Published in final edited form as:

Brain Res. 2010 December 13; 1365: 3–17. doi:10.1016/j.brainres.2010.07.045.

Timing using temporal context

Karthik H. Shankar and Marc W. Howard

Department of Psychology, Syracuse University

Abstract

We present a memory model that explicitly constructs and stores the temporal information about when a stimulus was encountered in the past. The temporal information is constructed from a set of temporal context vectors adapted from the temporal context model (TCM). These vectors are leaky integrators that could be constructed from a population of persistently-firing cells. An array of temporal context vectors with different decay rates calculates the Laplace transform of real time events. Simple bands of feedforward excitatory and inhibitory connections from these temporal context vectors enable another population of cells, *timing cells*. These timing cells approximately reconstruct the entire temporal history of past events. The temporal representation of events farther in the past is less accurate than for more recent events. This history-reconstruction procedure, which we refer to as timing from inverse Laplace transform (TILT), displays a scalar property with respect to the accuracy of reconstruction. When incorporated into a simple associative memory framework, we show that TILT predicts well-timed peak responses and the Weber law property, like that observed in interval timing tasks and classical conditioning experiments.

Keywords

Timing; Weber's law; trace conditioning; episodic memory

1. Introduction

Timing the interval between two events is one of the basic cognitive capacities we all share. This has been rigorously studied in a wide variety of classical conditioning experiments on animals [1,2] and explicit interval timing experiments on humans and animals [3,4,5,6]. One basic finding of these experiments is scalar variability in the underlying timing distributions. Suppose that subjects are trained to reproduce a time interval of a given duration, say d_o . The reproduced duration d generally forms a smooth probability distribution peaked approximately at d_o . Moreover the data shows that the standard deviation of the response distribution is proportional to d_o . That is the ratio of the interval to be timed and the standard deviation of the distribution of responses is approximately constant, a manifestation of Weber's law [7]. More specifically, the response distributions for different values of d_o overlap when they are scaled linearly. This is referred to as the scalar property. When the interval to be timed is short, the peak in the response distribution is narrow and the estimated duration is more accurate than when the interval to be timed is long. Superficially this appears fairly intuitive, but the underlying scalar property has very important implications for models of timing. Similar features are observed in classical conditioning experiments where animals are trained with a conditioned stimulus (CS) followed by an unconditioned

stimulus (US) after a latency period. It is observed that the peak of the conditioned response (CR), which we can think of as a measure of the animal's anticipation of the US, approximately matches the reinforcement latency during training. In addition, the time distribution of the CR activity approximately exhibits the scalar property described above [1,2].

In order to model these and related tasks, we first need an efficient timing mechanism. This timing mechanism then needs to be integrated with a memory mechanism in order to store and retrieve the timing information. It has been argued that in the 10 to 100 milliseconds range, relevant to speech and motor processing, the dynamically evolving pattern of activity in a spatially distributed network of neurons, is intrinsically sufficient as a timing mechanism [10], and hence it is unnecessary to postulate a specialized mechanism for timing. However, for longer time scales of the order of seconds to minutes, it seems necessary to have a specialized mechanism. There are many timing models developed over decades involving a variety of specialized mechanisms. They can be divided into two broad classes. See [8,9,10,11,12] for reviews.

The more prominent class of models of timing relies on an internal clock-like mechanism. Models in this class use different mechanisms to construct a scalar representation of elapsed time. Some [7,13,14] use a pacemaker whose pulses will be accumulated to represent perceived time, while others [15,16,17] use a population of neural oscillators of different frequencies. Still others [18,19] use a distributed idea of detecting the coincidental activity of different neural populations to represent the ticks of the internal clock.

The other class of models posit a distributed population of neural units, each of which responds to an external stimulus with a different latency. A straightforward approach is to use tapped delay lines [20] or chained connectivity between late spiking neurons [21]. In these models, the delays accumulated while traversing through each link of the chain add up, thereby making the different links of the chain respond with different latencies. A more sophisticated way to accomplish the same goal is to require the different members of the population to be intrinsically different and react to an external stimulus at different rates. The spectral timing model [22,23] and multi-time scale (MTS) theory [24] both share this property. MTS, for instance, assumes a cascade of leaky integrators where the activity in each unit exponentially decays following a stimulus with a distinct decay rate.

In this paper, we construct a timing mechanism in the framework of the temporal context model (TCM), an associative memory model that has been extensively applied to problems in episodic recall [25,26,27]. This timing model falls into the second class of timing models—it has no explicit clock system like a pacemaker or synchronous oscillators. Instead, the model requires a population of persistently firing neurons with a range of decay rates, similar in many respects to the cascade of leaky integrators of MTS [24]. We show that this population of leaky integrators implements the Laplace transform of the stimulus sequence. Using this insight, we approximate the inversion of the Laplace transform, constructing a separate population of “timing cells”. We refer to this procedure as timing from inverse Laplace transform, TILT. The approximation of the inverse Laplace transform can be accomplished using bands of alternating feedforward excitation and inhibition from the leaky integrators. In effect, the leaky integrators implement the Laplace transform of the stimulus history and the timing cells approximately invert this Laplace transform, thus generating an approximate reconstruction of the stimulus history. Each of the timing cells responds with peak activity at a different delay following a stimulus. The effect of this inversion is thus not unlike that generated by the spectral timing model or the delay line models. However, it turns out that the activity across the timing cells at any instant precisely shows the scalar property. When integrated into an analog of TCM's learning and retrieval

mechanisms, the model generates a prediction of the immediate future that reflects prior learning experiences. Rather than developing a detailed model of behavior, the focus of this paper will be on describing the qualitative features of the proposed timing mechanism. However, we do demonstrate that a simple behavioral model derived from this prediction qualitatively exhibits the Weber law property at the behavioral level.

We start with a brief description of the encoding and retrieval mechanisms of TCM. Following that, we construct the timing mechanism and discuss its neural representation. Finally, we integrate the timing mechanism with an analog of the learning and retrieval rules of TCM to qualitatively account for behavioral aspects of timing observed in classical conditioning experiments.

2. TCM

The initial goal of TCM was to account for the recency and contiguity effects observed in episodic recall tasks. The recency effect refers to the finding that, all other things being equal, memory is better for more recently experienced information. The contiguity effect refers to the finding that, all other things being equal, items experienced close together in time become associated such that when one comes to mind it tends to bring the other to mind as well [28]. The basic idea of TCM is that when a sequence of stimuli is presented successively, each stimulus is associated with a gradually-varying context state. Any two stimuli that have been experienced in close temporal proximity, though not directly associated, are indirectly linked as a consequence of associations to similar contexts.

The architecture of TCM can be formalized in terms of a two layer network comprised of a stimulus layer and a context layer with bidirectional connections between them as shown in Figure 1. Each node in the stimulus layer denoted by $\mathbf{f}$ corresponds to a unique stimulus. The external input drives the activity in this layer. At any instant, only the specific node corresponding to the stimulus being perceived is active in this layer. Each of these nodes can be viewed as a distributed set of neurons and different nodes could potentially share some neurons. But we assume this overlap to be sparse; for expository simplicity we treat each node in the stimulus layer as a grandmother cell for a specific stimulus.

The nodes in the context layer denoted by $\mathbf{t}$ are for simplicity assumed to be in one to one correspondence with the nodes in the stimulus layer, and the activity in the stimulus layer drives the activity in the context layer via an operator $\mathbf{C}$. In general the nodes in the stimulus and context layers need not be in one to one correspondence. From a pragmatic point of view, we would expect that the number of context nodes is far less than the number of distinct stimulus nodes, in which case we would require the $\mathbf{C}$ operator to map multiple nodes from the stimulus layer to the same node in the context layer. But here we shall take the stimulus and context nodes to be in one to one correspondence, and $\mathbf{C}$ to be the identity map for simplicity. When a stimulus node is activated, $\mathbf{C}$ immediately activates the corresponding context node. Unlike the stimulus node, the activity in the context node is not abruptly turned off when the next stimulus is presented. Instead it gradually decays. As the stimuli are sequentially presented, the activity in the context layer gradually drifts. If $\mathbf{t}_i$ is the context activity just before the presentation of a stimulus, and if $\mathbf{t}^{\text{IN}}$ is the activity induced by the stimulus through $\mathbf{C}$, then the context activity immediately following the presentation of that stimulus is given by

$$\mathbf{t}_{i+1} = \rho \mathbf{t}_i + \beta \mathbf{t}^{\text{IN}} \quad (1)$$

Here ρ is a number between zero and one that determines the rate of decay of the activity of a context node once it is activated, and β denotes the strength of activation of a context node by $\mathbf{C}$. To maintain a normalized context activity at all times, β can be chosen to appropriately depend on ρ and/or the relationship between $\mathbf{t}_i$ and $\mathbf{t}^{\text{IN}}$. In this paper we shall simply choose $\beta = 1$. Note that in the above equation, the context activity is expressed as a vector, which we shall sometimes refer to as the temporal context vector.

2.1. Encoding in Memory

Each of the context nodes is connected to each of the stimulus nodes, and the connection weights are denoted by the operator $\mathbf{M}$ which can be viewed as a matrix (see Figure 1). Each entry of $\mathbf{M}$ holds the connection strength between a specific context node and a specific stimulus node. When a stimulus node is activated, the connectivity of all the active context nodes to it is strengthened in a Hebbian fashion. In effect, each row of $\mathbf{M}$ corresponds to a specific stimulus, and when that stimulus is experienced, that row of $\mathbf{M}$ is additively updated with the $\mathbf{t}$ vector at that moment.

2.2. Retrieval from Memory

At any moment, the activity in the context layer can induce activity in the stimulus layer *via* the connections learned in $\mathbf{M}$. This internally stimulated activity in the stimulus layer corresponds to “what comes to mind next”. This is formalized by defining the induced activity as a prediction vector $\mathbf{p}$ in the stimulus layer to be the product of $\mathbf{M}$ and current context activity.

$$\mathbf{p} = \mathbf{M} \cdot \mathbf{t}_{\text{cue}} \quad (2)$$

This vector is heuristically interpreted as representing the probability that any stimulus would be experienced next (for details, see [29]). The component of $\mathbf{p}$ corresponding to a particular stimulus will be activated to the extent the cuing context $\mathbf{t}_{\text{cue}}$ is similar to the context activity at the time that stimulus was encountered. This is immediately obvious when we note that the above equation is a product of a matrix with a column vector. For example, in the right panel of Figure 1, the cue context is more similar to $\mathbf{t}_3$ than to $\mathbf{t}_2$, and more similar to $\mathbf{t}_2$ than to $\mathbf{t}_1$. As a result, the component of $\mathbf{p}$ corresponding to *pen* is higher than the component corresponding to *cat* which in turn is higher than the component corresponding to *tree*. This retrieval rule along with the gradually drifting property of context activity immediately yields the recency effect. In TCM, the stimulus-induced drift in the context activity, $\mathbf{t}^{\text{IN}}$, reflects the prior experience with the stimulus; this property is essential to generate a satisfactory contiguity effect in episodic recall. But for this paper, we shall consider $\mathbf{t}^{\text{IN}}$ to be the same each time a given stimulus is experienced. With $\mathbf{C}$ being an identity map, $\mathbf{t}^{\text{IN}}$ is simply $\mathbf{f}$, the stimulus being presented.

2.3. Predicting the imminent future

Although TCM was initially proposed to capture the recency and contiguity effects observed in episodic recall, it turns out that this framework can also be adapted to learn structure from sequential stimuli and predict a subsequent stimulus based on the prior sequence [29]. To illustrate this, let us consider adopting TCM to learn a lengthy sequence of words which are not just randomly assembled lists as in most free recall studies, but are generated by a simple probabilistic generating function. Each unique stimulus will occur at many different positions in the sequence and therefore in a variety of different context states. These states will be linearly superposed and stored in the corresponding row of $\mathbf{M}$, thus making each row of $\mathbf{M}$ proportional to the average context in which the corresponding stimulus was

experienced. After training the model on the study sequence, the retrieval rule (Eq. 2) can exploit the statistics gathered in $\mathbf{M}$ to serve as a sequence predictor. Consider a short sequence of stimuli presented to the model at the end of training. The context activity induced by this sequence would form the cue context $\mathbf{t}_{cue}$ to generate the prediction vector $\mathbf{p}$ according to Eq. 2. If the cue context matches the average context of any stimulus stored in $\mathbf{M}$, that stimulus will be well predicted to fit into the subsequent position. Hence $\mathbf{p}$ can be interpreted as the prediction vector that predicts the subsequent stimuli to be generated by the generating function.

Considering the events encountered in one's life as a sequence of stimuli generated by the environment (which is hopefully a structured generating function), we expect the prediction vector $\mathbf{p}$ to roughly predict the imminent future based on the entire history of past experiences. It has been shown [29] that for sufficiently simple generating functions, this approach can be used to generate a statistically accurate prediction of future stimuli. This strategy can even be used to generate a reasonable approximation to the semantic structure of English when trained on a corpus of naturally-occurring text [30].

3. Timing using a set of temporal context vectors

Although TCM has been useful in many domains, it is ultimately limited. The temporal context vector discards explicit information about the time at which stimuli are experienced. This creates a number of problems that will remain intractable within TCM without significant extension. For example consider a classical trace conditioning paradigm. During training, the conditioned stimulus (CS) is followed by an unconditioned stimulus (US) after n time steps. Every time the US is experienced, the context vector contains a degraded representation of the CS which is stored in the row of $\mathbf{M}$ that corresponds to the US. That is, when the US is experienced, the context vector contains the input caused by the CS multiplied by a number ρ^n . After learning, the US will be predicted by the CS to the extent that the CS is present as a part of the context cue at test. Hence during test, when the CS is re-presented, the US will be maximally predicted immediately following the repetition of the CS and then decline as the CS activity gets degraded in the context after presentation of the CS. However, experiments show that animals and humans predict the US maximally n time steps after the CS rather than immediately after presentation of the CS [1]. TCM implicitly stores partial timing information in the form of appropriate degradation in context activity. However, behavioral responses in the above mentioned experiment are consistent with explicit storage of detailed timing information about the reinforcement delay. TCM as currently formulated is clearly insufficient to explain the timing of the CR observed behaviorally.

In this section we generalize the definition of temporal context so as to enable the representation to carry explicit information about the temporal relationships between stimuli. Instead of collapsing the entire history of experience into a single context vector with a fixed decay rate ρ , we use the history to construct a *set* of context vectors, each with a different decay rate spanning the range of allowed values of ρ . We first show that the set of context vectors essentially stores the Laplace transform of the stimulus history. As such, it contains detailed information about the history of stimulus presentation. We illustrate a simple technique for reconstructing the timing information from the activity distributed over the set of context vectors. We then show that the reconstruction procedure demonstrates the scalar property and briefly discuss its neural representation.

3.1. Temporal context drifting in real time

In previous formulations of TCM, the temporal context vector drifts only when there is input provided. If there is no input, the context vector remains fixed. However, such a "pure

interference” version of contextual evolution falls apart if there are inputs uncorrelated with nominal, experimentally relevant stimuli. For instance, uncontrolled environmental variables could be inputs to Eq. 1 when a nominal stimulus is not provided, or there could exist a genuinely stochastic component to the input [31]. We shall suppress these uncontrollable components of the context, but realize that these would cause the experimentally relevant components of the context vector to drift even in the absence of a nominal stimulus. So, in what follows we will write the equations as if context drifts due to time *per se*, but this is equivalent to assuming that a stochastic component, constant in magnitude, is provided to the context vector during delays.

The evolution equation for a temporal context vector (Eq. 1) is designed to induce a drift by a factor of ρ (in appropriate units) at each discrete time step. We shall now generalize the sequence from being defined on discrete time steps to being defined on continuous real time. The stimulus presented at each time step can be considered as a delta function on the real time (τ) axis centered around the appropriate time step. If a stimulus occurred multiple times in the past at lags $m, n, \dots$, then it can be inferred from Eq. 1 that the activity in the corresponding context node will be a superposition of the activity induced by each occurrence, $(\rho^m + \rho^n + \dots)$. We shall require this basic property to be preserved in the continuous time context evolution.

Let us represent the activity of the stimulus layer at each instant of time τ by $\mathbf{f}(\tau)$. $\mathbf{C}$ induces an instantaneous drift in the context layer which we denote by $\mathbf{t}^{\text{IN}}(\tau)$. For a given ρ , we shall define the context vector as a function of time τ to be

$$\mathbf{t}(\tau) = \int_{-\infty}^{\tau} \mathbf{t}^{\text{IN}}(\tau') \rho^{(\tau-\tau')} d\tau'. \quad (3)$$

Since $\mathbf{C}$ is taken to be an identity map, $\mathbf{t}^{\text{IN}}(\tau')$ is simply $\mathbf{f}(\tau')$. Taking $\mathbf{f}(\tau')$ to be a collection of delta functions around multiple time points in the past, representing multiple discrete occurrences, we can immediately verify that the required property mentioned in the previous paragraph is well preserved. Analogous to the discretized evolution equation (Eq. 1), the differential equation that guides the evolution of the context vector in continuous real time can then be derived to be

$$\frac{d\mathbf{t}}{d\tau} = (\ln \rho) \mathbf{t} + \mathbf{t}^{\text{IN}}(\tau). \quad (4)$$

Under the assumptions used here, $\mathbf{t}^{\text{IN}}(\tau) = \mathbf{f}(\tau)$, each component of the above vector equation is independent of the others and can be decoupled. Hence it suffices to just focus on a single component corresponding to one stimulus. From Eq. 4, we can note a few basic properties of the context activity. First, if a particular stimulus is not experienced at some instant, the corresponding component of $\mathbf{f}(\tau)$, and hence $\mathbf{t}^{\text{IN}}(\tau)$ is zero at that moment. This leads to a decrease in the corresponding component of the context vector: the second term on the RHS of Eq. 4 vanishes and the first term with $\ln \rho$ is negative, thus making the derivative negative. If that component of $\mathbf{f}(\tau)$ stays zero for long enough, then the corresponding component of $\mathbf{t}(\tau)$ would eventually decay to zero. Secondly, if the stimulus function was a constant (say 1) for sufficiently long time, then the activity of the corresponding context node will increase and saturate at $-1/\ln \rho$. Figure 2 illustrates these properties by plotting the context activity corresponding to a stimulus presented three times for different durations and intensities. To summarize, a single temporal context vector is closely related to a low-pass filter of the stimulus history function.

3.2. Reconstructing the history of events from the context activity at an instant

At any moment τ_o , the entire history of a particular stimulus up to that moment is integrated into the activity of the corresponding component of $\mathbf{t}(\tau_o)$, as denoted by Eq. 3. Figure 2 illustrates the activity of a $\mathbf{t}$ node corresponding to a stimulus whose presentation history is illustrated at the top of the figure. Let us refer to the activity of the corresponding $\mathbf{f}$ node by $f(\tau)$ and the activity of the corresponding $\mathbf{t}$ node by $t(\tau)$. Note that the dotted line demonstrates that the level of activity of the context node is identical at six separate occasions. The stimulus history at each of these six moments is clearly different, yet the context activity at all of them is the same. It would be impossible to reconstruct the entire history of a stimulus based on the context activity at any particular moment in time. This is simply because we cannot reconstruct an entire function $f(\tau)$, for all $\tau < \tau_o$ from a single number $t(\tau_o)$. Although the number $t(\tau_o)$ is not sufficient to reconstruct the history $f(\tau)$, we will see that a set of values of $t(\tau_o)$ constructed from multiple values of ρ can in principle be used to reconstruct the entire $f(\tau)$.

To this end, let us consider a set of temporal context vectors, each evolving according to Eq. 4 but with a distinct decay rate ρ . We shall assume that this set spans all values of ρ between zero and one. It may be helpful to visualize this set of temporal context vectors by arranging them one below the other, ordered by their ρ values and such that the components corresponding to each distinct stimulus in each of these vectors line up (see Figure 3). The context layer can now be thought of as a 2 dimensional sheet of nodes with each column responding to a specific stimulus. Alternatively, its state at each moment can be thought of as a vector-valued function of ρ , $\mathbf{t}(\rho)$.

It turns out that $\mathbf{t}(\rho)$ at any moment, distributed over all these vectors, is simply the Laplace transform of the entire stimulus history. To see this, let us define $s \equiv -\ln \rho$ and label each vector by its value of s ranging between 0 and ∞ . Now let us rewrite Eq. 3 in terms of s instead of ρ :

$$\mathbf{t}(s) = \int_{-\infty}^{\tau_o} e^{s(\tau' - \tau_o)} \mathbf{f}(\tau') d\tau' \quad (5)$$

Clearly $\mathbf{t}(s)$, the $\mathbf{t}$ layer activity at the moment τ_o , is the Laplace transform of $\mathbf{f}(\tau)$, the stimulus history. Knowing this is extremely useful—it means that the Laplace transform could be inverted to accurately recover $\mathbf{f}(\tau)$ for all $\tau < \tau_o$. This would constitute the detailed history of stimulus presentation up to time τ_o as a function of past time. Of course a model that provides a perfectly accurate description of all of stimulus history is not a good model for memory, which is nothing if not imprecise. Our goal is thus to approximately reconstruct $\mathbf{f}(\tau)$. Moreover, we desire that this reconstruction be more accurate for recent events than for less recent events. It will turn out that the approximation we use not only reconstructs more recent events more accurately, but that the error in the reconstruction obeys the scalar property.

The recipe we adopt for approximating the inversion of the Laplace transform is based on the work of Emil Post [32]. Let's first define the following function.

$$\mathbf{T}^*(\tau) = \frac{(-1)^k}{k!} s^{k+1} \mathbf{t}^{(k)}(s) : \text{where } s = -k / \tau^* \\ \equiv \mathbf{L}_k^{-1}[\mathbf{t}(s)]. \quad (6)$$

Here k is any positive integer and $\mathbf{t}^{(k)}(s)$ is the k -th derivative $\mathbf{t}(s)$ with respect to s . The symbol $\mathbf{L}_k^{-1}$ indicates that this is an approximation to the inverse Laplace transform

operation. Post proved that in the limit as $k \rightarrow \infty$, $\mathbf{L}_k^{-1}$ becomes the inverse Laplace transform for appropriate functions. We understand the variable $\tau \square$ as the *internal time* that ranges from 0 to $-\infty$, representing the entire past up to the present moment at $\tau \square = 0$.

For convenience, let us shift the time axis such that the present moment is at $\tau_o = 0$. It turns out that the stimulus history $\mathbf{f}(\tau)$ is well approximated by the function $\mathbf{T}(\tau \square)$. That is, with $\tau \square = \tau$, we have $\mathbf{T}(\tau \square) \approx \mathbf{f}(\tau)$. This approximation grows increasingly accurate as k increases. It can be shown [32] that in the limit $k \rightarrow \infty$, $\mathbf{T}(\tau \square)$ exactly matches $\mathbf{f}(\tau)$.

Figure 4 provides an image to help understand the relationship between $\mathbf{t}(\rho)$ and $\mathbf{T}(\tau \square)$. Let us arrange the temporal context vectors in rows sorted by their value of ρ , as in Figure 3. Now, as before, a column of this array of temporal context vectors corresponds to a single stimulus element. From Eq. 6, observe that the variable $\tau \square$ is in one to one correspondence with the decay constants (s or ρ) of the $\mathbf{t}$ nodes. The reconstructed stimuli history $\mathbf{T}(\tau \square)$ can be understood as a separate timing layer, where each node indexed by $\tau \square$ is in one to one correspondence with the nodes in the $\mathbf{t}$ -layer. In Figure 4 we have organized the columns of $\mathbf{T}(\tau \square)$ sorted by the value of $\tau \square$. Presentation of a stimulus will activate the appropriate $\mathbf{t}$ column which will in turn activate the nodes in the corresponding $\mathbf{T}$ column according Eq. 6. The pattern of activity distributed over each $\mathbf{T}$ column at any instant, will approximately represent the history of the corresponding stimulus.

Since this mechanism yields timing information about when a stimulus was encountered in the past, we refer to it as timing from inverse Laplace transform - TILT.

3.3. Properties of the reconstructed stimulus history

Having formally specified $\mathbf{T}(\tau \square)$ above, we here describe some of its properties. These are emergent properties that follow from the method of reconstruction described above. The temporal representation $\mathbf{T}(\tau \square)$ has a single parameter, k . As a general rule, the quality of the reconstruction improves with increasing k . Following the description of the properties of $\mathbf{T}(\tau \square)$ we will describe qualitatively how to compute the k -th derivatives necessary to construct $\mathbf{T}(\tau \square)$.

3.3.1. The Weber fraction and k —Consider four different stimuli (each of duration 0.1 sec) presented at different moments in the recent past, namely 2, 5, 10, and 20 seconds ago. Four different columns in the $\mathbf{t}$ -layer of Figure 4 corresponding to these stimuli will be active representing the Laplace transform of the corresponding stimulus history. Consequently four columns of the $\mathbf{T}$ -layer will be active representing the corresponding reconstructed stimulus history. $\mathbf{T}(\tau \square)$ calculated from Eq. 6 is plotted in Figure 5. The different curves corresponding to the different stimuli are labeled by the corresponding delay.

First, observe that each curve peaks when $\tau \square$ approximately matches the corresponding delay. For instance, the curve for the stimulus presented ten units in the past peaks about at $\tau \square = -10$. Second, observe that as the stimulus is pushed farther back in time, the function $\mathbf{T}(\tau \square)$ is weaker in magnitude and more spread out. For instance, the function representing the stimulus presented two units in the past is higher and more sharply peaked than the function representing the stimulus presented five units in the past. It turns out that the area under each curve is the same. Third, the coefficient of variation of each of the curves is a constant determined only by k as illustrated in the table. This illustrates the scalar property of the function $\mathbf{T}(\tau \square)$. All three of these properties hold for all fixed values of k . It is essential to note that the scalar property holds only when k is fixed; varying k in the $\mathbf{L}_k^{-1}$ procedure as we traverse down the column would destroy this property.

The underlying reason for the emergence of a precise scalar property is hidden in the mathematical form of $\mathbf{L}_k^{-1}$ and the form of $\mathbf{t}(s)$. This operation entails taking the k -th derivative of exponentially decaying functions. The formal scale-invariance of the approximate reconstruction can be analytically demonstrated, but it is beyond the scope of this paper. For now, let us content ourselves with noting that the Laplace transform is scale-free. A single temporal context vector has a specific value which sets a preferred scale. The set of all possible temporal context vectors has all possible values of ρ ; none is preferred. The scale-invariance of $\mathbf{T}(\tau)$ is made possible by the scale-free nature of $\mathbf{t}(s)$.

The curves on the left panel of Figure 5 are plotted for $k = 12$. It should be noted that the qualitative features of the model are the same for any fixed k . For higher values of k , all four curves in general become sharper; this can be seen from the table which shows that the coefficient of variation decreases with increasing k .¹ Hence the reconstructed history as a function of internal time becomes an increasingly accurate representation of the actual history as k increases. The primary computational cost of increasing k is that it involves computing higher derivatives of $\mathbf{t}(s)$.

3.3.2. Representing multiple presentations of a stimulus—Figure 5 describes the basic properties of $\mathbf{T}(\tau)$ when presented with a very simple stimulus history—a single presentation of a stimulus at a single time in the past. $\mathbf{T}(\tau)$ is also able to represent more complex stimulus histories with a veridicality that increases with k . For example, Figure 6 shows the representation of a stimulus presented at two distinct points in the recent past. Note that when k is relatively large, as in the left panel of Figure 6, $\mathbf{T}(\tau)$ clearly identifies both of the presentations of the stimulus, with the time of the more recent presentation represented with greater accuracy. However, when k is relatively small, the second peak in $\mathbf{T}(\tau)$ is barely discernible, appearing merely as an inflection in the curve. For any value of k , the model is able to clearly distinguish separate presentations of a single stimulus if the two presentations are separated by a sufficiently large temporal distance that can be determined from the Weber fraction of a single presentation (Figure 5) for that value of k .

3.3.3. Stimulus-specific, Delay-specific cells in T-layer—Thus far we have examined $\mathbf{T}$ as a function of τ , the internal time, at a particular moment in time with a specific stimulus history. Of course, $\mathbf{T}(\tau)$ will change from moment to moment because the history being represented at each moment is different, at minimum having shifted backward in internal time as physical time changes. As described before, $\mathbf{T}(\tau)$ should be understood as the activity across a population of cells, each representing a different value of τ , at a specific moment. To further illustrate the properties of the model, here we examine how cells coding for specific values of τ change their activity as physical time passes.

At each moment, the array of temporal context vectors $\mathbf{t}(s)$, is used to generate the representation $\mathbf{T}(\tau)$. Restricting our attention to a single stimulus, one column of $\mathbf{t}(s)$ is used at each moment in time to generate the corresponding column of $\mathbf{T}(\tau)$. The top panel of Figure 7 shows a function describing the times at which a particular stimulus is presented. The middle panel samples the column of the temporal context vector $\mathbf{t}(s)$ corresponding to the stimulus at two values of s . Recall that each value of s corresponds to a value of ρ via the relationship $s = -\ln \rho$. As illustrated previously, the cells in the temporal context vectors compute something very similar to a low-pass filter of the stimulus function with a specific value of ρ . Each cell in $\mathbf{t}(s)$ representing a specific value of s corresponds to a cell in $\mathbf{T}(\tau)$ representing a specific value of τ . The bottom panel shows the response of two cells in $\mathbf{T}(\tau)$ derived from the activity of the temporal context vectors at each moment. With $k = 12$

¹As an aside, the distribution also become more nearly symmetric with increasing k .

the bottom panel of Figure 7 shows the response of cells in $\mathbf{T}(\tau_k)$ with $\tau_k = -3$ and $\tau_k = -6$ corresponding to the two cells in $\mathbf{t}(s)$, with $s = 2$ ($\rho = 0.14$) and $s = 4$ ($\rho = 0.02$), from the middle panel. Note that the activity of the $\mathbf{t}$ cells always peaks at the stimulus offset while the $\mathbf{T}$ cells peak roughly at an appropriate delay after the stimulus offset. That is, the activity of the cell in $\mathbf{T}$ with $\tau_k = -3$ peaks roughly 3 units after each presentation of the stimulus. In contrast, the $\mathbf{T}$ cell with $\tau_k = -6$ peaks roughly 6 units after each presentation of the stimulus.

Another important point illustrated by Figure 7 is that the temporal response of cells representing more remote values of τ_k lasts for a longer duration than that of cells corresponding to less remote values of τ_k . For instance, the activity of the $\tau_k = -6$ cell is more spread out in time than the activity of the $\tau_k = -3$ cell. In fact, explicit analytical calculations reveal that the spread in activity of a specific τ_k cell is directly proportional to τ_k . In other words, the model predicts that the scalar property should hold for the temporal response of the $\mathbf{T}$ layer cells.

3.4. Neural plausibility of the representation of internal time

As discussed in prior publications [26,33], temporal context vectors represented by the $\mathbf{t}$ -layer could be computed in a straightforward way using populations of persistent-firing integrator cells [34,35,36,37] equipped with divisive normalization [38]. Presumably network properties would control the value of ρ observed in any particular set of coupled integrator cells. It is also worth noting that temporal correlation in the firing of hippocampal cells over the scale of minutes has been observed in rats [39]. This suggests that cells that persist in firing over long periods of time may be present in the medial temporal lobe.

The primary characteristic of cells in the $\mathbf{T}$ -layer is that they respond at a characteristic delay to the stimulus that they code for. There is mounting evidence that this might reflect a primary function of the hippocampus. Pastalkova and colleagues [40] observed “time cells” in the hippocampus that exhibited a similar property. In their experiment, rats were trained to run on a maze. In one part of the maze, they remained stationary in allocentric space while they ran on a wheel for a fixed duration. During the delay interval, cells fired for a circumscribed period of time. Curiously, despite the fact that the animal was not moving during this period, these time cells displayed many of the same firing characteristics as place cells, suggesting that the spatial and temporal aspects of hippocampal function share a common computational basis.

The model makes three basic predictions about time cells. The model predicts that time cells 1) should be observed in non-spatial tasks, 2) should differentiate non-spatial stimuli presented in the recent past, and 3) should obey the scalar property. Recent evidence has shown dramatic evidence for the first two predictions [41]. Although the third prediction has not been confirmed quantitatively, it is certainly the case that the variability in the time at which these time cells fire, goes up with the delay they code for. It should be noted that the clock models of timing [7,15,19], do not predict such time cells that respond to stimuli at specific latencies. On the other hand, it should be noted that the existence of time cells does not uniquely support our model. Most models belonging to the second class mentioned in the introduction, for example [23,21], would also predict the existence of such time cells.

Overall, there appears to be some evidence for persistent-firing cells that could give rise to the $\mathbf{t}$ layer cells. There is also some evidence that the brain computes something similar to $\mathbf{T}$. However, the model presented here proposes that the latter is calculated from the former by a specific mechanism, the operator $\mathbf{L}_k^{-1}$. At first glance $\mathbf{L}_k^{-1}$ may seem neurally implausible because it involves computing the k -th derivative of the context activity along each column.

However, it turns out that calculation of the k -th derivative can be achieved through bands of feedforward excitation and inhibition between neighboring context cells.

Let us assume that cells composing the temporal context vector are topologically arranged so that their decay constants change monotonically from one cell to the next along a direction, as in Figure 3 and 4. First note that the context activity as a function of decay rate $t(s)$ can be considered continuous only at a very coarse-grained level. Zooming in, the individual cells are discrete by nature, and so will be their decay rates. At the cellular scale where the underlying space of decay rates s is discrete, the derivative of the function $t(s)$ is simply proportional to the difference in the activities of neighboring cells with slightly different values of s . More generally, the k -th derivative of the function is simply proportional to a linear combination of the activity of k neighboring cells. The calculation of the exact weights of this linear combination is somewhat involved. Here we highlight the key features of this connectivity using Figure 8. Note that the contributions from the k neighbors are inhibitory and excitatory in an alternating fashion and the magnitude of contribution from a neighbor falls off as the distance to that neighbor increases.

4. Timing at the behavioral level

The foregoing section described the timing mechanism TILT, the mathematical properties of $\mathbf{T}$ and explored the possibility that it is supported by neural evidence. Here we sketch a small number of behavioral applications to illustrate how $\mathbf{T}$ might contribute to timing behavior. For this we propose a simple associative memory model with encoding and retrieval rules adapted from TCM. We should emphasize that the goal of this is not to provide a detailed behavioral model of these tasks, but to illustrate the ability of the model of timing behavior to solve problems in accounting for behavioral data, especially the findings that would not be possible to account for within TCM.

We start by sketching an associative memory model that uses $\mathbf{T}$ as a cue for retrieval analogous to the way the temporal context vector is utilized in TCM. We then demonstrate that this behavioral model retains the scalar property at the behavioral level and then show that it is sufficient to exhibit well-timed behavior in a classical conditioning task.

4.1. An associative memory model using the distributed representation of internal time

Recall that in TCM, the $\mathbf{t}$ layer activity at the moment any stimulus is presented is encoded in the appropriate row of the matrix $\mathbf{M}$. In order to enable $\mathbf{T}$ to affect behavior, we have to modify the encoding process. In the associative memory model we develop here, the $\mathbf{t}$ layer is not directly associated with the $\mathbf{f}$ layer. Instead the $\mathbf{t}$ layer at each moment constructs $\mathbf{T}$ which is then associated with the stimulus in the $\mathbf{f}$ layer. This association will also be denoted by $\mathbf{M}$ (see Figure 9). However, rather than a vector, $\mathbf{T}$ is a vector-valued function of τ . As described earlier, if one considers τ to be discretely represented by separate cells, then $\mathbf{T}$ at any moment can be thought of as a two-dimensional sheet of nodes, with a stimulus dimension and a τ dimension. Thus $\mathbf{M}$ is no longer a matrix, but a tensor of rank three. Let us denote the associations in $\mathbf{M}$ for a given τ by $\mathbf{M}(\tau)$, which is a matrix. Each row of $\mathbf{M}(\tau)$ corresponds to a unique stimulus, and when any stimulus is encountered, the vector $\mathbf{T}(\tau)$ at that instant is stored in the appropriate row.

At the retrieval stage, $\mathbf{M}$ acts on the existing activity in the $\mathbf{T}$ layer to generate activity in the $\mathbf{f}$ layer. We refer to the output of this retrieval as $\mathbf{p}$ in analogy to Eq. 2. $\mathbf{p}$ can be understood as a prediction for what stimulus will be experienced at the next moment. In analogy to TCM, each stimulus is predicted to the extent the $\mathbf{T}$ state used to cue $\mathbf{M}$ resembles the $\mathbf{T}$ state when that stimulus was originally presented. However, now the similarity of the two-dimensional $\mathbf{T}$ states must be evaluated by integrating over τ :

$$\mathbf{p} = \int_{-\infty}^0 \mathbf{M}(\tau) \cdot \mathbf{T}(\tau) g(\tau) d\tau \quad (7)$$

The integral over τ means that information from all timescales is incorporated in generating the prediction. The cell density function $g(\tau)$ denotes the number of cells representing any particular value of τ . Since τ ranges from 0 to $-\infty$, and since there cannot be infinite number of cells, it is reasonable to assume that $g(\tau)$ will decay to zero as τ goes to $-\infty$. However, in the simulations that follow, we will assume that $g(\tau)$ is a constant as the most minimal assumption we could make.

4.2. The scalar property is observed in the prediction at the behavioral level

In the previous sections we have demonstrated that the activity in the $\mathbf{T}$ layer exhibits the scalar property. Because of the ubiquity of the scalar property in timing behavior [14], the existence of this property in $\mathbf{T}$ is nontrivial. However, it remains to be seen if the scalar property transfers to the behavioral level with the associative memory model we have just specified. Although one can construct arbitrarily complex behavioral models, the prediction vector $\mathbf{p}$ at any moment should largely control behavior at that instant. We content ourselves here with demonstrating that $\mathbf{p}$ retains the scalar property.

Let us examine the prediction $\mathbf{p}$ in a simple interval estimation paradigm where the duration between a START signal and a STOP signal has to be estimated. Consider the situation when the START signal is followed by the STOP signal after a delay d_o . At any moment after the START signal, only the column corresponding to START in both $\mathbf{t}$ and $\mathbf{T}$ layers will be active; all other columns will remain inactive. After a delay d_o , when STOP is encountered, let the $\mathbf{T}$ layer activity (of the START column) be $T_{d_o}(\tau)$. This gets stored in the STOP row of $\mathbf{M}(\tau)$. At the test (retrieval) stage, the START signal is re-presented and the task is to predict the moment STOP signal would have been repeated.

From Eq. 7, we can deduce that the STOP component of $\mathbf{p}$ at any instant, p_{stop} , is simply the integral of the product of $T_{d_o}(\tau)$ and the $\mathbf{T}$ layer activity at that instant:

$$p_{stop}(d) = \int_{-\infty}^0 T_{d_o}(\tau) \cdot T_d(\tau) g(\tau) d\tau \quad (8)$$

Here $T_d(\tau)$ is the $\mathbf{T}$ layer activity after a delay d following the START signal at test.

It turns out that the distribution $p_{stop}(d)$ becomes wider as d_o , the delay to be timed, increases. In addition, the peak value of p_{stop} is approximately at $d = d_o$ and the coefficient of variation is a constant for fixed k . In fact the function $p_{stop}(d)$ explicitly shows the scalar property, as summarized in Table 1.

From Eq. 8, note that the prediction is constructed from the product of the $\mathbf{T}$ layer activities representing different instances of internal time. Because $\mathbf{T}(\tau)$ exhibits the scalar property, this is transferred to the prediction vector that drives the behavior. It can be shown analytically that the scalar property holds for any power law choice of the cell number density $g(\tau)$. Although it is not our goal here to describe a detailed model of any particular behavioral task, because the scalar property is observed in the $\mathbf{p}$ vector, it can be transferred to behavioral predictions.

4.3. Timing of the response in Pavlovian conditioning

In Pavlovian conditioning, a conditioned stimulus (CS) is paired *via* some temporal relationship with an unconditioned stimulus (US) during learning. At test, the CS is re-presented and a conditioned response (CR) is observed, reflecting learning about the pairing between the CS and the US. Human and animal subjects can learn a variety of temporal relationships between stimuli, and respond in a way that reflects learning of the temporal relationships between the stimuli [14,42]. In an experiment on goldfish [1], during the learning phase, the US (shock) was presented 5 or 15 seconds after the onset of a CS (light). In the left side of Figure 10, the CR is plotted with respect to the delay since the onset of CS during the test phase. First, note that the peak CR approximately matches the reinforcement delay. Secondly, note that the CR becomes stronger as the the number of learning trials increases.

This pattern of results is qualitatively consistent with what would be predicted from the associative memory model based on **T** that we have described here. During learning, the US row of **M** stores the **T** activity representing the CS onset, which is peaked at the appropriate τ . During every learning trial, **M** gets reinforced by the same **T** activity, and thus gets stronger. Hence at the test phase, when the CS is re-presented, the component of the prediction vector corresponding to the US, p_{us} , automatically inherits the timing properties with the peak at an appropriate delay. Moreover, the fact that the **M** stores additional copies of **T** with additional learning trials immediately implies that p_{us} grows larger with learning trials.

The prediction vector by itself has several properties that render it inappropriate for treating it as a direct model of the CR. First, it starts out at zero where, in general, there is some spontaneous probability of a CR even prior to learning. Second, with **M** calculated as above, p_{us} grows without bound². Here we will use a minimal model to map p_{us} onto the probability of a CR, which can be behaviorally observed. We calculate the probability of a response as

$$\text{Probability of response} = \frac{p_{us} + \theta}{p_{us} + \theta + \varphi} \quad (9)$$

Here θ and φ are free parameters that control the background rate of responding and the scale over which the probability of a response saturates. In the absence of the CS, p_{us} is zero, and the baseline response probability is $\theta/(\theta + \varphi)$. We shall take the probability of conditioned response to be simply Eq. 9 with the baseline response probability subtracted from it.

The experiment corresponding to the left panel of Figure 10 is simulated and the probability of conditioned response is plotted in the right panel. In these simulations θ , φ , and k are the only free parameters. As long as φ is much bigger than θ , the simulation results make a good qualitative match to the experimental data. In these simulations, $k = 4$, $\theta = 0.1$ and $\varphi = 1$.

A simplifying assumption used in these simulations is that only the onset of CS gets encoded in the **t** and **T** layers and hence only the onset of the CS contributes to the prediction of the US. If this assumption were to be exactly true, it would imply identical CRs for both delay conditioning and trace conditioning as long as the total time interval between the CS onset

²This can be avoided by adopting a Rescorla-Wagner learning rule rather than the simple Hebbian association used here. As is well-understood, such a learning rule has numerous advantages in conditioning paradigms.

and US is held constant. To more accurately model the various classical conditioning CS-US pairing paradigms, we should consider not just the CS-onset, but also the whole CS duration and the CS-offset, and associate each of them with the US. This would of course generate well-timed responses, and potentially also distinguish the results from various experimental paradigms. But this would require a more detailed behavior model (as opposed to Eq. 9) and more free parameters such as the relative saliencies of the CS-onset, CS-duration, and CS-offset.

A qualitative feature that is immediately visible from the model predictions in Figure 10 is that the response distribution is skewed. The skew predicted by the model is ameliorated by larger values of k . This qualitative prediction from the model should be observed for timing experiments as well. This is not necessarily a counterfactual prediction. Although symmetric response distributions are sometimes observed in human peak-interval timing studies, careful examination of the methods of these studies reveals that symmetric distributions are observed only when subjects are provided feedback about the distribution of their responses and instructed to provide symmetric responses—see for instance Experiments 1 and 2 of [3]. When feedback is omitted, the response distributions are dramatically asymmetric—see Experiment 3 of [3]. Unlike the extremely simple Pavlovian experiment modeled here, most tasks used to examine animals' and humans' ability to time intervals are far too complex to yield to the analytic approach used here.

5. General Discussion

In this paper, we have described a timing mechanism (TILT) and integrated it into a memory model based on TCM. The model starts with a set of temporal context vectors, like those used in TCM, each of which exponentially decays with a distinct rate. At any instant, the information distributed across these context nodes is the Laplace transform of the function describing the stimulus history up to that point. We show that an elegant approximation to the inverse Laplace transform can be used to approximately reconstruct the entire temporal history of the stimulus function. The reconstruction functions as a representation of the temporal history of stimulus presentations up to that point in time. This temporal representation has greater accuracy for stimuli recently experienced and a less accuracy for stimuli presented further back in time. It exhibits the scalar property at several levels, 1) in the representation of prior events across the columns of **T** cells, 2) in the real time activity of the **T** cells following the presentation of a stimulus and 3) in the temporal distribution of the predictions it makes at the behavioral level.

5.1. Comparison to other timing models

As set out in the introduction, there are two broad classes of models that have been used to account for timing, clock models and delayed firing models. Clock models do not have neurons that exhibit delayed responding to a stimuli at various latencies. Such responding has been observed in the hippocampus [41,40]. The clock models do not provide a good account of this phenomenon. To the extent this phenomenon is important in timing at the behavioral level, this argues against clock-like models of timing. The temporal activity of **T** cells described by our model (see Figure 7) resembles to a large extent the activity of cells in the delayed firing models [21,20,23,22,24], although the underlying mechanisms are very different from those in these other models.

The most dramatic point of distinction between the current approach and previous models of timing behavior that rely on delayed firing comes from the fact that this model is derived from TCM, a model that has been extensively applied to modeling episodic memory [25,27,43] and semantic memory [29,30]. This leads naturally to the hypothesis that temporal effects in episodic recall, i.e., recency and contiguity effects, share a common

origin with the source of timing behavior. While other authors have certainly noted the connection between trace conditioning phenomena and other memory functions attributable to the hippocampus [44,45,46], previous quantitative models of timing behavior have not made the connection between timing and temporal effects in episodic recall. To the extent models have attempted to account for both classes of phenomena, they ascribed these to different sources. For instance, in spectral resonance theory, the recency effect in episodic recall emerges from the short term memory [47], while the timing behavior in classical conditioning emerges from the time varying activity of a different population of neural units [23].

5.2. Unifying episodic memory models and timing models

It remains to be seen if a strong connection between temporal effects in episodic memory and timing behavior is an advantageous property for a model to have or not. It is certainly the case that the present timing model provides a number of dramatic advantages over prior models of episodic recall, notably TCM, that would potentially enable it to account for a broad variety of phenomena. For instance, the judgment of recency task [48] could be accomplished in a relatively straightforward manner by reading off the information about the time of occurrence of prior events that is stored in the **T** layer. This method predicts that the confusability of two events separated by a fixed time interval should increase as the delay to the events increases, consistent with empirical findings [48]. The ability of **T** to separately represent multiple presentations of an item (see Figure 6) enables a behavioral model based on this to effectively function as a multitrace model [49,50]. This opens the possibility of a model of episodic memory based on **T** to account for numerous effects that would be extremely challenging for TCM [51,52,53]. Finally, consider the advantages of the **T** layer in developing a model of sequential learning. It has been shown that temporal context can be adapted to enable learning of stimulus generating functions as long as the “language” to be learned is sufficiently simple [29]. In particular a model of sequential learning based on just the **t** is sufficient to allow learning of bigram languages. Now that **T** includes explicit information about the time in the past at which stimuli were presented, it should enable learning of much more elaborate languages, perhaps even approaching the complexity of natural languages.

Adopting the **T** representation to construct a timing-memory model could not only advance our understanding of episodic memory performance, but could also help elaborate the ability to account for second order trace conditioning effects. TCM has been shown to develop compressed stimulus representations on the basis of temporal relationships among stimuli [26,54,29,30]. In the current paper we have assumed for simplicity that the input to the context vector caused by a stimulus is a constant. In TCM, this is not in general the case. Repeated stimuli have the capacity to recover the prior temporal context in which they were experienced. This enables the model to account for such phenomena as the ability to learn associations between stimuli that were never experienced in temporal proximity. For instance, suppose that a subject is presented with the double function pairs A–B and, much later B–C. Although A and C are not experienced in temporal proximity, they become associated to one another [55,56,57], presumably because both are experienced in the context of B. In TCM, this happens as a consequence of recovery of temporal context. For related accounts of similar phenomena, see [58,59,60,61,62]. This ability to integrate temporally discontinuous learning episodes to learn relationships among items not presented together, when combined with the inherently temporal representation described here, may enable an account of some otherwise puzzling results from second order conditioning studies. Ralph Miller and colleagues [63,64,65] have demonstrated that animals not only show well-timed conditioned responses, but also seem to construct temporal maps that integrate temporal information between stimuli that were never experienced in temporal

proximity. For example, if CS1 and CS2 are paired in the first phase of learning and CS2 and the US are paired in the second phase of learning, the animals show a conditioned response to CS1 even though it was never paired with US. The pattern of conditioned responding across conditions suggests that animals not only learn that CS2 and the US are associated to one another, but learn a particular temporal relationship between the two stimuli (e.g., CS2 precedes the US), despite the fact they are never presented in temporal proximity. It is possible that these apparently puzzling phenomena are tractable if the timing mechanism described here were combined with the ability to generalize across distinct learning episodes that comes from allowing the inputs to the temporal context vectors to change with learning.

Acknowledgments

The authors gratefully acknowledge support from AFOSR award FA9550-10-1-0149, and NIH award MH069938-01 and useful discussions with Brad Wyble, Ralph Miller, Jun Zhang, Howard Eichenbaum, Mike Hasselmo, Tom Brown, Steve Grossberg and Amy Criss.

References

1. Drew MR, Couvillon PA, Zupan B, Cooke A, Balsam P. Temporal control of conditioned responding in goldfish. *Journal of Experimental Psychology: Animal Behavior Processes*. 2005; 31:31–39. [PubMed: 15656725]
2. Smith MC. CS-US interval and US intensity in classical conditioning of rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*. 1968; 3:679–687. [PubMed: 5721496]
3. Rakitin BC, Gibbon J, Penny TB, Malapani C, Hinton SC, Meck WH. Scalar expectancy theory and peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes*. 1998; 24:15–33. [PubMed: 9438963]
4. Ivry RB, Hazeltine RE. Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance*. 1995; 7:242–268.
5. Wearden J. Temporal generalization in humans. *Journal of Experimental Psychology*. 1992; 18:134–144. [PubMed: 1532184]
6. Roberts S. Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*. 1981; 7:242–268. [PubMed: 7252428]
7. Gibbon J. Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*. 1977; 84(3):279–325.
8. Gibbon J, Malapani C, Dale CL, Gallistel CR. Toward a neurobiology of temporal cognition: advances and challenges. *Current Opinion in Neurobiology*. 1997; 7:170–184. [PubMed: 9142762]
9. Miall, RC. Models of neural timing. In: Pastor, MA.; Artieda, J., editors. *Time, Internal Clocks and Movements*. Elsevier Science; Amsterdam: 1996. p. 69-94.
10. Mauk MD, Buonomano DV. The neural basis of temporal processing. *Annual Review of Neuroscience*. 2004; 27:307–340.
11. Eagleman DM. Human time perception and its illusions. *Current Opinion in Neurobiology*. 2008; 18:131–136. [PubMed: 18639634]
12. Ivry RB, Schlerf JE. Dedicated and intrinsic models of time perception. *Trends in Cognitive Science*. 2008; 12:273–280.
13. Church, RM. Properties of the internal clock. In: Gibbon, J.; Allan, L., editors. *Timing and Time Perception*. New York academy of sciences; New York: 1984. p. 566-582.
14. Gallistel CR, Gibbon J. Time, rate, and conditioning. *Psychological Review*. 2000; 107(2):289–344. [PubMed: 10789198]
15. Church RM, Broadbent H. Alternative representation of time, number and rate. *Cognition*. 1990; 37:55–81. [PubMed: 2269008]

16. Treisman M, Faulkner A, Naish PL, Brogan D. The internal clock: evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*. 1990; 19:705–743. [PubMed: 2130371]
17. Miall RC. The storage of time intervals using oscillating neurons. *Neural Computation*. 1990; 37:55–81.
18. Matell MS, Meck WH. Corticostriatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive Brain Research*. 2004; 21:139–170. [PubMed: 15464348]
19. Buhusi CV, Meck WH. What makes us tick? functional and neural mechanisms of interval timing, *Nature Reviews. Neuroscience*. 2005; 6:755–765. [PubMed: 16163383]
20. Moore JW, Choi JS. Conditioned response timing and integration in cerebellum. *Learning & Memory*. 1997; 4:116–129. [PubMed: 10456058]
21. Tieu KH, Keidel AL, McGann JP, Faulkner B, Brown TH. Perirhinal-amygdala circuit level computational model of temporal encoding in fear conditioning. *Psychobiology*. 1999; 27:1–25.
22. Grossberg S, Schmajuk NA. Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks*. 1989; 2:79–102.
23. Grossberg S, Merrill J. A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. *Cognitive Brain Research*. 1992; 1:3–38. [PubMed: 15497433]
24. Staddon JE, Chelaru IM, Higa JJ. Habituation, memory and the brain: The dynamics of interval timing. *Behavioural Processes*. 2002; 57:71–88. [PubMed: 11947990]
25. Howard MW, Kahana MJ. A distributed representation of temporal context. *Journal of Mathematical Psychology*. 2002; 46(3):269–299.
26. Howard MW, Fotedar MS, Datey AV, Hasselmo ME. The temporal context model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. *Psychological Review*. 2005; 112(1):75–116. [PubMed: 15631589]
27. Sederberg PB, Howard MW, Kahana MJ. A context-based theory of recency and contiguity in free recall. *Psychological Review*. 2008; 115:893–912. [PubMed: 18954208]
28. Kahana, MJ.; Howard, M.; Polyn, S. Associative processes in episodic memory. In: Roediger, HL., III; Byrne, J., editors. *Cognitive psychology of memory*, Vol. 2 of *Learning and Memory - A Comprehensive Reference*. Elsevier; Oxford: 2008. p. 476-490.
29. Shankar KH, Jagadisan UKK, Howard MW. Sequential learning using temporal context. *Journal of Mathematical Psychology*. 2009; 53:474–485.
30. Howard MW, Shankar KH, Jagadisan UKK. Constructing semantic representations from a gradually-changing representation of temporal context. *Topics in Cognitive Science*.
31. Howard MW, Kahana MJ, Wingfield A. Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin & Review*. 2006; 13:439–445. [PubMed: 17048728]
32. Post E. Generalized differentiation. *Transactions of the American Mathematical Society*. 1930; 32:723–781.
33. Howard MW, Natu VS. Place from time: Reconstructing position from the temporal context model. *Neural Networks*. 2005; 18:1150–1162. [PubMed: 16198538]
34. Egorov AV, Hamam BN, Fransén E, Hasselmo ME, Alonso AA. Graded persistent activity in entorhinal cortex neurons. *Nature*. 2002; 420(6912):173–8. [PubMed: 12432392]
35. Fransén E, Tahvildari B, Egorov AV, Hasselmo ME, Alonso AA. Mechanism of graded persistent cellular activity of entorhinal cortex layer V neurons. *Neuron*. 2006; 49(5):735–46. [PubMed: 16504948]
36. Egorov AV, Unsicker K, von Bohlen und Halbach O. Muscarinic control of graded persistent activity in lateral amygdala neurons. *European Journal of Neuroscience*. 2006; 24(11):3183–94. [PubMed: 17156379]
37. Bang S, Leung VL, Zhao Y, Boguszewski P, Tankhiwale AA, Brown TH. Role of perirhinal cortex in trace fear conditioning: Essential facts and theory. *Society for Neuroscience Abstracts*.
38. Chance FS, Abbott LF, Reyes AD. Gain modulation from background synaptic input. *Neuron*. 2002; 35(4):773–82. [PubMed: 12194875]

39. Manns JR, Howard MW, Eichenbaum HB. Gradual changes in hippocampal activity support remembering the order of events. *Neuron*. 2007; 56:530–540. [PubMed: 17988635]
40. Pastalkova E, Itskov V, Amarasingham A, Buzsaki G. Internally generated cell assembly sequences in the rat hippocampus. *Science*. 2008; 321(5894):1322–7. [PubMed: 18772431]
41. MacDonald CJ, Eichenbaum H. Hippocampal neurons disambiguate overlapping sequences of non-spatial events. *Society for Neuroscience Abstracts*. 2009:101.21.
42. Balsam PD, Gallistel CR. Temporal maps and informativeness in associative learning. *Trends in Neuroscience*. 32(2)
43. Polyn SM, Norman KA, Kahana MJ. A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*. 2009; 116:129–156. [PubMed: 19159151]
44. Rawlins JN. Associations across time: The hippocampus as a temporary memory store. *Behavioral and Brain Sciences*. 1985; 8:479–528.
45. Wallenstein GV, Eichenbaum HB, Hasselmo ME. The hippocampus as an associator of discontiguous events. *Trends in Neuroscience*. 1998; 21:317–323.
46. Clark RE, Squire LR. Classical conditioning and brain systems: the role of awareness. *Science*. 1998; 280(5360):77–81. [PubMed: 9525860]
47. Grossberg S, Pearson LR. Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: toward a unified theory of how the cerebral cortex works. *Psychological Review*. 2008; 115(3):677–732. [PubMed: 18729596]
48. Yntema DB, Trask FP. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*. 1963; 2:65–74.
49. Hintzman DL. Judgments of frequency and recognition memory in multiple-trace memory model. *Psychological Review*. 1988; 95:528–551.
50. Shiffrin RM, Steyvers M. A model for recognition memory: REM — retrieving effectively from memory. *Psychonomic Bulletin and Review*. 1997; 4:145–166. [PubMed: 21331823]
51. Hintzman DL, Block RA, Summers JJ. Contextual associations and memory for serial positions. *Journal of Experimental Psychology*. 1973; 97(2):220–229.
52. Hintzman DL, Block RA. Memory for the spacing of repetitions. *Journal of Experimental Psychology*. 1973; 99(1):70–74.
53. Hintzman DL. How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*. 2010; 38(1):102–15.
54. Rao, VA.; Howard, MW. Retrieved context and the discovery of semantic structure. In: Platt, J.; Koller, D.; Singer, Y.; Roweis, S., editors. *Advances in Neural Information Processing Systems 20*. MIT Press; Cambridge, MA: 2008. p. 1193–1200.
55. Slamecka NJ. An analysis of double-function lists. *Memory & Cognition*. 1976; 4:581–585.
56. Bunsey M, Eichenbaum HB. Conservation of hippocampal memory function in rats and humans. *Nature*. 1996; 379(6562):255–257. [PubMed: 8538790]
57. Howard MW, Jing B, Rao VA, Provyn JP, Datey AV. Bridging the gap: Transitive associations between items presented in similar temporal contexts, *Journal of Experimental Psychology: Learning, Memory and Cognition*. 2009; 35:391–407.
58. O'Reilly RC, Rudy JW. Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychological Review*. 2001; 108(2):311–345. [PubMed: 11381832]
59. Frank MJ, Rudy JW, O'Reilly RC. Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. *Hippocampus*. 2003; 13(3):341–54. [PubMed: 12722975]
60. Levy WB. A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*. 1996; 6:579–590. [PubMed: 9034847]
61. Wu, XB.; Levy, WB. A hippocampal-like neural network model solves the transitive inference problem. In: Bower, JM., editor. *Computational Neuroscience: Trends in Research*. Plenum Press; New York: 1998. p. 567–572.
62. Wu X, Levy WB. Simulating symbolic distance effects in the transitive inference problem. *Neurocomputing*. 2001; 38–40:1603–1610.

63. Cole RP, Barnet RC, Miller RR. Temporal encoding in trace conditioning. *Animal Learning & Behavior*. 1995; 23(2):144–153.
64. Arcediano F, Miller RR. Some constraints for models of timing: A temporal coding hypothesis perspective. *Learning and Motivation*. 2002; 33:105–123.
65. Arcediano F, Escobar M, Miller RR. Bidirectional associations in humans and rats. *Journal of Experimental Psychology: Animal Behavior Processes*. 2005; 31(3):301–18. [PubMed: 16045385]

Figure 1.

TCM architecture. The left panel shows two layers with nodes representing the stimulus layer and the context layer. The external input activates a single node in the stimulus layer. The operator **C** in turn activates the corresponding node in the context layer. But the other nodes in the context layer corresponding to the stimuli encountered in recent past are still active. This is represented by the shaded activity in different context nodes. The lighter the shading, the farther back in time the corresponding stimulus was experienced. At each moment the context activity is associated with the incoming stimulus and is stored in the matrix **M**. An example is illustrated in the right panel. The context activity gradually changes from t_1 to t_2 to t_3 as three stimuli are sequentially presented. The context activity prior to the experience of a stimulus gets stored in a row of **M** uniquely associated with that stimulus. Hence t_1 is stored in the *tree* row of **M**, t_2 is stored in the *cat* row and t_3 is stored in the *pen* row. The stored information in **M** can be accessed through the context activity at any point in the future. At the retrieval phase, if the context activity is given by t_{cue} , the rows of **M** which are similar to t_{cue} get strongly activated and the rows which are less similar to t_{cue} get less activated. In this example, we have chosen t_{cue} to be more similar to t_3 , hence we see that the component of **p** vector corresponding to *pen* is stronger than the other components.

Figure 2.

Temporal decay of continuous-time context activity. The top curve represents the stimulus presented thrice with different durations and intensities. The bottom curve represents the activity of the corresponding context node. We have taken $\rho = 0.3$. The dotted line intersects the curve representing the context activity at six points, indicating that the context activity at these points is the same despite different stimulus history preceding each point. The y axis has arbitrary units and the two curves are not drawn to scale.

Figure 3.

Schematic representation of multiple context vectors stacked together. The different vectors are ordered by their ρ or s values with components corresponding to each stimulus lined up. All the nodes within each column of the **t** layer are activated by a specific **f** node. As an illustration, two columns of **t** are shaded in concordance with their corresponding **f** node.

Figure 4.

Schematic description of the one to one mapping between the context layer $\mathbf{t}$ and the timing layer $\mathbf{T}$. The activity in each $\mathbf{t}$ column is mapped on to the activity in the corresponding $\mathbf{T}$ column via the operator $\mathbf{L}_k^{-1}$ according to equation 6.

k	Delay			
	2	5	10	20
4	0.70	0.70	0.70	0.70
12	0.31	0.31	0.31	0.31
20	0.23	0.23	0.23	0.23
40	0.16	0.16	0.16	0.16

Figure 5.

Scalar property of the reconstructed stimulus history. Four stimuli of duration 0.1 sec was presented at various moments in the recent past. The curves show the reconstructed stimulus history for these stimuli, with a fixed value of $k = 12$. The coefficient of variation, the mean divided by the standard deviation, of each of these curves is exactly the same. The qualitative features of the graph is the same for all k , but the coefficient of variation decreases for higher values of k . The table gives the coefficient of variation of these curves for each stimulus (columns) for different values of k (rows).

Figure 6.

Reconstruction of complex stimulus history. A stimulus is presented twice in the recent past and the reconstructed history is plotted for $k = 12$ (left) and $k = 4$ (right). For the $k = 12$ case, the two peaks clearly occur at the appropriate positions, with the more recent stimulus being better represented. For the $k = 4$ case, the earlier peak is barely discernible.

Figure 7.

Time dependent activity of various layers of the model. A stimulus is presented twice, and the activity of two cells in the corresponding $\mathbf{t}$ and $\mathbf{T}$ columns are shown as a function of time. Note that the activity of the $\mathbf{T}$ cells peak roughly at the appropriate delay after each stimulus presentation.

Figure 8.

Neural Representation of the internal time. The left most panel shows a column of the context layer $\mathbf{t}$, and the associated column in the timing layer $\mathbf{T}$. The cells in these two columns are mapped in a one to one fashion. The activity of any cell in the $\mathbf{T}$ column not only depends on the activity of its counterpart in the $\mathbf{t}$ column, but also on the activity of k neighbors in the $\mathbf{t}$ column. This is a discretized approximation of $\mathbf{L}_k^{-1}$ from eq. 6. The right panel gives a pictorial representation of the connectivity between a cell in the $\mathbf{T}$ column and its k near neighbors in the $\mathbf{t}$ column. The contribution from the neighbors alternate between excitation and inhibition in either directions. The points above x-axis are excitations and the points below the x-axis are inhibitions. The tick marks on x-axis denote the position of the neighboring cells on either side. The dotted curve that forms an envelope simply helps to illustrate that the magnitude of the contribution falls off with the distance to the neighbor. With $k = 2$, we see an off-center-on-surround connectivity. With $k = 4$, we see a mexican-hat like connectivity, and $k = 12$ shows a more elaborate band of connectivity.

Figure 9.

Timing mechanism and memory. The external input at any moment activates a unique node in the stimulus layer **f**. Corresponding to each stimulus node is a column of cells in the context layer **t** that get activated via **C** according to Eq. 4. Each cell in this column has a distinct decay rate spanning $0 < \rho < 1$. Each column of the **t** layer is mapped on to the corresponding column in the **T** layer via L_k^{-1} described in Figure 8. The **T** layer activity at each moment is associated in a Hebbian fashion with the **f** layer activity and these associations are stored in **M**. After sufficient training with sequential external inputs, the associations in **M** can grow significantly strong and the **T** layer activity at any moment can induce activity in the **f** layer through **M**. This internally generated activity in the stimulus layer is interpreted as the prediction **p** for the next moment.

Figure 10.

Timing in goldfish. During training, the US (shock) was presented 5 sec (top panel) and 15 sec (bottom panel) after the onset of the CS (light). The rate of CR is plotted in the left panel as a function of the time after presentation of the CS in the absence of the US. This figure is reproduced from Drew et al (2005). The different curves represent different number of learning trials. Notice that the response gets stronger with learning trials. The right panel shows the probability of CR generated from simulations of the model. In these simulations, for simplicity, only the onset of CS is encoded into the context, not the entire CS. The parameters used in this simulation are $k = 4$, $\theta = 0.1$ and $\varphi = 1$.

Table 1

The prediction vector $\mathbf{p}$ exhibits the scalar property at the behavioral level. Coefficient of variation of $p_{stop}(d)$ (see text for details). The column headings give the value for d_o . The different rows give values for the corresponding value of k .

k	2	5	10	20
4	0.86	0.86	0.86	0.86
12	0.43	0.43	0.43	0.43
20	0.32	0.32	0.32	0.32
40	0.22	0.22	0.22	0.22