) in internal future time is also analogous to the temporal discounting function employed in studies of decision-making. It is unavoidable that predictions about the near future should be more accurate than for values of \( \delta \) further in the future. In analogy to the arguments about the density \( g(\hat{\tau}) \), why would we invest as much time predicting adjacent values of \( \delta \) for distant events where the prediction is changing slowly as a function of \( \delta \) as we would for events near in the future where the prediction changes rapidly as a function of \( \delta? \) Happily, we don’t need to specify the scanning rate for \( \delta \) in order to calculate the indifference point—the time at which the predicted peaks occur at the same value of \( \delta \). All that is necessary is that whatever distribution \( \delta \) has, it is the same for both trajectories.
Figure 15. The model’s account of the time-left procedure. Each panel shows the prediction function in internal future time $\delta$ for reward associated with each stimulus. In each panel, the prediction for the long stimulus (which predicts a reinforced keypeck after a time $C$) is shown as a heavy grey curve. The prediction for the short stimulus (which predicts a reinforced keypeck after a time $C/2$) is shown as a lighter grey curve. The curves for the short stimulus are more peaked than for the long stimulus because the stimulus predicts reward sooner in the future for the short stimulus, resulting in a tighter peak. Vertical lines show the locations of the peaks for the corresponding curves. The panels differ in $\tau_d$, the amount of time after the long stimulus was presented when the short stimulus was presented. Note that the curve for the long stimulus shifts backward as $\tau_d$ increases. 

a. The prediction if the choices are presented simultaneously, $\tau_d = 0$. The prediction for the short stimulus is much closer in internal future time than the long stimulus.

b. For a non-zero $\tau_d$ less than the critical value, the peak of the short stimulus is still earlier than the long stimulus. 

c. At the point of indifference, the peaks of the prediction curves align. This happens when $\tau_d = \tau_{1/2} = \frac{k}{k+2} \frac{C}{2}$. 

d. When $\tau_d$ is greater than the point of indifference, the peak of the long stimulus happens earlier in predicted future time than the peak for the short stimulus.
Figure 16. Results from the time-left procedure. The time-left procedure yields a linear relationship between the indifference point, the delay at which the subject chooses each stimulus equally often and the optimal switch point, $C/2$. a. Data from Gibbon and Church (1981). The heavy line is the diagonal. The lighter line is the result of a linear regression fit to the four points. The intercept was not significantly different from zero. The slope was less than one, $.76 \pm .08$ ($R^2 = .965$). b. Predictions from the model of protention, Eq. 26. The lower dark line is for $k = 4$. The grey lines are for progressively larger values $k$, up to $k = 12$.

The prediction of a linear relationship holds for all possible values as long as $g(\tau^*)$ is a power function. The slope is less than one for all finite values of $k$. The linear relationship between the indifference point and $C/2$ may seem like it is contradictory to the logarithmic mapping of physical time onto the cells representing $\tau$ implied by setting $g(\tau^*) \propto \tau^{-1}$. Indeed Gibbon and Church (1981) argued that their results contradict a logarithmic representation of future time. This model produces a linear function because the peak of the prediction moves forward linearly in physical time (Eq. 23). The model takes the match of one state of $T(\tau^*)$ (the encoding state) with another (the probe state). Each of those states is distributed logarithmically in internal past time (note the $g(\tau^*)$ in Eq. 9), but the match between them does not change logarithmically in physical time (Eq. 22).

Discussion. Gibbon, Church, Fairhurst, and Kacelnik (1988) suggested that their non-optimal result, i.e., the slope less than unity in Fig. 16a, could be due to an asymmetry between the way different types of errors are valued. One could also attribute the difference to an asymmetry between the two stimuli—for the long stimulus, the same key that is rewarded is available during the decision phase whereas for the short stimulus the animal must peck at a red key to cause the arrival of the subsequently rewarded green key. Perhaps this asymmetry introduces a bias away from the short stimulus. The present model predicts a slope less than one in Figure 16 for a different reason. According to the model, the slope less than unity happens because the peak in the protention function occurs before the actual time of reinforcement. The slope is less than one precisely because the model does
not predict veridical timing. This is consistent with results from temporal reproduction experiments when the intervals are randomized across trials (Fig. 13). We suggest that veridical timing is observed only in temporal reproduction tasks in which the subject is provided with feedback about their predictions.

This situation is consistent with the results from the peak interval procedure, in which subjects respond around the time they expect a practiced interval to end (e.g., Roberts, 1981; Rakitin et al., 1998). During training in the animal variant of the task, the first response to a lever press or a key peck is reinforced after a certain interval has elapsed since the previous reinforcement. On peak interval trials, the reinforcement is withheld and the distribution of responses is noted. In the human variant, there are a large number of training intervals that are studied without responding. Then on reproduction trials, the subject is asked to respond in such a way that the responses bracket the end of the trained interval. Peak responding in the peak interval procedure is generally reported to be at the right time, without systematic errors (e.g., Gallistel & Gibbon, 2000; Rakitin et al., 1998).

However, the learning experience in the peak interval procedure provides ample opportunity for feedback. In the human procedure even if there is not explicit feedback about the timing and distribution of the responses (which is often the case—e.g., Exp. 1, Rakitin et al., 1998) the subject can surreptitiously guess at the end of the interval during training trials and thus obtain feedback. In the animal variant of the task, there is also explicit and asymmetric feedback on the training trials. Responses made prior to the end of the fixed interval are not reinforced whereas the first response after the interval is reinforced.

The present model predicts that in the absence of feedback the reproduction of temporal intervals is non-veridical, with a systematic tendency to time early. We argue, somewhat paradoxically, that non-veridical timing is actually adaptive and reflects an asymmetric cost to timing errors. Consider the problem faced by a learner trying to predict some biologically relevant event, say the appearance of a predator, a certain time after a predictive stimulus. If the learner responds too early, this is perhaps suboptimal, but nearly as costly as if the learner predicts the appearance of the predator too late. This suggests that non-veridical timing such that events are predicted early is actually an adaptive feature.

Temporal mapping by anticipating the future

In temporal mapping experiments (e.g., Cole et al., 1995; Barnet, Cole, & Miller, 1997; Savastano & Miller, 1998; Arcediano & Miller, 2002; Arcediano, Escobar, & Miller, 2003), subjects both learn temporal relationships between pairs of stimuli and integrate those relationships into a coherent map that generalizes temporal relationships among stimuli that were not experienced together in time. In response to these results Miller and colleagues have developed the temporal encoding hypothesis (Arcediano & Miller, 2002; Matzel et al., 1988; Savastano & Miller, 1998) to describe learning in conditioning experiments. The temporal encoding hypothesis has two components that are particularly relevant for the model of internal time developed here. First, they argue that the temporal relations between stimuli forms an essential and unavoidable part of the learning event. Second, they argue that learners can integrate disparate learning events into a coherent temporal map by aligning different time lines on a common stimulus. We model core results from the temporal mapping paradigm using both protention and the ability to jump back in time to a previous state of history. Jumping back in time enables the integration of distinct learning episodes;
protention enables a conditioned response that is sensitive to the future occurrence of the US.

To make this more concrete, let us describe a specific experiment (Cole et al., 1995). Rats were trained to associate a 5 s CS1 with a US (shock). In one condition, the time between offset of the CS1 and the onset of the US was 0 s (Figure 17a, top). In the other condition, the time between the offset of CS1 and the US was 5 s (Figure 17a, bottom). Let us refer to these as the 0 s and 5 s conditions, respectively. After training the CS1-US association, a second-order association was formed between CS1 and another 5 s CS2. In both conditions, the onset of CS2 immediately followed the offset of CS1 (Figure 17). In neither condition did CS2 ever cooccur with the US. The first finding was, not surprisingly, that the CR to the CS1 was stronger in the 0 s condition than in the 5 s condition. If the relationships learned between the stimuli were atomic associations, we would expect the second order conditioning to CS2 would also be stronger in the 0 s condition than in the 5 s condition. After all, the association from CS2 to the US must be mediated by the association from CS1 to the US. However, exactly the opposite was observed. This result makes no sense from the perspective of simple associative strength. The temporal coding hypothesis (Matzel et al., 1988; Savastano & Miller, 1998) of Miller and colleagues reconciles these findings as follows. Note that if the two learning episodes were aligned on the CS1 (as in Figure 17c), then the CS2 would not predict the onset of the US in the 0 s condition. In the 5 s condition, CS2 strongly predicts the onset of the US when the two learning episodes are aligned on presentation of CS1.

The present framework aligns closely with the temporal encoding hypothesis, providing a mechanistic implementation of it. A mechanistic model must have two basic properties in order to account for this phenomenon. One is that the temporal relationships between stimuli, rather than a simple scalar associative strength, is learned. Second, some mechanism for integrating disparate episodes into a coherent synthetic representation is necessary. The representation of internal past and future time satisfies the first constraint. The ability to retrieve temporal contexts satisfies the second constraint (see also Howard, Fotedar, Datey, & Hasselmo, 2005; Rao & Howard, 2008).
It has been argued that jumping back in time by recovering a gradually-changing representation could serve the purpose of forming associations between items that were not presented together in time, but that were presented in similar temporal contexts. For instance, if subjects learn an association between two stimuli, e.g., A-B and then at a much later time learn B-C, subjects form associations between A and C despite the fact that they were never experienced close together in time (Primoff, 1938; Slamecka, 1976; Bunsey & Eichenbaum, 1996; Howard, Jing, et al., 2009). There is evidence suggesting that these bridging associations are not dependent on chaining across the pairs. That is, the association from A to C does not depend on surreptitiously recalling B as an intermediate step (Bunsey & Eichenbaum, 1996; Howard, Jing, et al., 2009). The jump back in time hypothesized to underlie the contiguity effect in episodic memory can also account for these transitive associations (Howard et al., 2005; Howard, Jing, et al., 2009). For instance if repeating B as part of the pair B-C recovers information from its previous presentation, i.e., information about A, then it is as if the temporal context of C is close in time to the temporal context of A. In this sense, jumping back in time enables learning to integrate different training episodes (see also Rao & Howard, 2008). If the learner were able to jump back to recover a genuinely temporal representation, this could enable an account of temporal mapping phenomena.

In the present framework when CS2 is given as a probe, the state of \( T(\tau) \) it recovers doesn't just provide a general sense that it has something to do with CS1. Rather, it recovers information specifying that CS1 was presented 5 s in the past when CS2 was presented. Given that state, the animal plays forward the future trajectory by manipulating \( \delta \) searching for a peak corresponding to the arrival of a shock. The animal is frightened by CS2 to an extent proportional to the magnitude of the predicted peak divided by the time in the future it will occur.

Let us consider the prediction trajectories going forward in time in each of the four situations corresponding to presentation of CS1 or CS2 in either the 0 s or 5 s conditions (see Figure 18). In the 0 s condition, presentation of CS1 predicts an imminent peak. In the 5 s condition, CS1 predicts a peak further in the future. This accounts for the finding of less fear to CS1 after learning in the 5 s condition compared to the 0 s condition. Now let us consider the effect of presenting CS2 after learning in the two conditions. In both conditions, CS2 recovers a state with CS1 5 s in the past. In the 0 s condition, this state generates a monotonically-decreasing prediction function for the US. In this case, the peak detector will never find a peak going forward, placing the time to the peak at infinity. However in the 5 s condition, the state of \( T \) with CS1 5 s in the past is an excellent predictor of an imminent US, much like CS1 was in the 0 s condition. The magnitude is less than that caused by CS1 in the 0 s condition because CS2 does not perfectly recover its prior context. This account is sufficient to describe the degree of freezing elicited by both stimuli in both conditions.

Figure 19 shows concrete model predictions. For this figure we set \( \gamma \), which controls the amount of recovery of the temporal context vectors and thus the stimulus history (see Eq. 20), to 0.8 and calculated the time to the peak of the US predicted by each stimulus. The amount of fear caused by a stimulus was taken as the magnitude of that peak divided by the time to its instance. In this experiment the dependent measure is the cumulative time taken to drink water for five seconds. The conditioned response is manifest as an extension in the time taken to drink due to freezing. Because the behavioral dependent
Figure 18. Schematic illustration of the prediction function for the US going forward in time after the onset of CS1 or CS2 in the two conditions described in Figure 17. The top row shows the prediction following CS1 (trace, TR). The bottom row shows the prediction following CS2 (second order conditioning, SOC). The left column shows the predictions that obtain after both phases of learning in the 0 s condition (top row of Figure 17); the right column shows the predictions for the 5 s condition (bottom row of Figure 17). In each condition, the magnitude of the peak and the time to reach the peak are indicated by the vertical and horizontal double headed arrows. In the second order conditions, even though CS2 was not explicitly paired with the US, the prediction for the US is nonzero because it recovers a temporal history that includes CS1 5 s in the past, the delay between the onsets of CS1 and CS2 during training (see Figure 17b). The prediction following CS2 is the same as the prediction following CS1 with a time shift of 5 seconds (for this schematic figure, we have assumed perfect reinstatement, that is $\gamma = 1$). Note that there is no peak going forward from presentation of CS2 in the SOC-0 condition.
Figure 19. Results from a temporal mapping experiment. Left: Data from Cole, Barnet & Miller (1995). The CR after training to CS1 (the TR conditions, black bars) or CS2 (the SOC conditions, light bars) separated by the condition (0 or 5), corresponding to the delay in seconds between the CS1 and the US in the first phase of learning (see Figure 19a). The CR in each condition is the log of the the mean time taken by the rats to drink water for five cumulative seconds. Fear of the shock US disrupts drinking. Right: The height of the peak (see Figure 18) divided by the time taken to play forward to the peak of the prediction is taken as a heuristic measure of conditioned response generated by the model. We have used $\gamma = 0.8$ and added a small constant to account for the fact that the CR cannot be zero.

measure cannot be zero, we added a small constant to the ratios in constructing the results in Figure 19.

The qualitative pattern across conditions does not depend at all on the choice of parameters. For all non-pathological choice of parameters, there is a greater CR to the CS1 in the 0 s condition than in the 5 s condition. Similarly, all non-pathological choices of parameters with $\gamma > 0$ result in the reversal across conditions for CS2. For all values of $\gamma$ greater than zero and less than one, the level of the CR to the CS2 (light bars) is reduced relative to the CS1 (dark bars).

Discussion. Temporal mapping is not limited to the findings from the Cole et al. (1995) experiment. For instance, if the order of the training phases is reversed, training the relationship between CS1 and CS2 prior to learning the relationship between CS1 and the US, the same results are obtained (Experiment 2, Cole et al., 1995). The framework described here would predict the same outcome even if the stages of training were reversed.
Other variants of the paradigm (Arcediano et al., 2003; Arcediano, Escobar, & Miller, 2005; Barnet et al., 1997) may be considerably more challenging, requiring a significant elaboration of the model. For instance, in Experiment 2 of Arcediano et al. (2003), rats first learned that CS2 preceded CS1 by five seconds. In a second phase of learning, they learned that the US preceded CS1 by four seconds. Integrating the phases of learning by aligning on CS1, the only element common to both phases of learning, would result in the offset of CS2 immediately preceding the US onset. Indeed, there was robust conditioned responding to CS2 relative to control conditions. Because the US is never preceded by either of the conditioned stimuli, this experiment constitutes a challenge to mechanistic models that try to predict the US. However, this experiment can be addressed in this framework if one is willing to assume off-line integration of the learning episodes. Suppose that CS2 can cause recovery of both of the temporal histories in which it was experienced. That is, remembering CS2 recovers a state of temporal history that includes both CS1 five seconds in the past and the US four seconds in the past. If the subject can then scan through this history in the forward direction, CS2 could be remembered prior to remembering the US. Because scanning through memory preserves temporal relationships between remembered events, it can serve as a proxy for actual experiences, allowing distinct events to be integrated into a coherent map of temporal relationships among stimuli.

While the temporal representation is essential to account for these data, it is possible to fashion an account of the Cole et al. (1995) experiment without appeal to a jump back in time that recovers a prior state of the temporal history. For instance, if the subject were able to surreptitiously retrieve the US at the appropriate time during the second order conditioning (Figure 17b), then if the animal could use this surreptitious retrieval to associate the CS2 to the US at an appropriate time one could account for the bridging between learning episodes.

There is no result from the Cole et al. (1995) experiment that rules out this alternative account of this experiment. However, in the case of transitive associations in double-function lists, there are several pieces of evidence that argue against surreptitious retrievals as a general account of integrating distinct learning episodes. For instance, Bunsey and Eichenbaum (1996) did an experiment in which rats learned to predict reward among odor stimuli. Rats initially learned that odor A predicted reward of odor B over odor Y. In a separate phase of the experiment, they learned that odor B predicted the reward of odor C over odor Z. When presented with A and a choice between C and Z, normal rats preferred C over Z even though A and C were never explicitly paired. Rats with hippocampal lesions, were able to learn each of the pairwise associations, A goes with B and B goes with C as rapidly as control animals. However, hippocampal lesions prevented animals from bridging across these experiences and exhibiting an association between A and C. Retrieved temporal context can account for the dissociation between the intact pairwise associations and the transitive association that bridges across episodes. It is worth noting that in the temporal mapping experiment, hippocampal lesions would also be expected to eliminate pairwise trace associations (Solomon, Vander Schaaf, Thompson, & Weisz, 1986; Clark & Squire, 1998). If the bridging across episodes manifest in the temporal mapping experiments results from the same process that enables bridging across episodes in double function lists, then this suggests that surreptitious retrieval is not an attractive option (see also Howard, Jing, et al., 2009).
Neurophysiological evidence for a representation of internal time

We’ve just reviewed four behavioral applications constructed using the representation of internal time from four relatively distinct fields of cognitive psychology. If a representation of internally-experienced time is at the center of a wide variety of cognitive functions, then we would expect to see signatures of this representation in a wide variety of brain regions. There are three primary features hypothesized by the current mathematical framework to represent internal time that we will discuss here. First, the model predicts long-range correlations driven by the to-be-remembered stimuli. Second, it should be possible to recover states of this gradually-changing representation. The jump-back-in-time was essential for accounting for the contiguity effect in episodic memory as well as the integration of learning episodes in the temporal mapping experiments. Third, the framework requires not only that the representation change over time in a way that is sensitive to the stimuli. It also requires that the representation contain explicit information about when stimuli were experienced. This history consists of a set of cells that fire to their preferred stimulus at a characteristic lag after the stimulus was experienced.

Rather than attempting to argue for the neural plausibility of these computations by making circuit-level arguments, here we look for direct evidence of the product of these computations in the brain. Of course, it would remain possible that the representations are constructed using a different computation, but from a cognitive perspective, that is of secondary importance. While the story is still quite incomplete, there is a growing body of evidence that provides at least some support for each of these properties in the brain.

Neural evidence for long time correlations. One of the core motivations for the mathematical model of internal time presented here is that internal time should be scale-invariant. There is evidence that neurons from a variety of regions show scale-free behavior. For instance, recordings from visual cortex while monkeys watched natural videos (R. Baddeley et al., 1997) and recordings from auditory nerve in anesthetized cats (Teich, 1989) show long-range correlations in the interspike interval. In humans, units recorded from the hippocampus and amygdala of epileptic patients (Bhattacharya, Edwards, Mamelak, & Schuman, 2005) as well as voxels in fMRI experiments (Thurnera, Windischberger, Moser, Walla, & Barth, 2003) show long-range time correlations. This scale-free behavior does not, however, imply that active maintenance of information in memory is scale invariant. Scale-free behavior is widely observed in a variety of physiological (see Goldberger et al., 2002, for a review) systems, but no one would argue that heart rate stores information about recent stimuli to support performance in the JOR task. At minimum, we would like to demonstrate that slow changes in neural activity reflect some correlation with memory performance and the to-be-remembered stimuli.

The vast majority of neurophysiology and neuroimaging experiments are designed to look for neural correlations on a specific scale—typically on the order of hundreds of milliseconds to a few seconds—and ignore the effect of previous stimuli. There is thus ample evidence that to-be-remembered stimuli cause changes in brain activity over the short time scales most experiments are designed to examine. If our hypothesis is correct and correlations extending thousands of seconds are important in cognition, they would be invisible to most conventional neurophysiological experiments. Rather than trying to directly establish scale-invariance at this stage, let us first just try to establish that there is also a meaningful
neural signal supporting memory at long time scales. Note that $t(s)$, for small values of $s$, and $T(\tau)$, for large values of $\tau$, should both show evidence for autocorrelation over long time scales. For instance, a cell representing a particular value of $\tau$ should persist in firing for a duration proportional to $\tau$, leading to autocorrelation over the same scale. A few studies have examined long-range correlations in neural activity. These studies have shown strong evidence that activity is autocorrelated up to at least a few thousand seconds in brain regions believed to be important in various forms of memory.

The prefrontal cortex (PFC) is a brain region that has long been implicated in JOR (e.g., Milner, Corsi, & Leonard, 1991; Chiba, Kesner, & Reynolds, 1994). Recent evidence suggests that JOR performance is supported by a PFC representation that changes gradually over time. Jenkins and Ranganath (2010) observed BOLD response using fMRI while subjects encoded a set of stimuli for a subsequent absolute JOR task. They found that activity in the MTL, including the hippocampus during study predicted subjects' later accuracy on the task. Regions in the PFC also predicted performance on a JOR task. For the PFC regions, the compared the multivoxel pattern of activity during study of the initial presentation of the probe and neighboring items within the list. They found that the pattern of activity in PFC regions changed gradually across presentations of the list items. Moreover, the similarity of the pattern in one region, the rostrolateral PFC (RLPFC), predicted subsequent performance on the JOR task. That is, when the pattern changed rapidly in the RLPFC during study, performance on the JOR task was better than when the pattern did not change as dramatically.

The foregoing study shows that PFC activity changes gradually over a few seconds and is correlated with performance on the JOR task in humans. Rodent work has established that PFC activity changes gradually over the time scale of thousands of seconds. Hyman, Ma, Balaguer-Ballester, Durstewitz, and Seamas (2012) examined neural activity in ensembles of neurons in the medial PFC while rodents foraged randomly in one of two distinct spatial environments. They examined ensemble distance between population vectors as a function of the time between those two events both within and between environments. Hyman et al. (2012) found that PFC ensembles distinguished the two environments. However, this effect was superimposed on a dramatic effect of temporal distance. The distance between the ensemble response at two times continued to change even when those events were separated by a few thousand seconds.

Performance in the JOR task is affected by the integrity of both the PFC and the MTL (Chiba et al., 1994; Kesner, Gilbert, & Barua, 2002; Fortin, Agster, & Eichenbaum, 2002) and may depend on connections between the regions (Hannesson, Howland, & Phillips, 2004; Devito & Eichenbaum, 2011). Manns, Howard, and Eichenbaum (2007) measured the firing rate of ensembles of neurons recorded simultaneously from the hippocampus of rats performing a JOR task. Rats sampled a list of five odors. After each list, the rats were presented a pair of probes and rewarded for choosing the odor that was presented earlier in the list. Manns et al. (2007) computed the population vector averaged over a four second period of time around the time the animal sniffed each odor in the list. They found that the distance between the ensemble vectors corresponding to study of the list odors increased with the distance within the list, indicating neural autocorrelation over a few seconds. Notably, the within-list change predicted performance in the memory task.
Manns et al. (2007) also compared the activity of ensembles across lists. Despite an interlist interval of more than 2 minutes, ensemble activity continued to change gradually between events separated by at least a dozen lists. This means that the ensemble of neurons in the hippocampus continued changing gradually over more than a thousand seconds.

The foregoing results indicate that the PFC and MTL have access to a neural representation that changes gradually over at least a few thousand seconds and that is correlated with performance in the JOR task. However, this does not demonstrate that the representation is storing information about the stimuli per se. As such, the foregoing results do not demonstrate that these ensembles maintain a representation of recent experience. It is possible that the findings from each of these studies could simply be due to random fluctuations that happen to be correlated in time rather than being caused by the identity of the stimuli presented. We will discuss other findings that argue for a sensitivity of the ensemble response in the MTL to the identity of the stimuli below. For now we note two additional results that demonstrate stimulus-specific activity persisting across trials. Schoenbaum and Eichenbaum (1995b, 1995a) presented rats with a series of odors. The observed that some neurons in the rodent MTL responded on a trial to the identity of the stimulus on the preceding trial. Additionally, Yakovlev, Fusi, Berman, and Zohary (1998) showed that neurons in the monkey perirhinal cortex showed stimulus-specific activity during a delayed-match-to-sample task that persisted across trials. This is despite the fact that the requirements of the delayed-match-to-sample task do not require the animal to remember the sample stimulus after the test phase of the trial.

Neural evidence for jumping back in time during performance of episodic memory tasks. In several of the cognitive applications described here, we assumed that prior states of internal time can be recovered. In free recall, this accounted for the contiguity effect. Recovery of previous states of internal time was also essential for bridging across disparate experiences in the temporal mapping application. Previous work has argued that one of the functions of the MTL, and the hippocampus in particular, is to enable this jump-back-in-time (Howard et al., 2005). Although there is not yet definitive evidence that the brain is able to recover prior states of a temporal history, we discuss three recent neuroimaging studies in humans that are suggestive of this hypothesis.

Zeithamova, Dominick, and Preston (In press) had subjects learn simultaneously-presented pairs of pictures in an fMRI. The pairs shared an overlapping stimulus, i.e., A-B and B-C. Although A and C were not learned at the same time, they were both experienced in the context of B. Previous work has shown that bridging associations between A and C depend on the integrity of the hippocampus (Bunsey & Eichenbaum, 1996; Greene, Gross, Elsinger, & Rao, 2006). It has been argued previously that recovery of prior temporal context is sufficient to describe bridging associations (Howard et al., 2005; Howard, Jing, et al., 2009). If, during study of B-C, B recovers its prior temporal context, that is, it recovers components caused by A, it is as if A were encoded simultaneously with C. In the Zeithamova et al. (In press) study, the stimuli were chosen from different visual categories such that the category could be distinguished using multivoxel pattern classification. During study of B-C, the cortical pattern associated with the category of stimulus A was preferentially activated even though A was not physically present. The degree to which A was activated was correlated with the degree to which the anterior MTL, including the hippocampus, was
Figure 20. Neural recency and contiguity effects. Multiple neurons were recorded from human medial temporal lobe during performance of a continuous recognition task. **a.** We compared the vector across cells averaged over the 3 s each stimulus was presented to preceding states, excluding comparisons to the same stimulus. The ensemble changed gradually over macroscopic periods of time. **b.** When an item was repeated, we compared that state to the neighbors of the original presentation, subtracting out the contribution due to recency. There was a contiguity effect in both directions. After Howard, Viskontas, Shankar & Fried (in press).

activated by learning of the pairs.

A limitation of the Zeithamova et al. (In press) study, at least for the purpose of determining if the brain jumps back in time, is that the members of the pairs were presented simultaneously. A genuine jump back in time would require that a gradually-changing representation is recovered. It is possible that recovery of \( A \) in the Zeithamova et al. (In press) study is only recovery of instantaneously available information and simply reflects direct associations between \( B \) and \( A \). Manning, Polyn, Litt, Baltuch, and Kahana (2011) recorded intracranial EEG from patients with epilepsy while they performed a free recall experiment. Manning et al. (2011) demonstrated that the pattern of oscillatory components changed gradually over time during study of the list. Notably, the pattern of activity just before recall of a word from the list resembled the pattern during study of neighbors of the original presentation of the word. This similarity fell off gradually in both the forward and the backward direction and was greater for subjects who showed a larger behaviorally-observed contiguity effect. This finding is just what one would expect if there was a jump back in time during the period prior to recall of the item.

One limitation of the Manning et al. (2011) study is the fact that the sequence of recalls was uncontrolled. That is, while the pattern of activity preceding recall of a word list resembled the neighbors of that word from the study list, the recalled word may have been preceded by recall of other words from the list. Because of the contiguity effect, these preceding recalls are likely to have been the neighbors of the recalled word. This shortcoming was addressed by a reanalysis of single-unit recordings from the medial temporal lobe of
epileptic patients performing a continuous recognition task on images (Howard, Viskontas, Shankar, & Fried, In press; Viskontas, Knowlton, Steinmetz, & Fried, 2006). The ensemble pattern of activity after the presentation of each stimulus in the continuous recognition study was used to assemble a population vector. These ensemble states were then compared to one another. MTL ensembles changed gradually across at least tens of seconds (Fig. 20a), demonstrating a neural recency effect. When a stimulus was repeated as an old item, the gradually-changing part of the ensemble resembled the neighbors of the original presentation of that stimulus (Fig. 20). The ensemble similarity fell off gradually in both the forward and backward direction. That is, the MTL ensemble changed gradually over tens of seconds. When an item was repeated, the ensemble “jumped back in time” to resemble the state prior to the initial presentation of that stimulus.

Of course the Howard et al. (In press) study also has limitations. First, the extra-cellular recording technique used with human patients is not nearly as reliable in isolating individual neurons as the techniques used regularly in animal studies. Second, because the repeated stimuli were repeated at a variety of recencies to their initial presentation, it was necessary to statistically isolate the relatively small effect of contiguity from the relatively larger effect of recency. Finally, because performance was very high, it was not possible to establish that the neural jump back in time was correlated with successful memory retrieval of the repeated item.

While each of the studies here have some limitations, taken as a group they present a much stronger story. Although the Manning et al. (2011) result could be due to correlated retrieval strategies, neither the Howard et al. (In press) result nor the Zeithamova et al. (In press) result are susceptible to that concern. While the Zeithamova et al. (In press) experiment involved stimuli simultaneously presented, both the other two papers were able to establish a gradually-changing representation. While the Howard et al. (In press) paper relied on statistically isolating the contiguity effect from the recency effect, the methods of both of the other two papers were such that there was little or no recency effect and the neural contiguity effect could be directly measured. Of course, very little can be gleaned about the nature of the gradually-changing representation that is recovered from any of these studies, other than the fact that it changes gradually over stimulus presentations.

**Evidence for a representation of internal past time.** Any number of methods would result in a representations that changes gradually over time—for instance a strength model changes gradually over time. There are, however, distinctive predictions of the representation of internal past time that can be measured. First, it should be possible to reconstruct not only what stimulus was experienced in the past, but also how long in the past that stimulus was experienced. Second, cells should respond to their preferred stimulus not immediately, but after some characteristic time controlled by its value of $\tau^*$ (Fig. 4a). Cells with different values of $\tau^*$ will peak at different times after the stimulus was presented (Figure 21a). Third, the spread in the time a cell fires should increase with the delay at which it fires. We review two studies, one from the rodent hippocampus, the other from the monkey putamen, that show striking evidence for all of these properties over delays of a few seconds. First, though, we briefly note another recent paper that shows evidence for widespread temporal and stimulus coding throughout the MTL.

Naya and Suzuki (2011) recorded from various regions in the monkey MTL while they
Figure 21. Time cells in the brain. Each plot provides a heat map that describes the responsiveness of multiple cells as a function of time. Each row shows the normalized firing rate for actual neurons (b & c) or activity for model nodes (a) as a function of time within a trial, which runs from left to right. The rows are organized such that cells that fire earliest are at the top and the cells that fire latest are at the bottom. 

- **a.** Heat map showing the normalized activity of a set of simulated cells representing different values of internal time as a function of physical time since a stimulus was presented. After Shankar & Howard (2012).

- **b.** MacDonald et al. (2011) recorded neurons from region CA1 of the dorsal hippocampus during the delay period of a non-spatial memory task. The heat map shows the firing rate of a set of simultaneously recorded neurons as a function of time during the delay interval. These time cells fire during circumscribed periods of the delay. After MacDonald, et al. (2011).

- **c.** Adler et al., (2012) recorded from medium spiny neurons in the putamen during performance of a simple stimulus-response learning task. On each trial, a stimulus was presented. Two seconds later, an outcome was presented. The heatmap shows all of the medium spiny neurons recorded from the putamen in two monkeys across recording sessions.

Studied a list of two visual stimuli. The monkeys were required to remember the stimuli that were presented as well as their order. Naya and Suzuki (2011) looked for neurons that distinguished the stimulus presented and for neurons that distinguished the time within the trial. Naya and Suzuki (2011) found both types of cells, with stimulus-specific cells more prominent in regions like the perirhinal cortex and inferotemporal cortex. Cells that distinguished time within a trial were more prominent in central MTL regions such as the hippocampus and the entorhinal cortex. There were also a large number of cells that were sensitive to both the identity of the stimulus and the time within trial. Although this study did not look for cells that respond to a particular stimulus at a characteristic delay, they did show that the hippocampus and entorhinal cortex contained enough information to decode time within the trial and at least the currently-presented stimulus.

MacDonald, Lepage, Eden, and Eichenbaum (2011) simultaneously recorded many neurons from dorsal CA1, a subfield of the hippocampus, during the delay period of a memory task. The animal explored one of two objects prior to a delay period. The identity of the object shown prior to the delay predicted which of two odors would be paired with reward after the delay. Cells fired at circumscribed periods of time within the delay period (Figure 21b). These cells have been dubbed “time cells” because they fire in circumscribed regions of time analogous to hippocampal “place cells” which fire in circumscribed regions of allocentric space (O’Keefe & Dostrovsky, 1971; Wilson & McNaughton, 1993; see also Pastalkova, Itskov, Amarasingham, & Buzsaki, 2008). Time cells show several properties that are qualitatively consistent with the nodes that constitute $T$. Different cells fire max-
imally at different delays (Figure 21b). As a consequence, at one moment it is possible to examine the pattern of activity across the ensemble and determine how long has passed since the delay period began. The temporal spread of firing of time cells also increases with the latency at which they respond (Figure 21a, unpublished secondary analyses of MacDonald et al., 2011). The profile of firing rate as a function of time during the delay interval is positively skewed, as predicted by the model (unpublished secondary analyses of MacDonald et al., 2011). Critically, a representation of stimulus history requires that time cells differentiate the specific stimuli presented in the past. A few of the time cells responded differentially based on the identity of the object that had been explored prior to the delay, consistent with the requirement of stimulus-specificity (MacDonald et al., 2011).

Adler et al. (2012) recorded from medium spiny neurons in the putamen while monkeys were presented with simple stimulus-response pairings. On each trial the monkey was presented with a visual stimulus; two seconds later one of three outcomes—an aversive airpuff, delivery of rewarding water or a neutral tone—was presented. The identity of the stimulus predicted the outcome. Figure 21c shows the response of each of the medium spiny neurons as a function of time since the stimulus was presented. As with the hippocampal ensemble shown in Figure 21b, the medium spiny neurons show several properties predicted for cells participating in $T(\tau)$. First, different cells respond at different latencies after the stimulus. From examining which cells are active we could determine the amount of time that has passed since the stimulus was presented. Second, the temporal spread in the firing of cells increases with the latency at which they respond. Third, the response profile is asymmetric, with the tail extending longer after the time of maximal responding longer than the leading edge of the temporal spread (Adler et al., 2012, also secondary analyses). Finally, although Adler et al. (2012) did not analyze the individual stimuli separately, they did show that medium spiny neurons responded differently on trials with different outcomes. Because each stimulus was paired with one type of outcome, demonstrating outcome specificity is not distinguishable from stimulus specificity without further analyses.

It is striking the hippocampus and putamen are anatomically well-separated regions believed to be part of separate memory systems. The hippocampus is part of the medial temporal lobe, which is believed to be part of the declarative memory system (Squire, 1992; Cohen & Eichenbaum, 1993), which is responsible for episodic and semantic memory. The putamen, in contrast, is part of the striatum, which is usually understood to be part of a non-declarative habit learning memory system (e.g., Knowlton, Mangels, & Squire, 1996; Squire & Zola, 1996). The function of the hippocampus and striatum are literally covered in different chapters of textbooks on the cognitive neuroscience of memory (e.g., Eichenbaum, 2012). The commonalities in the properties of the neurons in the hippocampus (Fig. 21b) and the putamen (Fig. 21c) are even more striking when one considers how distinct the function of these regions is believed to be. This is just what one would expect, however, if a variety of memory systems have access to a common representation of internal time. Perhaps the differences between the different memory systems can be understood largely in terms of the different operations they make on a common representation of internal time.
General Discussion

We described a mathematical model that provides a principled method to construct a representation of internal past time and internal future time. Internal past time is constructed by taking the Laplace transform of experience and then approximating the inversion of the Laplace transform. The representation of internal time is mathematically scale-invariant. This property is extremely important for a learner faced with a world that exhibits correlations over a variety of time scales (Shankar & Howard, submitted). Behavioral applications showed that this mathematical model of internal time is sufficiently rich to describe applications in what are typically thought of as distinct fields of cognitive psychology, including working memory, episodic memory, conditioning and timing.

If Husserl’s account of internal time is correct, it requires that the brain maintain a representation that retains information about what stimuli were experienced and how far in the past those experiences took place. In many applications in cognitive psychology memory for when and memory for what are treated as separate problems. Clock models keep track of a representation of time relative to some starting point (Brown et al., 2000; Gibbon, 1977; Matell & Meck, 2004). Clock models have found applications not only in describing performance in timing tasks (Rakitin et al., 1998), but also in applications such as serial recall. For instance, the OSCAR model of Brown et al. (2000) constructs a representation of time since the beginning of a list using the phase of oscillators with different frequencies. During encoding, each list item is associated to this clock signal. When the clocks are reset, they run through the same sequence of states, resulting in recall of the list items. The important distinction is that clock models are independent of the stimuli experienced during that interval. In contrast, strength models do not retain any information about time per se, but do retain information about recently-presented stimuli. The temporal context vector in TCM is a strength model that illustrates the utility of retaining information about which stimuli have been recently presented. Because temporal context contains information about recently-experienced stimuli, this enables the description of associative contiguity effects (Howard, Venkatadass, et al., 2007; Howard, Jing, et al., 2009).

The representation of internal past time can be thought of as a hybrid of a clock model and a strength model. In this sense, because it retains continuous information about what and when the model of internal past time has much in common with spectral resonance theory (Grossberg & Merrill, 1992, 1996). One could construct a clock model from the representation of internal time by summing \( T(\tau) \) over the columns corresponding to the start signal. Similarly, one could construct a strength model by summing over values of \( g(\tau) \). It would be extremely difficult to account for contiguity effects using a clock model. Similarly, it is very difficult to account for timing effects with a strength model. Starting with a representation that includes both a clock model and a strength model as limiting cases, one can account for a greater range of phenomena. While this representation of internal time is undoubtedly richer than a clock model or a strength model this flexibility is not a weakness if our goal is to develop unified theories of cognition.

Note that a shift-register model of short-term memory (e.g., Gureckis & Love, 2010) also contains information about what happened when. A shift register with capacity \( n \) retains the identity and order of the last \( n \) items by holding them in \( n \) slots. The slot an item representation resides in provides information about how long ago it was presented.
The problem with a shift register solution is that no information at all is available for item \( n + 1 \). In order to get information about long time scales from a shift register, the capacity must go up linearly with the longest delay. Note that if one restricts their attention to short time scales where the spread in \( T(\hat{\tau}) \) is narrow relative to the temporal spacing of the stimuli, \( T(\hat{\tau}) \) behaves like a shift register. However, the scale-invariance of \( T(\hat{\tau}) \) provides a distinct advantage over a shift register. The scale-invariance of \( T(\hat{\tau}) \) can be understood as an attempt to sacrifice accuracy in a way that optimally distributes temporal information across a finite set of resources (Shankar & Howard, submitted).\(^{18}\) Scale-invariance means that the representation can be used to describe behavior across a variety of time scales, from one or two seconds (the Hacker scanning paradigm, immediate free recall) to much longer time scales (Yntema & Trask, 1963; final free recall). Critically, although performance functions can give rise to qualitative differences across scales, the representation itself transitions smoothly across scales. This property enables models built using this framework to account for scale-invariant behavioral phenomena in timing and conditioning (see also Shankar & Howard, 2012).

The focus in this paper has been on understanding the implications of a mechanistic representation of internal time for cognition considered broadly rather than working out detailed behavioral models of any particular task. A potential disadvantage of this approach is that the memory representation is only weakly constrained by any single piece of behavioral data. Fortunately, neurophysiological results can provide extremely strong constraints on the memory representation. Because the representation is concrete and mechanistically generated, direct physical predictions can be directly evaluated in the brain. Given the relative newness of this particular hypothesis, the degree of preliminary neural evidence in support of this representation of internal time is encouraging at this stage of model development. Going forward, relatively subtle properties of the mathematics, such as scale-invariance and the distribution \( g(\hat{\tau}) \) could be definitively falsified with sufficient interest and patience from neurophysiologists and neuroimagers using existing technology.

Theoretical issues

The current framework may prove to be a good starting point for behavioral models of a broad variety of findings in learning and memory. Going forward, there are several important issues that are likely to arise in developing these behavioral models.

Interference vs time per se. One of the oldest questions in the study of memory is whether forgetting is caused by interference or the passage of time per se. Although the way Eq. 1 is written suggests a strong commitment to time per se, this is not a strong constraint. The units of \( s \) control the rate at which internal time flows as a function of physical time \( \tau \). If we were to change all of the values of \( s \) in register by multiplying them all by the same value, \( s \rightarrow \beta s \), then we would have changed the rate at which internal time flows relative to physical time. We can write a generalization of Eq. 1 as follows:

\[
\frac{dt}{d\tau} = -s \beta(\tau) \ t + f(\tau)
\]  

\(^{18}\)Note that both clock models (e.g., Miall, 1989; Brown et al., 2000) and strength models (e.g., Staddon & Higa, 1999, the power law forgetting in Anderson, Bothell, Lebiere, & Matessa, 1998) can yield scale invariant behavior.
As long as $\beta$ is not a function of $s$, it serves to control the flow of internal time. Constructing an interference-based account of forgetting simply requires choosing $\beta(\tau)$ to depend on the amount of information presented to the network. For instance, if the stimulus function $f(\tau)$ is a series of discrete stimulus presentations separated by arbitrary delays, then choosing $\beta(\tau) = b|f(\tau)|$, where $b$ is a constant, will generate an ordinal representation. When there is no stimulus presented, $|f(\tau)| = 0$ and the right-hand-side of Eq. 27 is zero. Choosing $\beta(\tau)$ in this way makes internal past time “stand still” when no stimulus is presented.

So, what is the appropriate choice of $\beta(\tau)$? It does not seem likely that there is a single answer to this question. Just restricting our attention to the two episodic memory tasks used here, it is not difficult to find contradictory results. Some results suggest that memory should be sensitive to time per se. For instance Hintzman (2004b) presented subjects with a continuous absolute JOR task in which the items were presented for different times. The ratings of recency depended on the intervening time rather than the number of intervening items. This suggests that $\beta(\tau)$ is a constant function of time. Other evidence shows that the recency effect in recall tasks is unaffected by an unfilled delay (A. D. Baddeley & Hitch, 1977; Murdock, 1963). This is consistent with $\beta(\tau)$ being a function of the stimuli being attended to. It seems that the discussion should not center on whether interference or time per se causes forgetting—the answer is apparently “both.” The discussion should center on whether there are multiple parts of $t_s$ with different functions $\beta(\tau)$ that are brought to bear in different experimental circumstances or whether $\beta(\tau)$ changes in response to task demands.

Temporal effects over very long time scales. We argued that the upper limit $\tau_{\text{max}}$ should be chosen such that events as much as ten thousand seconds in the past could be retained. What then are we to make of the finding of recency effects over very long time scales (e.g., Moreton & Ward, 2010; A. D. Baddeley & Hitch, 1977)? The free recall model can be thought of as an activity-dependent $T$ coupled with weights stored in $M$ (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005). In the present treatment, $T$ is solely responsible for change over time; associations stored in $M$ are treated as constant. This is almost certainly not a reasonable assumption over very long time scales. It is possible that over time scales much longer than those considered here, changes in associations stored in $M$ dominate forgetting. It is also possible that very long-term recency effects are caused by autocorrelations in the stimulus function. For instance, if $f(\tau)$ is autocorrelated over very long time scales, this autocorrelation will naturally be reflected in the states of $T$ observed over very long time scales, resulting in associations. For instance, suppose one lives in Alaska for several years, where one frequently sees grizzly bears and then moves to Syracuse, where there are no grizzly bears. If we looked at the part of $f(\tau)$ coding for grizzly bears over the scale of many years we would see an elevation in activation during the period of time the subject was in Alaska. That is, there would be an autocorrelation in $f(\tau)$ on the scale of a year. A behavioral association between one event in Alaska and another event a year later may be a consequence of both being experienced in the temporal context of grizzly bears. Another, not mutually exclusive possibility is that recursive reminding or active retrieval of past experiences constructs an effective stimulus function that is autocorrelated over time scales much longer than the largest time scale represented in $T$. For instance, suppose that $\tau_{\text{max}}$ is 10,000 s, but every 5,000 s, the subject decides to take a minute and reflect on the
events of the last hour or so. In this case, after the first presentation of a stimulus, it can be retrieved via scanning (or currently-available cues). However, the event of remembering the initial presentation of the stimulus is itself part of the mnemonic record. This later memory can then be retrieved at future times. In this way, memory for a stimulus could be extended over very long time scales.

Connection to other theories

The present mathematical framework can be seen as continuous with verbally-specified theories from several fields. In some cases it is a straightforward mechanistic implementation of these ideas; in other cases it at least provides a mathematical language to begin discussing how to implement them.

Conditioning. In conditioning, the temporal encoding hypothesis (Matzel et al., 1988; Savastano & Miller, 1998) describes even apparently simple associations as the result of learning inherently temporal relationships between stimuli. We have argued that the mathematical framework described here is sufficiently rich to describe at least the basic phenomena from temporal mapping experiments (e.g., Cole et al., 1995). Balsam and Gallistel (2009) make two extremely important further points. First, they note that it does not appear that there is a characteristic scale for formation of associations. Shankar and Howard (2012) showed that the scale-invariance of $T$ and $p$ make it well-suited for describing this aspect of learning. Second, Balsam and Gallistel (2009) argued that the results of the standard conditioning paradigms can be understood as a natural consequence of the informativeness about time of presentation of stimuli. For instance, in the blocking paradigm, CS1 is paired with the US during a first phase of training. In a second phase of training, CS1 and CS2 are paired with the US. In this view, the finding that CS2 does not cause a CR is understood as a consequence of the fact that its occurrence does not provide additional information about when the US occurs. Expanding the current framework to encompass these learning paradigms is a significant challenge for future work. More subtle learning rules for $M$ than the simple Hebbian learning rule used here would be an excellent starting point for this investigation.

Episodic memory. Human memory research has long considered the verbally-specified theory that that episodic memory consists of mental time travel (Tulving, 1983, 1985b). According to Tulving, (e.g., Tulving, 1985b), episodic memory is accompanied by a conscious experience he calls auto-noetic consciousness. That is, the subject’s experience of an episodic memory is accompanied by the sense that the previous event is part of the subject’s personal timeline. This timeline extends into the past and will also persist into the future. This concern with the distant future in the field of memory research seems somewhat counterintuitive, but follows from the definition of auto-noetic consciousness. Tulving found that the amnesic patient K.C. does not show evidence for either vivid recollection of past events nor the ability to imagine future events (Tulving, 1985b); additional evidence from behavioral, neuroimaging and neuropsychological studies suggest that this definition makes sense (see Schacter, Addis, & Buckner, 2007, for a review). The mathematical framework we have described provides a representation of past time and a trajectory of predictions extending into the immediate future. However, it is descriptive with respect to why there is a contiguity effect and silent with respect to the quality of auto-noetic consciousness.
According to the model, the contiguity effect—presumably corresponding to mental time travel to an event in the personal past—is accompanied by partial recovery of a prior state of $T$. This mechanistically causes a contiguity effect that resembles that observed in the laboratory and will also cause the recovery of a predictive trajectory that resembles that when initially presented. But it does not say how the recovery of $T$ (or $t$) is accomplished nor does it say anything about why that would correspond to a particular type of consciousness. Similarly, one could describe the process of future events (e.g., imagine what your next birthday party will be like) by postulating that $T$ is set to a state that resembles what the subject infers $T$ will be like at that point in the future. But that doesn’t mention how setting $T$ to that state could be accomplished nor why that is experienced in a particular way. Filling in these gaps should a important goal for future research.

Internal future time and temporal difference learning. The inclusion of internal future time here is a significant departure from previous models of human verbal learning and memory. The mechanism to predict trajectories in internal future time could prove to be extremely useful in a variety of fields. For instance, in reinforcement learning, the temporal difference (TD) learning algorithm (Sutton & Barto, 1981, 1990) has been extremely influential, having been suggested as an explanation for the firing of dopaminergic neurons at the heart of the brain’s reward system (Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997). TD-learning is an attempt to solve the credit assignment problem. Given a variety of stimuli that precede a reward in real time, how does one determine which stimulus predicted the reward? The solution TD-learning proposes is that “credit” diffuses backwards over many learning trials along a compound serial representation of the stimulus. But, notice that the credit assignment problem is only a challenge if the stimuli preceding the reward are meaningfully separated in time. If one can predict future trajectories from the current state of the history, the credit assignment problem becomes much easier to solve. Each time a stimulus is presented, one just “plays forward” the future trajectory. If the trajectory includes reward, that stimulus predicts reward. The first stimulus in a series to predict the reward is the one that deserves the credit. This distinction between TD-learning and the approach we just sketched is extremely important. The models make very different predictions about second order conditioning. Given the importance of the dopamine system in brain disorders, decision-making, ADHD, and addiction the question of how stimuli come to be associated to reinforcers is not of solely theoretical interest.

Cognitive neuroscience of the MTL. Cognitive neuroscientists studying the role of the hippocampus in declarative memory, have long suggested that it plays a special role in encoding sequences that are experienced along a spatiotemporal trajectory (e.g., Levy, 1996; Sohal & Hasselmo, 1998; Jensen & Lisman, 1996; Lisman, 1999; Eichenbaum, 1999; Hasselmo, 2009, 2012). This view has been expressed through a number of computational network models of hippocampal function (Jensen & Lisman, 2005; Hasselmo & Eichenbaum, 2005; Koene & Hasselmo, 2007; Hasselmo, 2009; Hasselmo, Giocomo, Brandon, & Yoshida, 2010). A variety of neuropsychological (Agster, Fortin, & Eichenbaum, 2002; Fortin et al., 2002), neuroimaging (Lehn et al., 2009; Ross, Brown, & Stern, 2009; Kumaran & Maguire, 2006), and neurophysiological findings (Wood, Dudchenko, Robitsek, & Eichenbaum, 2000; Pastalkova et al., 2008; Lisman & Redish, 2009) can be seen in this light.

$T$ is a scale-invariant representation of the set of stimuli that led up to the present
moment. As such it can be understood as containing information about the sequence, or temporal trajectory, leading up to the present moment.\textsuperscript{19} The current approach differs from circuit level models of sequences not so much in the resulting representation, but in the mechanisms by which the representation of the sequence is constructed. \( T \) is described mechanistically, but at a much more abstract level and without explicit reference to neural circuits. Perhaps as a consequence, there are numerous differences in the mechanisms used in the two approaches. For instance, over time nodes of \( T(\tau^*) \) are sequentially activated. However, there are no direct connections between these nodes as in many circuit-level models of sequences. The advantages of the more abstract approach are mathematical tractability and relative ease in connecting to relatively detailed behavioral findings. But the more abstract approach also brings with it weaknesses. Even if we knew with certainty that a representation exactly like \( T(\tau^*) \) was maintained in the brain, that would not tell us if it was constructed using \( t(s) \) and \( L^{-1}_k \) as intermediate steps. Even if we also knew with certainty that it was constructed using \( t(s) \) and \( L^{-1}_k \) as intermediate steps, this would not tell us how \( t(s) \) or \( L^{-1}_k \) are implemented at the neural circuit level. A fully specified neural circuit model constructing the representation of internal time would be much more easily falsified, at multiple levels of description than the more abstract hypothesis developed here. This is a significant challenge for future work. For instance, Wood et al. (2000) had animals navigate a maze along a sequence of movements that had an overlapping section. They showed that hippocampal neurons fired along the common section differentially according to which sequence the animal was following. This is a natural prediction of the hypothesis that part of the function of the hippocampus is to generate a different code for two distinct sequences that share an overlapping element (Levy, 1996; Sohal & Hasselmo, 1998; Lisman, 1999). We note that \( T \) is a scale-invariant representation of the sequence of stimuli leading up to the present moment and is at least roughly consistent with this result. Consider the states of \( T \) that obtain after studying the sequence \( A \, B \, L \, M \) and after studying the sequence \( Y \, Z \, L \, M \). Even though the most recent stimuli, \( L \) and \( M \), are the same in each sequence, there would be a number of cells coding for earlier stimuli that would be differentially active after studying the two sequences. For instance, a cell that coded for a stimulus dimension differentiating \( B \) from \( Z \) with \( \tau^* = -3 \) would be active after studying the first sequence but not after the second sequence. In this way, \( T \) is close in spirit to the idea that coding of sequences is an essential part of declarative memory. However, it does so using a very different computational framework than the buffer models (Jensen & Lisman, 2005; Koene & Hasselmo, 2007) and models that construct conjunctive sequential codes (Lisman, 1999; Levy, 1996).

Conclusions

We described a mathematical framework for understanding the role of time in memory. The mathematical framework provides a mechanistic description of a scale-invariant representation of stimulus history as a function of internal time. The representation of

\textsuperscript{19} It is possible to generalize the mathematical framework developed here for representing internal time to represent position relative to environmental landmarks (Howard, Du, & Shankar, 2012). That is, more the current framework can be utilized to not only represent temporal trajectories, but more general spatiotemporal trajectories through internal past and future time.
the stimulus history can be used to construct a representation of internal future time—a trajectory heading forward from the present. We showed that this framework is sufficient to provide a concise description of fundamental behavioral results from a variety of paradigms, including an account of JORs over both short (Hacker, 1980) and long time scales (Yntema & Trask, 1963; Hintzman, 2011), recency and contiguity effects in episodic memory (Glenberg et al., 1980; Kahana, Howard, & Polyn, 2008), the time-left procedure in the field of interval timing (Gibbon & Church, 1981) and temporal mapping from conditioning (Cole et al., 1995). We reviewed preliminary neural evidence consistent with the idea that the brain directly implements something like this representation of internal time. The framework provides a common mathematical language in which to discuss ideas about time and memory in what have been thought of as separate fields, including episodic memory as mental time travel (Tulving, 1983), the temporal encoding hypothesis from conditioning (Matzel et al., 1988; Savastano & Miller, 1998; Balsam & Gallistel, 2009). This raises the possibility that in the future we could have models of memory and cognition that are both mathematically elegant and instantiated in the physical brain.

References


**Appendix A: Motivating the choice for the number density**

Let a delta function stimulus have been presented at some time $\tau$ in the past and consider the activity across time cells, $T(\tau, \ast \tau)$:

$$T(\tau, \ast \tau) = \frac{C}{\tau} \left( \frac{\tau}{\ast \tau} \right)^{k+1} e^{k(\ast \tau)}$$  (28)

where $C \equiv \left( \frac{-k}{k+1} \right)^{k+1}$. Consider the peak in this curve as a function of $\ast \tau$. As $\tau$ gets bigger, the peak in relative activity as a function of $\ast \tau$ gets more broad. If we think of $\ast \tau$ as being supported by a discrete set of nodes, then if the number density were constant, then the relative change from one node to the next around the peak goes down with increasing $\tau$. This means that there is increasing redundancy in adjacent values of $\ast \tau$ with increasing $\tau$. This is wasteful—a node coding for a small value of $\ast \tau$ carries more information different from its neighbors than a node coding for a $\ast \tau$ further in the past. If we could choose $g(\ast \tau)$ in
such a way that the shape of $T$ across nodes was constant regardless of the delay, this would be an optimal situation. The distribution of information across nodes would be invariant with the delay.

Let the nodes in the $T$ column be arranged as follows

\[
\tau_o, \tau_1, \tau_2, \ldots, \tau_n, \ldots
\]

Now as the activity passes through the $T$ column nodes, we require that the shape of the distribution stays the same. This requirement essentially implies that for any two time points $\tau_1$ and $\tau_2$ following the occurrence of the stimulus, the shape of the representation across the $T$ nodes is only shifted and perhaps multiplied by a factor that is constant for all values of $n$. That is, we require that there exists some constant translation $c$ that makes

\[
T(\tau_1, \tau_n) \propto T(\tau_2, \tau_n+c)
\]

for all $n$. The requirement of translation-invariance boils down to requiring that the ratio $T(\tau_1, \tau_n)/T(\tau_2, \tau_n+c)$ is independent of $n$. Using eq. 28 to evaluate this ratio, we have

\[
\frac{T(\tau_1, \tau_n)}{T(\tau_2, \tau_n+c)} = \left(\frac{\tau_1}{\tau_2}\right)^k \left[\frac{\tau_n+c}{\tau_n}\right]^{k+1} e^{\left[\frac{\tau_1}{\tau_n} - \frac{\tau_2}{\tau_n+c}\right]}
\]

The right-hand-side is the product of a power law function and an exponential. If we can make both the power law function and the exponential independent of $n$, then this makes the product independent of $n$ and satisfies our requirement. Considering the power law function, it is clear that if $\tau_n \propto (1+a)^n$, where $a$ is some number controlling the relative spacing of the nodes, the power law term is independent of $n$. That is, the values of the nodes should be given by

\[
\tau_o, \tau_o(1+a), \tau_o(1+a)^2, \ldots, \tau_o(1+a)^n, \ldots
\]

The choice of $\tau_n \propto (1+a)^n$ leads immediately to $g(\tau_n) \propto \tau_n^{-1}$. To see this note that $g(\tau_n)$ is given by the inverse of the change from $\tau_n$ to $\tau_{n+1}$. We find that difference is:

\[
\tau_{n+1} - \tau_n \propto a (1+a)^n \propto \tau_n
\]

so $g(\tau_n) \propto \tau_n^{-1}$.

It remains to be shown that Eq. 31 is invariant with $n$. We still require the exponential function to be invariant with $n$. This is accomplished if the argument of the exponential function is zero. This is satisfied if $c$ is chosen such that $(1+a)^c = \tau_2/\tau_1$. If we take the discreteness of the nodes seriously, this requires that $c$ is an integer. In general this is not possible, but if $a$ is small the error is not dramatic. Moreover, given a specific $\tau_1$, as we wait longer and longer, there are an infinite number of $\tau_2$s that will satisfy the requirement and be a precisely shifted version of $T(\tau_1, n)$. In between the values of $\tau_2$ that make $c$ an integer there will be some deviation due to the discreteness of the nodes, but the deviation across nodes is well-behaved even as $\tau_2$ grows without bound.
Appendix B: Calculating self-terminating scanning model

We assume that the subject can examine a column of $T$, starting from the present and extending backward in time. Let us denote the probability of the query returning at a particular value of $\tau$ as $p_r(\tau)$. We assume that the instantaneous probability of returning information from a query of $T$ is given by

$$p_r(\tau) = \alpha T(\tau),$$

(33)

where $\alpha$ is a free parameter. For a delta function stimulus presented at time $\tau_0$ in the past, the expectation value of the number of values the query returns is given by

$$\int_0^\infty p_r(\tau) g(\tau) d\tau = \alpha \int_0^\infty T(\tau) \frac{1}{|\tau|} d\tau = \frac{\alpha}{\tau_0},$$

(34)

where we have used the fact that we have set $g(\tau) = |\tau|^{-1}$. Now in a race model, we let the decision terminate as soon as the query returns information. The probability that the query returns information for the first time while scanning at $\tau$ is given by

$$p_f(\tau) = \left[1 - \int_0^{\tau} p_f(\tau') d\tau'\right] p_r(\tau),$$

(35)

This has the general solution

$$p_f(\tau) = p_d(\tau) \exp \left[-\int_0^{\tau} p_d(\tau') d\tau'\right],$$

(36)

where $p_d$ is the instantaneous probability of a query returning information multiplied by the number density $p_d(\tau) \equiv p_r(\tau) g(\tau)$. Equation 35 ensures that the integral of $p_f$ is not greater than one, allowing it to be understood as the probability that the query returns useful information.

When $\alpha$ is small relative to $\tau_0$, the time in the past at which the item was presented, then $\int_0^\infty p_d(\tau) d\tau$ is much less than one and $p_f(\tau) \approx p_d(\tau)$. When $\tau_0$ is much less than $\alpha$, then $\int_0^\infty p_d(\tau) d\tau$ is much greater than one. This means that the probability of the search returning at least some information is nearly one and more or less independent of $\tau_0$. When $\tau_0$ is much greater than $\alpha$, the probability goes down like $\alpha \tau_0$. When $g(\tau) = |\tau|^{-1}$ and $\tau_0 \gg \alpha$, the mean number of cells that must be scanned to obtain information goes up with $\log(\tau_0)$. If the rate of scanning is constant as a function of cell, and if the rate-limiting step in making the decision is the availability of information from the scanning process, then the mean time to terminate a query should also go up like $\log(\tau_0)$.

In the absolute JoR task, we report the queried value, expressed in units of cells. In the relative JoR task, we have two queries racing to be selected. Suppose we have two probes, $A$ and $B$. In this case, if we have the probability of first return for stimulus $A$ as $p_{f,A}(\tau)$, and $p_{f,B}(\tau)$, then the probability of selecting stimulus $A$ at time $\tau$ will be given by

$$p_A(\tau) = p_{f,A}(\tau) \left(1 - \int_0^{\tau} p_{f,B}(\tau') d\tau'\right),$$

(37)
In order to generate predictions, all that is necessary is to pick a value for $\alpha$, put explicit values of $T$ given the experimentally-relevant delays into Eq. 33 and go forward. Unfortunately, the resulting integrals cannot be calculated analytically. Accordingly, we have estimated them numerically using the R function \texttt{integrate()} in generating predicted results.

Appendix C: Derivations of quantities used in modeling free recall

Notation and assumptions

Let’s denote the list items by $f_n$, where $n$ denotes the position of the item from the end of the list. Let the $n$-th item be presented at time $\tau_n$. For simplicity, we shall take the time interval between successive items to be a constant $\delta \tau$. We will further assume that there is a never-ending stream of stimuli, each presented with an interstimulus interval of $\delta \tau$. This enables us to ignore edge effects. The $T$ layer activity at the end of the list is $|T(\tau_o)\rangle$ and prior to the presentation of the item $f_n$ is $|T(\tau_n)\rangle$. This is represented pictorially in the top of Figure 22. At the time when an item is presented, only the columns corresponding to preceding items contain information. This means that only the predecessors of $f_n$ will leave an image in $|T(\tau_n)\rangle$. As a consequence, only the predecessors of $f_n$ are stored in $M$.

Recency: Computing the activation of a word $n$ steps in the past

To estimate how well the item $f_n$ is predicted by the state of $T$ at the end of the list, $T_{\tau_o}$, we have to compute the inner product $\langle T(\tau_n)|T(\tau_o)\rangle$. Note that although $|T(\tau_o)\rangle$ has non-zero activations in the $T$-columns corresponding to all the items in the list, $|T(\tau_n)\rangle$ only contains activity in those columns corresponding to the items that were presented prior to $f_n$. Their inner product will hence only involve those columns corresponding to items presented prior to $f_n$. To compute the inner product, let’s denote the $T$ column of an item $f_{n+m}$ to be $T^m(\tau_o, \tau)$ and $T^m(\tau_n, \tau)$ at times $\tau_o$ and $\tau_n$ respectively. At the end of the list, the component of the prediction $p_n$ corresponding to the item $f_n$ is given by

$$p_n = \langle T(\tau_n)|T(\tau_o)\rangle = \sum_{m=1}^{\infty} A_n^m$$

(38)
where we have defined
\[ A^m_n = \int T^m(\tau_o, \tau) T^m(\tau_n, \tau) g(\tau) d\tau \] (39)
as the inner product of the T layer activities at times \( \tau_o \) and \( \tau_n \), restricted to the column corresponding to the item \( f_{n+m} \). \( A^m_n \) measures the contribution of the image of \( f_{n+m} \) in \( T_{\tau_o} \) towards predicting \( f_n \).

The value of the coefficients \( A^m_n \) with \( g(\tau) = |\tau|^w \) can be calculated analytically. It turns out that for any given \( n \), the summation of \( A^m_n \) over all values of \( m \) converges only for \( w < 0 \). The prediction for \( p_n \) when \( w < 0 \) is
\[ p_n \simeq n^w(\delta \tau)^{(w-1)} \] (40)

To compute the prediction component \( p_n \) from eqns. 38 and 39, we start with
\[ T^m(\tau_o, \tau) = \frac{-1}{\tau} \frac{(k)^{k+1}}{k!} \left( \frac{-m \delta \tau}{\tau} \right)^k e^{k\left(\frac{m \delta \tau}{\tau}\right)} \] \[ T^m(\tau_o, \tau) = \frac{-1}{\tau} \frac{(k)^{k+1}}{k!} \left( \frac{-(n + m) \delta \tau}{\tau} \right)^k e^{k\left(\frac{(n + m) \delta \tau}{\tau}\right)} \] (41)
Substituting into eq. 39, the coefficients, \( A^m_n \) can be written as
\[ A^m_n = \frac{k^{2k+2}}{k!^2} (1 + n/m)^k (m \delta \tau)^{2k} \int \left( \frac{1}{\tau} \right)^{2k+2} e^{k(2+n/m)\left(\frac{m \delta \tau}{\tau}\right)} g(\tau) d\tau \] (42)
With \( g(\tau) = |\tau|^w \), the above integral can be evaluated to
\[ A^m_n = \frac{k^{2k+2}}{k!^2} (1 + n/m)^k (m \delta \tau)^{2k} \left[ \frac{\Gamma(2k + 1 - w)}{(k(2 + n/m)m \delta \tau)^{2k+1-w}} \right] \]
\[ = \frac{k^{(1+w)} \Gamma(2k + 1 - w)}{k!^2} \left[ \frac{(1 + n/m)^k}{(2 + n/m)^{2k+1-w}} \right] (m \delta \tau)^{(w-1)} \] (43)
\[ = \frac{k^{(1+w)} \Gamma(2k + 1 - w)}{k!^2} \left[ \frac{(n/m)^{1-w}(1 + n/m)^k}{(2 + n/m)^{2k+1-w}} \right] (n \delta \tau)^{(w-1)} \] (44)
These coefficients are all finite and well defined except when both \( m = n = 0 \). An important property to note is that for any nonzero \( n \), \( A^0_n = 0 \), but for nonzero \( m \), \( A^m_0 \neq 0 \).

To compute \( p_n \) for any given \( n \), we have to sum over the contributions from all its predecessors; that is, sum the coefficients \( A^m_n \) over all \( m \). Note from eq. 43 that for \( m \gg n \), \( A^m_n \sim m^{(w-1)} \). The summation will converge for any \( w < 0 \) resulting in a finite \( p_n \), while for \( w \gg 0 \), the summation will diverge. To obtain a qualitative functional form of \( p_n \), it is convenient to approximate the summation by an integral.
\[ p_n \sim (n \delta \tau)^{(w-1)} \sum_{m=1}^{\infty} \frac{(n/m)^{1-w}(1 + n/m)^k}{(2 + n/m)^{2k+1-w}} \longrightarrow (n \delta \tau)^{(w-1)} \int_0^\infty dm \frac{(n/m)^{1-w}(1 + n/m)^k}{(2 + n/m)^{2k+1-w}} \]
\[ = n^w(\delta \tau)^{(w-1)} \int_0^\infty dz \frac{z^{(w-1)}(1 + z^{-1})^k}{(2 + z^{-1})^{2k+1-w}} \] (45)
In the last step, \( m/n \) is relabeled as \( z \). This integral converges for all \( w < 0 \) and diverges for \( w \gg 0 \). Thus we see that for \( w < 0 \), \( p_n \sim n^w(\delta \tau)^{(w-1)} \).
Contiguity calculations

In order to calculate the contiguity effect caused by the recalled item in isolation from the recency effect due to the preceding state of history, let us now assume that a lot of time has elapsed since the end of the study list. Our aim is to predict the probability that other list items would be recalled next. Let the recalled item be labeled as 0, and let us relabel the neighboring study list items by successive integers (Figure 22, bottom). Now we can calculate the probability that each of the neighboring items will be recalled subsequently.

It will be necessary to keep track of the activation for items that followed \( f_0 \), namely \( f_0 + 1 \), \( f_0 + 2 \), and so on, separately from the activation of the preceded \( f_0 \), namely \( f_0 - 1 \), \( f_0 - 2 \), etc.

Immediately after repetition of \( f_0 \), the input consists of two components, the prior state of the temporal history, \( |T_0\rangle \), weighted by \( \gamma \), and the item itself located at \( \tau = 0 \), weighted by \( 1 - \gamma \). Note that the latter component is not an effective cue for recall of any other list item. That is because no item in the list was experienced with the recalled item \( f_0 \) s in the past. The following item was experienced with the recalled item \( \delta \tau \) seconds in the past. On the other hand, \( |T_0\rangle \) is an effective cue for other list items immediately after recovery.\(^{20}\)

The prediction for an item in the forward direction due to \( |T_0\rangle \) is given by

\[
p_n = \gamma \langle T_n | T_0 \rangle = \gamma \sum_{m=1}^{\infty} A_m^n
\]

(46)

If we wait for some time following the context reinstatement, the lag-CRP will become asymmetric favoring the forward direction. This happens because both \( T_0 \) and the item representation entering at \( \tau = 0 \) evolve in time. First let’s consider the effect on \( T_0 \). Consider the state obtained by evolving \( T_0 \) for a time \( \tau = r \delta \tau \). Let us denote it by \( T_{0+r} \).

Then the prediction for the \( n \)-th item in the forward and backward direction can be deduced to be

\[
p_{-n} = \gamma \langle T_{-n} | T_{0+r} \rangle = \gamma \sum_{m=1}^{\infty} A_m^{n+r}
\]

\[
p_n = \gamma \langle T_n | T_{0+r} \rangle = \gamma \sum_{m=1+\min(n,r)}^{\infty} A_m^{n-r}
\]

(47)

(48)

In addition, the pattern of activity in the \( T \) column corresponding to the recalled item \( f_0 \) will generate further asymmetry in the forward direction because the activity in this \( T \) column is only associated with the items in the forward direction. The prediction for the item \( f_n \) due to this activity is simply \( A_{n-r}^r \) when \( n > r \) and is \( A_{r-n}^n \) when \( r > n \). When \( r = 0 \), this quantity is zero, hence immediately after the recall the CRP would still be symmetric and the asymmetry would grow with time. Combining the prediction from

\(^{20}\)For simplicity, we are assuming that \( |T_0\rangle \) is recovered instantaneously.
this part along with the prediction from the reinstated context, we have

\[ p_{-n} = \gamma \sum_{m=1}^{\infty} A^m_{n+r} \quad \text{and} \quad (49) \]

\[ p_n = \gamma \sum_{m=1+\mu}^{\infty} A^m_{|n-r|} + (1 - \gamma)A^\mu_{|n-r|} \quad (50) \]

where \( \mu = \min(n, r) \), the minimum of the numbers \( n \) and \( r \).