# Learning temporal relationships between symbols with Laplace Neural Manifolds

Marc W. Howard[1*], Zahra Gh. Esfahani[1], Bao Le[2],
Per B. Sederberg[2]

[1*]Department of Psychological and Brain Sciences, Boston University, 610 Commonwealth Ave, Boston, 02215, MA, USA.
[2]Department of Psychology, University of Virginia, 409 McCormick Road, Charlottesville, 22904, VA, USA.

*Corresponding author(s). E-mail(s): marc777@bu.edu;
Contributing authors: ghasemi.zahra@gmail.com; uac6qw@virginia.edu;
pbs5u@virginia.edu;

## Abstract

Firing across populations of neurons in many regions of the mammalian brain maintains a temporal memory, a neural timeline of the recent past. Behavioral results demonstrate that people can both remember the past and anticipate the future over an analogous internal timeline. This paper presents a mathematical framework for building this timeline of the future. We assume that the input to the system is a time series of symbols—sparse tokenized representations of the present—in continuous time. The goal is to record pairwise temporal relationships between symbols over a wide range of time scales. We assume that the brain has access to a temporal memory in the form of the real Laplace transform. Hebbian associations with a diversity of synaptic time scales are formed between the past timeline and the present symbol. The associative memory stores the convolution between the past and the present. Knowing the temporal relationship between the past and the present allows one to infer relationships between the present and the future. With appropriate normalization, this Hebbian associative matrix can store a Laplace successor representation and a Laplace predecessor representation from which measures of temporal contingency can be evaluated. The diversity of synaptic time constants allows for learning of non-stationary statistics as well as joint statistics between triplets of symbols. This framework synthesizes a number of recent neuroscientific findings including results from dopamine neurons in the mesolimbic forebrain.

**Keywords:** Temporal memory, prediction, Laplace transform, convolution

1

# 1 Introduction

Consider the experience of listening to a familiar melody. As the song unfolds, notes feel as if they recede away from the present, an almost spatial experience. According to Husserl (1966) "points of temporal duration recede, as points of a stationary object in space recede when I 'go away from the object'." For a familiar melody, Husserl (1966) argues that events predicted in the future also have an analogous spatial extent, a phenomenon he referred to as *protention*. This experience is consistent with the hypothesis that the brain maintains an inner timeline extending from the distant past towards the present and from the present forwards into the future. In addition to introspection and phenomenological analysis, one can reach similar conclusions from examination of data in carefully controlled cognitive psychology experiments (Tiganj, Singh, Esfahani, & Howard, 2022).

The evolutionary utility of an extended timeline for future events is obvious. Knowing what will happen when in the future allows for selection of an appropriate action in the present. Indeed, much of computational neuroscience presumes that the fundamental goal of the cortex is to predict the future (Clark, 2013; Friston, 2010; Friston & Kiebel, 2009; Palmer, Marre, Berry, & Bialek, 2015; Rao & Ballard, 1999).

In AI, a great deal of research focuses on reinforcement learning (RL) algorithms that attempt to optimize future outcomes within a particular planning horizon (Dabney et al., 2020; Ke et al., 2018) without a temporal memory. From the perspective of psychology, RL is a natural extension of the Rescorla-Wagner model (Rescorla & Wagner, 1972) an associative model for classical conditioning (Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1981; Waelti, Dickinson, & Schultz, 2001). Associative models describe connections between a pair of stimuli (or stimulus and an outcome etc) as a simple scalar value. In simple associative models, variables that affect the strength of an association, such as the number of pairings between stimuli, or attention, etc, must all combine to affect a single scalar value. Thus, although the strength of an association can fall off with the time between stimuli, the association itself does not actually convey information about time *per se* (C. Gallistel, 2021a).

Cognitive psychologists have argued that classical conditioning does not reflect atomic associations between stimuli, but rather explicit storage and retrieval of temporal relationships (Arcediano, Escobar, & Miller, 2005; Balsam & Gallistel, 2009; Cohen & Eichenbaum, 1993; C. Gallistel, Craig, & Shahan, 2019; Namboodiri, 2021). In this view, behavioral associations in classical conditioning reflect learning of temporal *contingencies* between stimuli, such that knowing that a particular stimulus was experienced in the present changes our expectations for the time at which an outcome will be experienced (Floeder, Jeong, Mohebi, & Namboodiri, 2024; Jeong et al., 2022). Such a theory clearly requires a temporal memory in order to learn temporal relationships between stimuli.

This paper presents a formal hypothesis for how populations of neurons could learn and express temporal relationships between symbols, ignoring similarity structure within stimuli. We assume the existence of a temporal memory expressed in the firing of neurons with an effectively continuous spectrum of time constants, forming the Laplace transform of the recent past (Atanas et al., 2023; Bright et al., 2020;

Kanter, Lykken, Moser, & Moser, 2024; Tsao et al., 2018; Zuo et al., 2023). Neurophysiological results suggest that the temporal memory expressed in neural firing extends at least several minutes (Tsao et al., 2018). We additionally hypothesize a neural timeline of the future expressed as Laplace transform (Cao, Bright, & Howard, in press). The present is part of both the past and the future, so that the current symbol is simultaneously the most recent part of the past and the most imminent part of the future. Hebbian associations between the Laplace transform for the past and the present symbol store temporal relationships between symbols. In addition, a continuous spectrum of *synaptic* time scales enable learning of temporal relationships over time scales much longer than a few minutes. This spectrum of synaptic time constants also enables learning of higher-order relationships among symbols expressed as their joint statistics.

# 2 Constructing neural timelines of the past and future

We take as input a finite set of discrete symbols, X, Y, etc., that are occasionally presented for an instant in continuous time. There are consistent temporal relationships between some of the symbols, such that knowing one symbol was presented at time $t$ may provide information about the occurrence of another symbol at time $t + \tau$. For convenience we assume that the time between repetitions of any given symbol is much longer than the temporal relationships that are to be discovered and much longer than the longest time constant $1/s_{\min}$. Much like the assumptions necessary to write out the Rescorla-Wagner model (C. Gallistel, 2021b; Namboodiri, 2021), this set of assumptions allows us to imagine that experience is segmented into a series of discrete trials and that each symbol can be presented at most once per trial. This assumption allows easy interpretation of quantities that we will derive.

## 2.1 The present

Let us take as input a vector valued function of time $\mathbf{f}(t)$. The notation $\mathbf{v}$ refers to a vector with each element a real number, $\mathbf{v}'$ is a transposed vector, so that $\mathbf{u}'\mathbf{v}$ is the inner product, a scalar, and $\mathbf{u}\mathbf{v}'$ is the outer product, a matrix. We assume a tokenized representation between symbols, so that $\mathbf{y}'\mathbf{x} = \delta_{y,x}$ where $\delta_{ij}$ is the Kronecker delta function. We write $\mathbf{f}_t$ for the symbol available at time $t$. At instants $t$ when no stimulus is presented, $\mathbf{f}_t = \mathbf{0}$, the zero vector. If we present a specific symbol at a specific time $t_o$, this adds to $\mathbf{f}(t)$ the basis vector for that symbol multiplied by a delta function over time centered at $t_o$. At most times, the input is zero. We will occasionally refer to the moment on a particular trial when X is presented, $\mathbf{f}_t = \mathbf{x}$ as $t_x$.

We write $\mathbf{f}_t(\tau)$ to describe the true past that led up to time $t$. The continuous variable $\tau$ runs from $0^-$, corresponding to the moment of the past closest to the present backwards to $-\infty$, corresponding to the distant past. Whereas $\mathbf{f}_t$ is the symbol available in the present at a particular instant $t$, $\mathbf{f}_t(\tau), \tau \in (-\infty, 0)$ is the timeline that led up to time $t$. Under the assumption that every symbol is presented at most once per trial, each component of $\mathbf{f}_t(\tau)$ over the interval $\tau < 0$ is either a delta function at some particular $\tau$ or zero everywhere. The goal of the associative memory is to provide

3

a guess about the future that will follow time $t$, $\mathbf{f}_t(\tau), \tau \in (0, \infty)$ (Figure 1) given the symbol available in the present.

The symbol provided in the present $\mathbf{f}_t$ is available to both the past and the future. The present enters the past timeline at its most recent point. In this formulation, the present is also available as the most rearward portion of the future timeline. By associating the past to the rearward portion of the future, we can learn temporal relationships between symbols separated in time. By probing these associations with the present—as the most recent part of the past timeline—we can construct an extended estimate of the future.

## 2.2 Laplace neural manifolds for the past and the future

We estimate both the past and the future as functions over neural manifolds. Each manifold is a population of processing elements—neurons—each of which is indexed by a position in a coordinate space. We treat the coordinates as effectively continuous and locally Euclidean. At each moment, each neuron is mapped onto a scalar value corresponding to its firing rate over a macroscopic period of time on the order of say 100 ms. We propose that the past and the future are represented by separate manifolds that interact with one another.

The representations for both the past and the future each utilize two connected manifolds. We refer to one kind of manifold, indexed by an effectively continuous variable $s$, as a Laplace space. The other kind of manifold, indexed by an effectively continuous variable $\overset{*}{\tau}$, is referred to as an inverse space. Taken together, we refer to these paired representations as a Laplace Neural Manifold. The representations of the past follow previous work in theoretical neuroscience (Howard et al., 2014; Shankar & Howard, 2013), cognitive psychology (Howard, Shankar, Aue, & Criss, 2015; Salet, Kruijne, van Rijn, Los, & Meeter, 2022), and neuroscience (Bright et al., 2020; Cao, Bladon, Charczynski, Hasselmo, & Howard, 2022).

### 2.2.1 Laplace spaces for remembered past and predicted future

The Laplace space corresponding to the past, which we write as $\mathbf{F}_t^-(s)$ encodes the Laplace transform of $\mathbf{f}_t(-\tau)$, the past leading from the present at time $t$ back towards the infinite past:

$$\mathbf{F}_t^-(s) = \int_0^\infty e^{-s\tau}\mathbf{f}_t(-\tau)d\tau = \mathcal{L}\left\{\mathbf{f}_t(-\tau)\right\}(s), \ \tau \leq 0 \tag{1}$$

We restrict $s$ to real values on the positive line (but see Aghajan, Kreiman, & Fried, 2023).[1] Many neurons tile the $s$ axis continuously for each symbol. To the extent that we can ensure a set of exponential receptive fields with a continuous spectrum of $s$ values, we have established that $\mathbf{F}_t^-(s)$ is the Laplace transform of the past. Exponential receptive fields over the past with a continuous spectrum of time constants

---

[1]The sign convention here is distinct from prior papers that did not require $\tau$ (and $\overset{*}{\tau}$ which will be introduced shortly) to be defined on both sides of zero (Fig. 1).
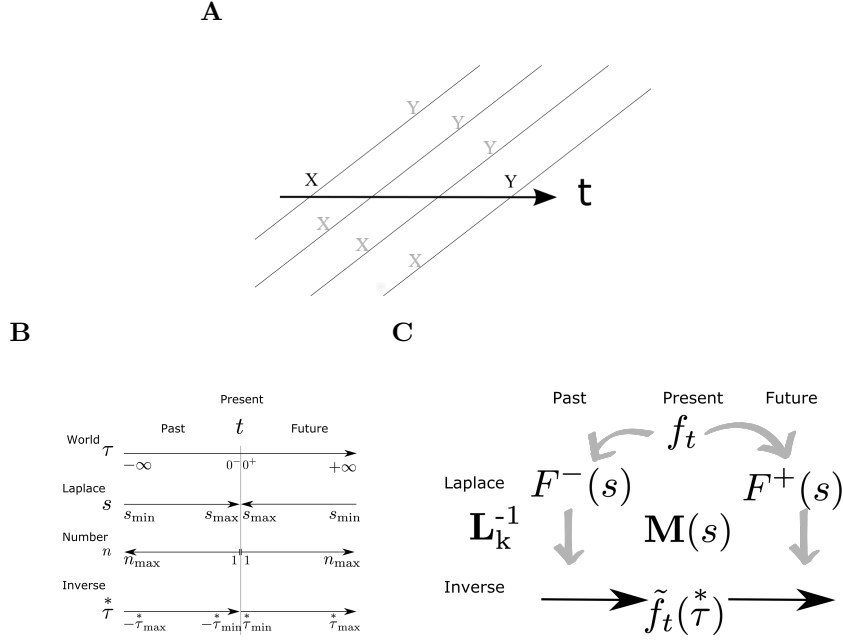
**Fig. 1** Guide to notation. **A.** Time measured externally is drawn as a horizontal line; the "internal timeline" available to the agent at each moment is drawn as a diagonal line. The remembered past at time $t$ is drawn below the horizontal; the predicted future is drawn above the horizontal. Locations on the internal timeline are spaced to suggest logarithmic compression. Consider a case in which X and Y are presented many times with a consistent temporal relationship. If X is presented at $t_x$ and then Y is presented at some later time $t_y$. After X is presented at $t = t_x$, it recedes into the past, so for $t > t_x$ we find that $\mathbf{f}_t(t - t_x) = x$. At the moment Y is presented, $f_{t_y}(t_y - t_x) = x$. After the relationship between X and Y is learned, then after X is presented Y is predicted a time $t_y - t_x$ in the future. As time proceeds after presentation of X, the predicted occurrence of Y should approach closer and closer to the present. **B.** Sign conventions. At the present moment $t$, objective time $\tau$ runs from $-\infty$ to $\infty$. $\tau = 0$ corresponds to time $t$. The real Laplace domain variable $s$ runs from $0^+$ to $+\infty$ for both past and future, approximated as $s_{\min}$ and $s_{\max}$. The units of $s$ are $t^{-1}$; the values corresponding to different points of the timeline are shown in the same vertical alignment. Cell number for Laplace and inverse spaces $n$ are aligned with one another. The variable $\overset{*}{\tau}$ describes position along the inverse spaces. It is in register with $\tau$ and derived from $s$. **C.** The stimulus available in the present, $\mathbf{f}_t$ provides input to two sets of neural manifolds. One set of neural manifolds represents the past; the other estimates the future. $\mathbf{M}(s)$ stores temporal relationships between events.

have been observed in many brain regions and species (Atanas et al., 2023; Bright et al., 2020; Cao et al., in press; Danskin et al., 2023; Tsao et al., 2018; Zuo et al., 2023).

The index $s$ assigned to a neuron corresponds to the inverse of its functional time constant. The Laplace space corresponding to the future, which we write as $\mathbf{F}^+(s)$ is an attempt to estimate the Laplace transform of the future, $\mathcal{L}\{\mathbf{f}_t(\tau)\}(s)$ over the interval $\tau \geq 0$. Thus, there is a natural mapping between $1/s$ and $|\tau|$ within both the past and the future. By convention, $s$ is positive for both the past and the future so that $\mathbf{F}_t^-(s)$ is the Laplace transform of $\mathbf{f}_t(-\tau)$ for $\tau \leq 0$ whereas $\mathbf{F}_t^+(s)$ is the Laplace transform of $\mathbf{f}_t(\tau)$ for $\tau \geq 0$.

5

Although $s$ is effectively continuous, this does not require that neurons sample $s$ evenly. Following previous work in psychology (e.g., Chater & Brown, 2008; Howard & Shankar, 2018; Piantadosi, 2016), neuroscience (Cao et al., 2022; Guo, Huson, Macosko, & Regehr, 2021), and theoretical neuroscience (Lindeberg & Fagerström, 1996; Shankar & Howard, 2013), we assume that $s$ is sampled on a logarithmic scale. Let $n$ be the neuron number, starting from the largest value of $s_{\max}$ nearest $\tau = 0$ and extending out from the present. We obtain a logarithmic scale by choosing $ds/dn = -s$.

### 2.2.2 Updating Laplace spaces in real time

Suppose that we have arranged for one particular component of $\mathbf{F}_t^-(s)$ or $\mathbf{F}_t^+(s)$ to hold the Laplace transform of one particular symbol, which we write as $f_t(\tau)$. Suppose further that $f_t(\tau)$ is zero in the neighborhood of $\tau = 0$. Consider how this component, which we write as $F^-(s)$ or $F^+(s)$, should update as time passes. Let us pick some minimal increment of time $\Delta t$ on the order of, say, 100 ms. At time $t + \Delta t$, information in $f_t(\tau)$ for $\tau \leq 0$ recedes further away from the present, so that $F_{t+\Delta t}^-(s) = \mathcal{L}\{f_t(\tau + \Delta t)\}$. In contrast, at time $t + \Delta t$, information in the future $f_t(\tau)$ for $\tau > 0$ comes closer to the present, so that $F_{t+\Delta t}^+(s) = \mathcal{L}\{f_t(\tau - \Delta t)\}$. More generally, suppose that $F_t(s)$ is the Laplace transform of a function over some variable $x$, $F_t(s) = \mathcal{L}\{f_t(x)\}(s)$. Defining $\alpha \equiv \frac{\Delta x}{\Delta t}$, we can update $F_t(s)$ as

$$F_{t+\Delta t}(s) = \mathcal{L}\{\mathcal{T}_{\alpha \Delta t} f_t(x)\}(s) = e^{-s\alpha \Delta t} F_t(s) \tag{2}$$

where $\mathcal{T}$ is the translation operator, $\mathcal{T}_a f(x) = f(x+a)$ and we have used the expression for the Laplace transform of translated functions. Equation 2 describes a recipe for updating both $F_t^{\pm}(s)$ with $\alpha_{\pm}$ in the absence of new input. Using the sign convention developed here, we fix $\alpha_- = 1$ for $F^-(s)$ and fix $\alpha_+ = -1$ for $F^+(s)$. It is possible to incorporate changes into the rate of flow of subjective time by letting $\alpha_{\pm}$ change in register, such that $\alpha_+(t) = -\alpha_-(t)$ for all $t$. The expression in Eq. 2 holds more generally and can be used to update Laplace transforms over many continuous variables of interest for cognitive neuroscience (Howard & Hasselmo, 2020; Howard, Luzardo, & Tiganj, 2018; Howard et al., 2014).

We are in a position to explain how $\mathbf{F}_t^-(s)$ comes to represent the Laplace transform of $\mathbf{f}_t(-\tau)$ over the interval $\tau \in (-\infty, 0)$; a discussion of how $\mathbf{F}_t^+(s)$ comes to estimate the future requires more development and will be postponed. When a symbol is presented at time $t$, it enters timeline of the past at $\tau = 0$. So, incorporating the Laplace transform of the most recent part of the past with the past that is already available and then evolving to time $t + \Delta t$ we have

$$\begin{aligned}
\mathbf{F}_{t+\Delta t}^-(s) &= e^{-s\Delta t}\left[\mathbf{F}_t^-(s) + \mathcal{L}\{\delta(0)\,\mathbf{f}_t\}\right] \\
&= e^{-s\Delta t}\mathbf{F}_t^-(s) + e^{-s\Delta t}\mathbf{f}_t.
\end{aligned} \tag{3}$$

At time $t + \Delta t$, the input from time $t$ is encoded as the Laplace transform of that symbol a time $\Delta t$ in the past. At each subsequent time step, an additional factor of $e^{-s\Delta t}$ accumulates. As time passes, the input from time $t$ is always stored as Laplace transform of a delta function at the appropriate place on the timeline. Because this

is true for all stimuli that enter $\mathbf{F}^-(s)$, we conclude that $\mathbf{F}_t^-(s)$ encodes the Laplace transform of the past $\mathbf{f}_t(-\tau)$ over the interval $\tau \leq 0$.

The middle panel of Figure 2 illustrates the profile of activity over $\mathbf{F}_t^-$ and $\mathbf{F}_t^+$, shown as a function of cell number $n$, resulting from the Laplace transform of a delta function at various moments in time. In the middle panel, the $s$ axis for the past is reversed to allow appreciation of the relationship between past time $\tau \leq 0$ and $F^-$. Note that the Laplace transform of a delta function has a characteristic shape as a function of cell number that merely translates as time passes. Note that the magnitude of the translation of $F^{\pm}[n]$ depends on the value of $\tau_o$. It can be shown that for a delta function $F_{t+\Delta t}^{\pm}[n] = F_t^{\pm}[n + \Delta n]$ with $\Delta n = \alpha_{\pm}\frac{\Delta t}{\tau_o}$. This can be appreciated by noting that the distances between successive lines in the middle panel of Fig. 2 are not constant despite the fact that they correspond to the same time displacement. Whereas $\Delta n$ goes down as time passes for $F^-[n]$ as the past becomes more remote from the present, $\Delta n$ increases with the passage of time for $F^+[n]$ as the future grows closer to the present.

There are implementational challenges to building a neural circuit that obeys Eq. 2; these challenges are especially serious when $\alpha < 0$, which requires activation to grow exponentially. If one is willing to restrict the representation of each symbol to the Laplace transform of a delta function at a single point in time, it is straightforward to implement a continuous attractor network (Khona & Fiete, 2021) to allow the "edge" in the Laplace transform as a function of $n$ to translate appropriately. Daniels and Howard (submitted) constructed a simple continuous attractor network to demonstrate the feasibility of this approach.

### 2.2.3 Inverse spaces for remembered past and predicted future

The mammalian brain also contains "time cells" with circumscribed receptive fields (MacDonald, Lepage, Eden, & Eichenbaum, 2011; Pastalkova, Itskov, Amarasingham, & Buzsaki, 2008; Schonhaut, Aghajan, Kahana, & Fried, 2023; Tiganj, Cromer, Roy, Miller, & Howard, 2018). Time cells resemble a "direct" estimate of the past and are reasonably well approximated as:

$$\tilde{f}_t(\overset{*}{\tau}) = \int_0^{\infty} \Phi\left(\frac{\tau}{\overset{*}{\tau}}\right) f_t(-\tau)d\tau \tag{4}$$

where $\Phi(x)$ is a unimodal function with its maximum at 1 and $\overset{*}{\tau}$ is here defined to be negative (Fig. 1). $\tilde{f}_t(\overset{*}{\tau})$ estimates the true past in the neighborhood of $\overset{*}{\tau}$. As $\Phi()$ becomes more and more sharp, approaching a delta function, $\tilde{f}_t(\overset{*}{\tau})$ goes to $f_t(\overset{*}{\tau})$. In this sense $\tilde{f}(\overset{*}{\tau})$ is like the inverse Laplace transform of $F_t(s)$. However, because receptive fields depend only on the ratio of $\tau/\overset{*}{\tau}$, and because neurons sample the $\overset{*}{\tau}$ axis logarithmically, $\tilde{f}(n)$ is a convolution of $f(\log \tau)$ and another function of $\log \tau$ that controls the blur.

The bottom panel of Figure 2 shows a graphical depiction of the inverse space for the past and the future during the interval between presentation of X and Y. The inverse spaces approximate the past, $\tilde{f}_t(\overset{*}{\tau})$ for $\overset{*}{\tau} < 0$ and the future, $\tilde{f}_t(\overset{*}{\tau})$ for $\overset{*}{\tau} > 0$
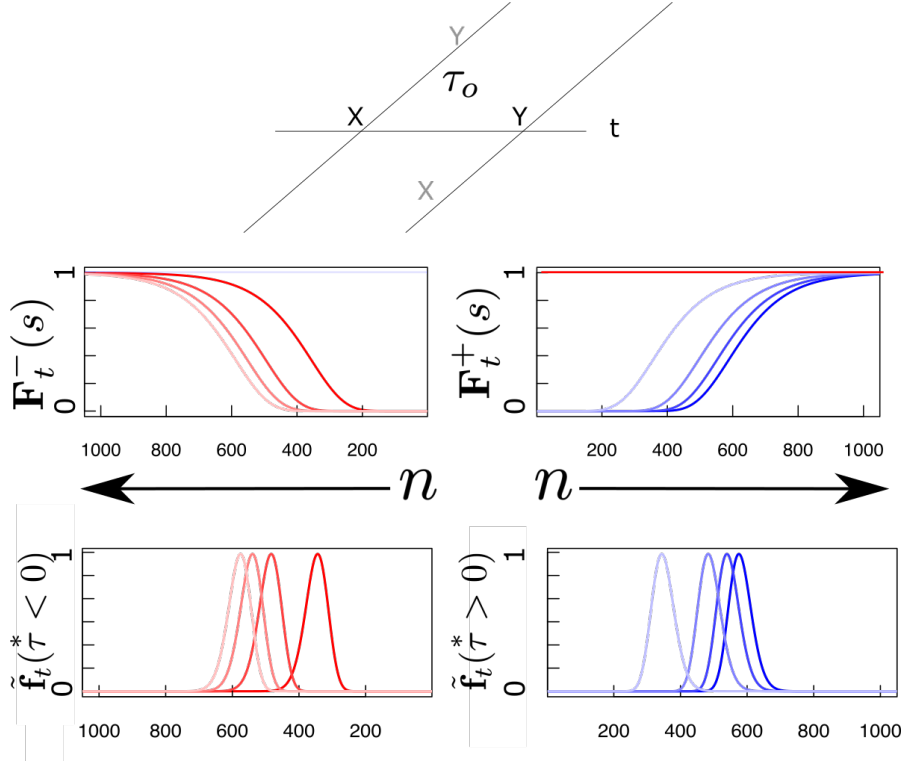
**Fig. 2** Neural manifolds that construct a logarithmically-compressed internal timeline of the past and the future. Top: A temporal relationship exists between X and Y such that Y always follows X after a delay of $\tau_o$ seconds. Consider how the internal timeline ought to behave after X is presented at $t = 0$. At time $t$, the past should include X $t$ seconds in the past and Y $\tau_o - t$ seconds in the future. Middle and bottom: Samples of the timeline at evenly-spaced moments between zero and $\tau_o$. At each moment, there is a pattern of activity over neurons indexed by $n$. The state of the timeline at earlier moments, closer to $t = 0$, are darker and later moments closer to $t = \tau_o$ are lighter. Red lines are neurons coding for X (primarily in the past except precisely at $t$), blue lines are neurons coding for Y (primarily in the future except precisely at $t = \tau_o$). Middle: Laplace spaces for the past (left) and future (right) shown as a function of cell number $n$; Bottom: inverse spaces, constructed using the Post approximation, for the past (left) and future (right) shown as a function of log time. Exactly at time $t = 0$, X is available a time $0^+$ in the future (dark horizontal red line, middle right). Similarly, exactly at $t = \tau_o$, Y is available a time $0^-$ in the past (light horizontal blue line, middle left).

on a log scale. As the delta function corresponding to the time of X recedes into the past, the corresponding bump of activity in $\mathbf{x}'\tilde{\mathbf{f}}_t(n)$ also moves, keeping its shape but moving more and more slowly as X recedes further and further into the past. In the future, the delta function corresponding to the predicted time of Y should start a time $\tau_o$ in the future and come closer to the present as time passes. As the prediction for Y approaches the present, the corresponding bump of activity in $\mathbf{y}'\tilde{\mathbf{f}}_t(n)$ keeps its shape but the speed of the bump accelerates rather than slowing with the passage of time.

It is in principle possible to construct the inverse space from the Laplace space *via* a linear feedforward operator. Previous papers (e.g., Shankar & Howard, 2013) have made use of the Post approximation to the inverse Laplace transform to construct the

inverse space from the Laplace space. This is not neurally reasonable (Gosmann, 2018); the Post approximation is difficult to implement even in artificial neural networks (e.g., Jacques, Tiganj, Howard, & Sederberg, 2021; Tano, Dayan, & Pouget, 2020). A more robust approach would be a continuous attractor network (for a review see Khona & Fiete, 2021) that takes input as the derivative of $F$ with respect to $n$. The width of the bump in $\tilde{f}$ would depend on internal connections between neurons in $\tilde{f}$ and global inhibition would stabilize the activity over $\tilde{f}$. In this case, moving the bump in different directions, corresponding to $\alpha > 0$ and $\alpha < 0$ is analogous to moving a bump of activity in a ring attractor in different directions. A companion paper fleshes out these ideas (Daniels & Howard, submitted).

## 2.3 Predicting the future from the past

The previous subsection describes how to evolve the Laplace manifold for the past. If we could somehow initialize the representation of the future appropriately then we could use the same approach to evolve the Laplace manifold for the future during periods when no symbol is experienced. Initializing the future will be accomplished *via* learned temporal relationships between the past and the future.

The model has access to the Laplace transform of the past, as described above. We define the present so that it overlaps with both the most recent part of the past and the most imminent, or "rearward," part of the future. We form Hebbian associations between the Laplace transform of the past and the Laplace transform of the rearward portion of the future. Recall that products of Laplace transforms are the Laplace transform of the convolution of these functions. Because there is a reflection between the definition of $\mathbf{F}_t^-(s)$ and $\mathbf{F}_t^+(s)$, the convolution of these two functions measures distances between time points in the past and the present. Later the present stimulus, taken as the Laplace transform of the most recent part of the past, can be used to recover the Laplace transform of an extended future timeline.

There are two sets of weights storing these associations, $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$. Each of these weights learn associations between the Laplace transform of the past, $\mathbf{F}_t^-(s)$, and the present stimulus $\mathbf{f}_t$. The two sets of weights are normalized differently. Roughly speaking, $\mathbf{M}(s)$ stores the Laplace transform of the future conditionalized on the present symbol. In contrast $\bar{\mathbf{M}}(s)$ stores the Laplace transform of the past conditionalized on the present symbol. With the assumptions that let us consider discrete trials, these transforms are understandable as pairwise statistics of events corresponding to a presentation of each symbol on a trial. We will see that taken together $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$ enable us to estimate the associative and temporal contingency between each pair of symbols conditionalized on each other symbol.

The learning rate and forgetting rate for the sets of weights fixes a time horizon for learning over trials. By choosing a continuous spectrum of forgetting rates $\rho$ and learning rates $1 - \rho$, both $\mathbf{M}(\rho, s)$ and $\bar{\mathbf{M}}(\rho, s)$ retain a memory for the history as a function of trials. Continuous forgetting allows the weights to implement a discrete approximation to the Laplace transform. This property of $\mathbf{M}(\rho, s)$ and $\bar{\mathbf{M}}(\rho, s)$ means that it is in principle possible to aggregate joint statistics between stimuli.

9
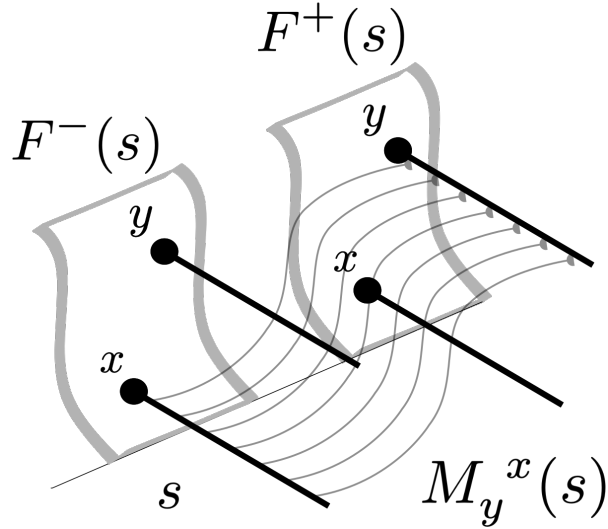
**Fig. 3** Schematic figure illustrating $M_y{}^x(s)$. $F^-(s)$ and $F^+(s)$ components for all the possible symbols, here shown schematically as sheets. Two symbols X and Y are shown in both $F^-(s)$ and $F^+(s)$. Each symbol is associated with a population of neurons spanning a continuous set of $s$ values, shown as the heavy lines in this cartoon. $\mathbf{M}(s)$ describes the connections between each symbol in $\mathbf{F}^-(s)$ to each symbol in $\mathbf{F}^+(s)$ for each value of $s$. The curved lines $M_y{}^x(s)$ illustrate the set of weights connecting units corresponding to X in $F^-$ to units corresponding Y in $F^+$. Connections exist only between units with the same values of $s$. The strength of the connections in $M_y{}^x(s)$ vary as a function of $s$ in a way that reflects the pairwise history between X and Y.

### 2.3.1 Encoding $\mathbf{M}(s)$

The moment a nonzero stimulus $\mathbf{f}_t$ is experienced, we assume it is available to both $\mathbf{F}^-$ and $\mathbf{F}^+$, triggering a number of operations which presumably occur sequentially within a small window of time on the order of 100 ms. First, the present stimulus updates a prediction for the future *via* a set of connections $\mathbf{M}$ organized by $s$. Then these connections are updated by associating the past to the present. Finally the present stimulus is added to the representation of the past. For ease of exposition we will first focus on describing the connections between the past and the future.

We write $\mathbf{M}(s)$ for a set of connections that associates the Laplace transform of the past to the Laplace transform of the future (Fig. 3). We postpone discussion of the other set of weights $\bar{\mathbf{M}}(s)$. For any particular value $s_o$, $\mathbf{M}(s_o)$ is a matrix describing connections from each symbol in $\mathbf{F}^-(s_o)$ to each symbol in $\mathbf{F}^+(s_o)$. For each pair of symbols, say X and Y, we write $M_y{}^x(s_o)$ for the strength of the connection *from* the cell corresponding to $\mathbf{x}$ with $s = s_o$ in $\mathbf{F}^-$ *to* the cell corresponding to $\mathbf{y}$ in $\mathbf{F}^+$ with $s = s_o$. $\mathbf{M}(s)$ does not include connections between neurons with different values of $s$. On occasion it will be useful to think of the set of connections between a pair of symbols over all values of $s$, which we write as $M_y{}^x(s)$. Similarly, we write $\mathbf{M}^y(s)$ for the set of connections *from* Y in $F^-$ to all stimuli in $F^+$ over all values of $s$. We write $\mathbf{M}_y(s)$ for the set of connections *to* Y in $F^+$ from all symbols and all values of $s$. In

10

this paper, the superscripting and subscripting of $M_y{}^x(s)$ has no significance beyond a visual aid to help keep the indices straight.

When a particular stimulus Y is presented the connections to and from that stimulus in $\mathbf{M}(s)$ are updated. When Y is presented, the connections from Y in the past towards all stimuli in the present are updated as

$$\mathbf{M}^y(s) \rightarrow \rho \mathbf{M}^y(s) \tag{5}$$

That is, the connections from $y$ to every other stimulus for each value of $s$ are all scaled down by a value $\rho$. Later we will consider the implications of a continuous spectrum of $\rho$ values; for now let us just treat $\rho$ as a fixed parameter restricted to be between zero and one. When Y is presented, it momentarily becomes available at the "rearward part" of the future. In much the same way that the present enters the past (Eq. 3) at $\tau = 0^-$, we also assume that the present is also available momentarily in the future at $\tau = 0$. When Y is presented, the connections from each symbol in the past to Y in the future are updated as

$$\mathbf{M}_y(s) \rightarrow \mathbf{M}_y(s) + (1 - \rho)\mathbf{F}_t^-(s) \tag{6}$$

Connections involving symbols that are not present in the history retained by $\mathbf{F}_t^-(s)$ are not updated. We can understand Eq. 6 as a Hebbian association between the units in $\mathbf{F}^-(s)$, whose current activation is given by $\mathbf{F}_t^-(s)$ and the units in the future $\mathbf{F}^+(s)$ corresponding to the present stimulus Y (see Fig. 4). More generally, we can understand this learning rule as strengthening connections from the past $\mathbf{F}_t^-(s)$ to the rearward part of the future, $\mathcal{L}\{\delta(0)\}(s)\mathbf{f}_t = e^{-s0}\mathbf{f}_t = \mathbf{f}_t$. Because the second term is the product of two Laplace transforms, it can also be understood as the Laplace transform of a convolution, here, the convolution of the present with the past.[2] Convolution has long been used as an associative operation in mathematical psychology (Jones & Mewhort, 2007; Kato & Caplan, 2017; Murdock, 1982), neural networks (Blouw, Solodkin, Thagard, & Eliasmith, 2016; Eliasmith, 2013; Plate, 1995), and computational neuroscience (Steinberg & Sompolinsky, 2022).

### 2.3.2 M(s) is a Laplace successor representation

From examination of Eqs. 5 and 6, we see that after each trial $M_y{}^x(s)$ is multiplied by $\rho$ when X was presented. For trials on which Y was also presented, $(1-\rho)e^{-s\tau_o}$ is added to $M_y{}^x(s)$. Writing $h[i]$ as an indicator variable for the history of presentations of Y on the trial $i$ steps in the past we find at the conclusion of a trial that

$$M_y{}^x(s) = (1-\rho)e^{-s\tau_o}\sum_i \rho^i h[i]. \tag{7}$$

---

[2]Because of the sign conventions adopted here, $F_t^-(s)$ is the Laplace transform of $f_t(-\tau)$ whereas $F_t^+(s)$ is the transform of $f_t(\tau)$. Viewed in this light it is more precise to think of Eq. 6 as learning the Laplace transform of the cross-correlation between the present and the past.
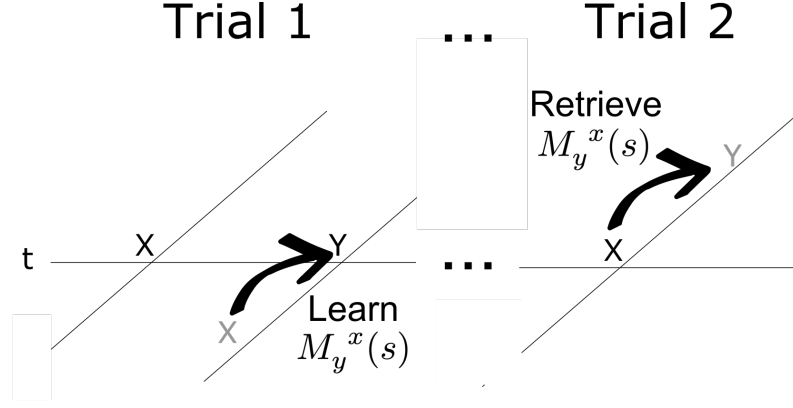
**Fig. 4** Learning and expressing pairwise associations with $\mathbf{M}(s)$. The horizontal line is time; the diagonal lines indicate the internal timeline at the moments they intersect. Memory for the past is below the horizontal line; prediction of the future is above. When X is presented for the first time, it predicts nothing. When Y is presented, the past contains a memory for X in the past. When Y is presented, $M_y{}^x(s)$ stores the temporal relationship between X in the past and Y in the present—the rearward part of the future. In addition to storing learned relationships, connections from each item decay each time it was presented (not shown). When X is repeated much later in time, the stored connections in $M_y{}^x(s)$ retrieve a prediction of Y in the future.

Note that if $P(y|x) = 1$, then after an infinitely long series of trials $\sum_i h[i]\rho^i = \frac{1}{1-\rho}$ and $M_y{}^x(s) = e^{-s\tau_o}$ for all choices of $\rho$. Following similar logic, if we relax the assumption that $P(y|x) = 1$ and take the limit as $\rho$ goes to 1, we find that $M_y{}^x(s) = P(y|x)e^{-s\tau_o}$.

Now let us relax the assumption that the time lag between X and Y always takes the same value. Let the lag be a random variable $\tau_{xy}$ subject to the constraint that $\tau_{xy}$ is always $> 0$. This is not a fundamental restriction; if $\tau_{xy}$ changed sign, those observations would contribute to $M_x{}^y(s)$ instead of $M_y{}^x(s)$. Now, again taking the limit as $\rho \to 1$, we find

$$M_y{}^x(s) = P\left(y|x\right) E\left[e^{-s\tau_{xy}}\right] = P\left(y|x\right) \mathcal{L}\left\{\tau_{xy}\right\}(s) \tag{8}$$

where we have used the definition for the Laplace transform of a random variable, again with the understanding that we restrict $s$ to be real and positive.

Equation 8 illustrates several important properties of $\mathbf{M}(s)$. First, we can see that $M_y{}^x(s)$ provides complete information about the distribution of temporal lags between X preceding Y. This can be further appreciated by noting that the Laplace transform of the random variable on the right hand side is the moment generating function of $-\tau_{xy}$. Keeping the computation in the Laplace domain means that there is no blur introduced by going into the inverse space as in previous attempts to build a model for predicting the future (Goh, Ursekar, & Howard, 2022; Shankar, Singh, & Howard, 2016; Tiganj, Gershman, Sederberg, & Howard, 2019). Second, because $\mathcal{L}\left\{\tau_{xy}\right\}(s = 0) = 1$ as long as the expectation of $\tau_{xy}$ is finite, $M_y{}^x(s = 0) = P(y|x)$ and $\mathbf{M}(s = 0)$ captures the pairwise probabilities between all symbols.

In the limit as $\rho \to 1$, $\mathbf{M}(s)$ is closely related to the successor representation (Carvalho, Tomov, de Cothi, Barry, & Gershman, 2024; Dayan, 1993; Gershman, Moore, Todd, Norman, & Sederberg, 2012; Momennejad et al., 2017; Stachenfeld, Botvinick, & Gershman, 2017) with a continuous distribution of discount rates (Kurth-Nelson & Redish, 2009; Masset et al., 2023; Momennejad & Howard, 2018; Sousa et al., 2023; Tano et al., 2020). More precisely, if one assumes a complete compound serial representation of the past and a fixed action policy, then computes the successor representation from RL (Dayan, 1993; Gershman et al., 2012), but with a continuous spectrum of discount rates $\gamma$, one would obtain $\mathbf{M}(s)$ with the identification $s = -\log \gamma$. However, computing $\mathbf{M}(s)$ does not require temporal difference learning. In RL language, $\mathbf{F}^-(s)$ is an ensemble of eligibility traces with a continuous spectrum of forgetting rates. Associating this multiscale eligibility trace to outcomes is sufficient to compute $\mathbf{M}(s)$, which we might refer to as a Laplace successor representation.

### 2.3.3 $\bar{\mathbf{M}}(s)$ is a Laplace predecessor representation

It is straightforward to construct a Laplace predecessor representation (Namboodiri & Stuber, 2021) using $\mathbf{F}^-(s)$, the Laplace transform of the past, and Hebbian learning. We write out a new set of connections $\bar{\mathbf{M}}(s)$. Adapting Eqs. 5 and 6, when each item Y is presented

$$\bar{\mathbf{M}}^y(s) \to \rho \bar{\mathbf{M}}^y(s) + (1 - \rho)\mathbf{F}_t^-(s) \tag{9}$$

That is, when Y is presented at time $t$ and X is available in $\mathbf{F}_t^-(s)$, $\bar{M}_x{}^y(s)$ is incremented. Following similar steps as for $\mathbf{M}(s)$, in the limit as $\rho \to 1$, we get

$$\bar{M}_x{}^y(s) = P(x|y)\, \mathcal{L}\{\tau_{xy}\}(s), \tag{10}$$

which can be compared to Equation 8. Thus, with learning as in Eq. 9, we can refer to $\bar{\mathbf{M}}(s)$ as a Laplace predecessor representation.

Note that the convention of $\bar{\mathbf{M}}(s)$ is different than $\mathbf{M}(s)$. Whereas $M_y{}^x(s)$ describes relationships between X preceding Y, $\bar{M}_y{}^x(s)$ describes relationships between Y preceding X. In this sense $\bar{\mathbf{M}}(s)$ is like $\mathbf{M}^T(s)$. In addition one must also account for the reflection operator involved in the definition of $\mathbf{F}_t^-(s)$ as compared to $\mathbf{F}_t^+(s)$ and the different marginalization.

The foregoing makes clear that if the brain has access to $\mathbf{F}^-(s)$—an eligibility trace with a continuum of time horizons—it is straightforward to compute either a successor representation or a predecessor representation in a way that maintains complete information about the temporal relationships between stimuli. This approach does not require selecting a single time horizon or time constant for either representation (Floeder et al., 2024).

### 2.3.4 Measures of contingency using $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$

Information contained in $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$ can be used to not only describe pairwise relationships between stimuli but also to assess contingency between symbols, allowing

13
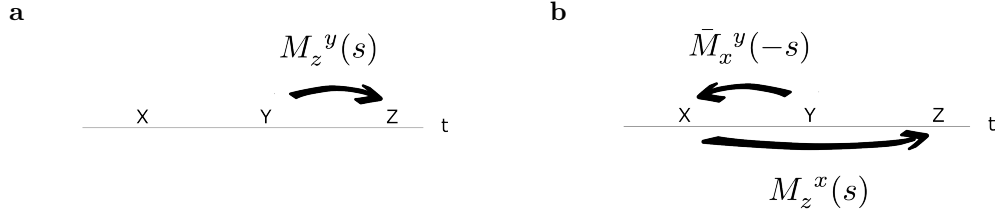
**Fig. 5** Measuring contingency by comparing pairwise relationships between Y and Z to pairwise relationships conditionalized on X. **a.** Equation 12 captures the Laplace transform of the random variable $\tau_{yz}$. By assumption, on each trial $\tau_{yz} = \tau_{xz} - \tau_{xy}$. **b.** Equation 14 captures the convolution of $\tau_{xz}$ and $-\tau_{xy}$. If these intervals are independent across trials then $M_z{}^y(s) = M_z{}^x(s)\bar{M}_x{}^y(-s)$.

solutions to the temporal credit assignment problem. The goal here is not to propose a specific measure of contingency—there are undoubtedly a multiplicity of such rules that could be used for cognitive and neural modeling—but simply to sketch out the properties of $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$. We continue attending to the limit as $\rho \to 1$.

For this illustration, let us restrict our attention to relationships between three symbols X, Y and Z. We assume for simplicity that, if they are presented on a trial, the three stimuli are presented in order on each trial. Let us refer to the time lags between symbols as random variables $\tau_{xy}$, $\tau_{yz}$; on trials where all three symbols are observed $\tau_{xz} = \tau_{xy} + \tau_{yz}$. For convenience let's assume that the distributions are chosen such that the relative times of presentation do not overlap. We denote the probabilities of each symbol occuring on a trial such that $P(z|y)$ gives the conditional probability that Z is observed on a trial given that Y is also observed on that trial.

We are interested in how much "credit" to allocate Y for the occurrence and timing of Z, taking into account X. We will compare $M_z{}^y(s)$, which describes the future occurrences of Z conditionalized on Y in the present to $M_z{}^x(s)\bar{M}_x{}^y(-s)$ (Fig. 5). This quantity is $M_z{}^x(s)$—the future of Z predicted by X—multiplied by $\bar{M}_x{}^y(-s)$—the past occurrence of X predicted by knowing that Y is in the present. That is, $M_z{}^x(s)\bar{M}_x{}^y(-s)$ describes the future of Z predicted by the past occurrence of X that is observed when Y is in the present. The reflection operator allows the integration of these two timelines in a way that can be compared to the future of Z given that Y was observed in the present (Cole, Barnet, & Miller, 1995).

We will work through the implications of this high level description under very simple circumstances. Recall that under the circumstances described in this subsection,

$$M_z{}^y(s) = P(z|y)\ \mathcal{L}\left\{\tau_{yz}\right\}(s) \tag{11}$$

$$= P(z|y)\ \mathcal{L}\left\{\tau_{xz} - \tau_{xy}\right\}(s) \tag{12}$$

Using properties of the Laplace transform we can rewrite $M_z{}^x(s)\bar{M}_y{}^x(-s)$ as

$$M_z{}^x(s)\bar{M}_y{}^x(-s) = P(z|x)\mathcal{L}\left\{\tau_{xz}\right\}(s)\ P(x|y)\mathcal{L}\left\{-\tau_{xy}\right\}(s) \tag{13}$$

$$= P(z|x)P(x|y)\ \mathcal{L}\left\{\tau_{xz} * (-\tau_{xy})\right\}(s). \tag{14}$$

14

The second term describes the Laplace transform of the convolution of $\tau_{xz}$ and $-\tau_{xy}$. Because the sum of two independent random variables is equal to their convolution, the Laplace transforms in Eqs 12 and 14 will enable us to assess the dependence between the times of presentations of X, Y, and Z.

***Associative contingency at $s = 0$***

$\mathbf{M}(s = 0)$ gives information about the pairwise probabilities between each pair of symbols. Suppose that X, Y and Z occur on different trials. Is the occurrence of Z predicted by Y or X? Or some more complex situation?

Suppose that Z only occurs if it is caused by X or Y. From Eqs. 8 and 10 and basic properties of random variables, we could compare

$$M_z{}^y(s = 0) = P(z|y) \tag{15}$$

to

$$M_z{}^x(s = 0)\bar{M}_x{}^y(s = 0) = P(z|x)P(x|y) \tag{16}$$

If Eqs 15 and 16 are equal to one another, then credit for Z should go to X rather than Y. To the extent they differ, then Y should get credit for the occurrence of Z.

Of course there are limits to how well the future can be predicted with pairwise statistics. More generally, we would like to consider joint statistics. This requires estimating higher order probabilities, e.g., $P(x, z|y)$. We establish later that joint statistics can be estimated from $\mathbf{M}(\rho, s)$. In an environment where joint statistics are important, predicting the future using simple pairwise relationships is untenable. However, it should be possible to recode the symbols into a new set of symbols that can be used to predict the future using pairwise relationships.

***Temporal contingency***

So that we can focus on temporal contingency, let us assome that all three stimuli are presented on each trial so that $P(y|x) = P(z|y) = P(z|x) = 1$. Because $M_y{}^x(s)$ contains information about every moment of the distribution $\tau_{xy}$ it is straightforward to ask whether the distribution of times for Z conditionalized on Y is higher or lower entropy than the distribution conditionalized on X. It is also possible to use $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$ to capture more subtle temporal relationships.

Recall that the distribution of the sum of two random variables equals the convolution of those random variables if they are independent of one another. Thus comparing the distribution of $\tau_{yz} = \tau_{xz} - \tau_{xy}$ to the distribution of the convolution $\tau_{xz} * (-\tau_{xy})$ allows us to assess the dependency across trials of the timing of the three stimuli. Equation 12 shows that the Laplace transform of $\tau_{xz} - \tau_{xy}$ is stored in $M_z{}^y(s)$, whereas Equation 14 shows that the Laplace transform of $\tau_{xz} - \tau_{xy}$ is stored in $M_z{}^x(s)\bar{M}_x{}^y(-s)$. Comparing these two quantities allows us to assess the dependence between the times of occurrence of X and Z conditionalized on Y in the present.

## 2.4 Continuum of $\rho$ allows a temporal memory *across* trials

For the past several subsections we have considered the limit where $\rho \to 1$. That limit is not physically realizable. How should we choose the value of $\rho$? The answer is that we

should not choose a single value of $\rho$. In much the same way we treat $s$ as a continuous variable rather than treating it as a parameter to be estimated from the data, we can also treat $\rho$ as a continuous variable. Continuous $s$ means that $\mathbf{F}^-(s)$ maintains a temporal memory of the entire past. Similarly, continuous $\rho$ enables $\mathbf{M}(\rho, s)$ to retain complete information about pairwise relationships as a function of trial history. Similar relationships can be worked out for $\bar{\mathbf{M}}(\rho, s)$ but we focus on $\mathbf{M}(\rho, s)$ here for simplicity.

Equation 7, which describes the situation where $\tau_{xy}$ is equal to $\tau_o$ on each trial, can be rewritten as

$$M_y{}^x(\rho, s) = (1 - \rho)e^{-s\tau_o}\mathcal{Z}\left\{h[i]\right\}\left(\rho^{-1}\right)$$

where $\mathcal{Z}\left\{\right\}(z)$ is the Z-transform, the discrete analog of the Laplace transform (Ogata, 1970). An analogous relationship can be written for $\bar{\mathbf{M}}(s)$.

Although the notation is a bit more unwieldy, allowing $\tau_{xy}$ to vary across trials we see that the trial history of timing is also retained by $\mathbf{M}(\rho, s)$. Writing the delay between X and Y on the trial $i$ steps in the past as $\tau[i]$, and $H[i](s) \equiv h[i]e^{-s\tau[i]}$ we can write

$$M_y{}^x(\rho, s) = (1 - \rho)\mathcal{Z}\left\{H[i](s)\right\}\left(\rho^{-1}\right). \tag{17}$$

We understand the Z-transform to be taken over the discrete variable $i$ and not the continuous variable $s$.

Because the Z-transform is in principle invertible, information about the entire trial history has been retained by virtue of having a continuum of forgetting rates $\rho$. Figure 6 illustrates the ability to extract the trial history including timing information of events that follow X from $\mathbf{M}(\rho, s)\mathbf{x}$.

This illustrates a remarkable property of Laplace-based temporal memory. Although each synaptic matrix with a specific value of $\rho_o$ forgets exponentially with a fixed time horizon (the time constant is given by $(-\log \rho_o)^{-1}$), the set of matrices with a continuum of $\rho$ retains information about the *entire* trial history. Although each matrix has a specific time horizon, the set of all matrices with continuous values of $\rho$ has a continuity of time horizons, tiling the entire trial history. In practice there must be some numerical imprecision in the biological instantiation of $\mathbf{M}(\rho, s)$. In principle however, a continuum of forgetting rates $\rho$ means that the past is not forgotten. Instead the past, as a function of trial history, has been written across the continuum of $\rho$.

## 2.5 Estimating three point correlation functions from Z-transform

A great deal of information can be extracted from the trial histories encoded in $\mathbf{M}(\rho, s)$ and $\bar{\mathbf{M}}(\rho, s)$. $M_z{}^y(s)$ contains the two-point probability distribution of Y and Z. It would be preferable to predict the occurrence and timing of Z using the three-point probability distribution.

Because $\mathbf{M}(\rho, s)$ contains information about the paired trial history, in principle we can extract information about the three-point correlation function. The problem

16
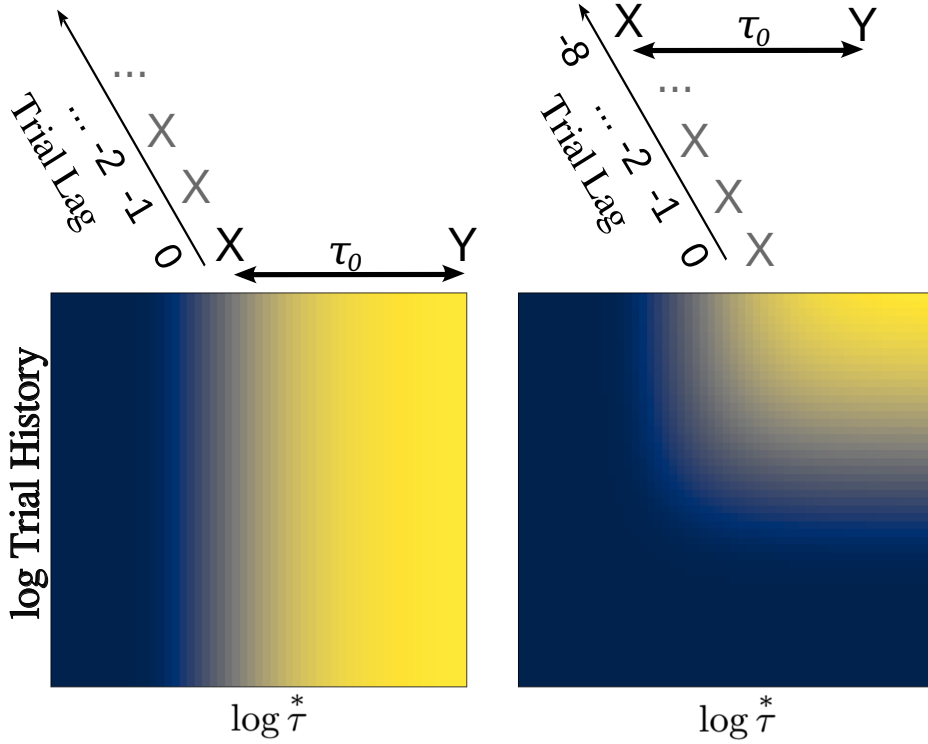
**Fig. 6** $\mathbf{M}(\rho, s)$ contains information about both time within a trial and trial history. Left: Consider a single pairing of X and Y on the most recent trial. The heatmap shows the degree to which Y is cued by X by $\frac{\mathbf{y}'\mathbf{M}(\rho,s)\mathbf{x}}{1-\rho}$ projected onto log time. The profile as a function of $\log \tau$ is identical to the profile for future time in Figure 2. If the pairing between X and Y had a longer delay, the edge would be further to the right. Right: The single pairing of X and Y is followed by an additional series of trials on which X was presented by itself. Now there is an edge in both trial history and time within trial. Additional trials with only X would push this edge further towards the top of the graph. Additional trials with X and Y paired would be added to this plot with a time delay that reflects the timing of the pairing.

of estimating the three-point correlation function between stimuli is straightforward if one has access to the trial history of both the past conditionalized on the present and the future conditionalized on the present. This information is contained in $\bar{M}_x{}^y(\rho, s)$ and $M_z{}^y(\rho, s)$ respectively.

For instance, if Z only occurs on trials on which both X and Y are presented, but not on trials when only one of them are presented, then we should observe a positive correlation between the trial history encoded in $M_z{}^y(\rho, s = 0)$ and $\bar{M}_x{}^y(\rho, s = 0)$. Similarly, one can imagine that the joint timing of the presentations of X and Y predicts the timing of Z, as if all three symbols are being generated by a process that can unfold at different rates.

17

Access to the joint statistics between symbols can in principle be leveraged to provide a much more complete prediction of the future, especially when integrated into deep networks that recode the symbols into new sets of symbols. Moreover, continuous values of $\rho$ may allow networks built using $\mathbf{M}(\rho, s)$ and $\bar{\mathbf{M}}(\rho, s)$ to respond to non-stationary statistics.

## 2.6 Updating the future

Let us return to the problem of generating a prediction of the immediate future. We again restrict our attention to the limit as $\rho$ goes to 1 and assume the system has experienced a very long sequence of trials with the same underlying statistics. Moreover, we assume for the present that only pairwise relationships are important, so we can neglect the temporal credit assignment problem, and construct the Laplace transform of the future that predicted solely on the basis of the present stimulus.

There are two problems that need to be resolved to write an analog of Eq. 3 for $\mathbf{F}_{t+\Delta t}^+(s)$. First, we can only use Eq. 2 to update $\mathbf{F}_t^+(s)$ if $\mathbf{F}_t^+(s)$ is already the Laplace transform of a predicted future; we must create a circumstance that makes that true. Second, we need to address the situation where a prediction reaches the present. Because of the discontinuity at $\tau = 0$ special considerations are necessary to allow the time of a stimulus to pass from the future to the past.

### 2.6.1 Predicting the future with the present

Equation 8 indicates that the weights in $M_y{}^x(s)$ record the future occurrences of Y given that X occurs in the present. $M_y{}^x(s)$ captures both the probability that Y will follow X as well as the distribution of temporal delays at which Y is expected to occur. This information is encoded as a probability times the Laplace transform of a random variable. If we only need to consider X in predicting the future, then $M_y{}^x(s)$ is precisely how we would like to initialize the future prediction for Y in $\mathbf{F}_t^+(s)$ after X is presented (Fig. 4).

We probe $\mathbf{M}(s)$ with the "immediate past." When X is presented it enters $\mathbf{F}_t^-(s)$ as $\mathcal{L}\left\{\delta\left(0\right)\mathbf{x}\right\}(s)$. Multiplying $\mathbf{M}(s)$ from the right with the immediate past, yields a prediction for the future.

$$\mathbf{M}(s) \ e^{-s0} \ \mathbf{x} = \mathbf{M}(s)\mathbf{x} = P\left(y|x\right)\mathcal{L}\left\{\tau_{xy}\right\}(s) \ \mathbf{y} \tag{18}$$

More generally, the input to the future at time $t$ should be given by $\mathbf{M}(s)\mathcal{L}\left\{\delta(0)\mathbf{f}_t\right\}$. For concision we write this as $\mathbf{M}(s)\mathbf{f}_t$. Because the past stored in $\mathbf{M}(s)$ was a probability times the Laplace transform the distribution of a random variable, the future recovered in this way is also understandable as a probability times the Laplace transform of a random variable. If only Laplace transforms of delta functions can be represented in $\mathbf{F}_t^+(s)$, then we can imagine sampling from this distribution of future times, perhaps with a preference for times more near to the future.

18

### 2.6.2 Continuity of the predicted future through $\tau = 0$

The neural representation described here approximates a continuous timeline by stitching together separate Laplace neural manifolds for the past and the future. With the passage of time, information in the future moves ever closer to the present. As time passes and a prediction reaches the present, this discontinuity must be addressed. Otherwise, the firing rates will grow exponentially without bound.

We can detect predictions that have reached the present by examining $\mathbf{F}_t^+(s = \infty)$, which only rises from zero when $\tau \to 0$. In practice, we would use $s_{\max}$ which should be on the order of $(\Delta t)^{-1}$. If the future that is being represented is the Laplace transform of a delta function, then we can simply take components for which $\mathbf{F}_t^+(s_{\max}) > 0$ to zero for all $s$ at the next time step. More generally, if the future that is represented is not simply a delta function, the linearity of the Laplace transform allows us to subtract $\mathbf{F}_t^+(s = \infty)$ from all $s$ values without affecting the evolution at subsequent time points.

If a prediction reaches the present and is observed, then no further action is needed. If a prediction reaches the present, but is not observed, we can trigger an observation of a "not symbol", written e.g., x̃ to describe the observation of a failed prediction for a stimulus x. Although we won't pursue it here, one could allow "not symbols" to be predicted by stimuli and to predict other stimuli, allowing for the model to provide appropriate predictions for a relatively complex set of contingencies using only pairwise relationships.

### 2.6.3 Evolving $\mathbf{F}_{t+\Delta t}^+(s)$

Integrating these two additional factors allows us to write a general expression for evolving $\mathbf{F}_t^+(s)$ to $\mathbf{F}_{t+\Delta t}^+(s)$.

$$\mathbf{F}_{t+\Delta t}^+(s) = e^{s\Delta t}\mathbf{F}_t^+(s) - \mathbf{F}_t^+(s = \infty) + \mathbf{M}(s)\mathbf{f}_t. \tag{19}$$

If the future is expressed as a delta function, continuous attractor networks with an edge are sufficient to support this evolution (Daniels & Howard, submitted). Because the future is in general more complex than a delta function, and predictions for distant parts of the future can change as events happen in the present, additional considerations are necessary.

## 3 Neural predictions

Regions as widely separated as the cerebellum (De Zeeuw, Lisberger, & Raymond, 2021; Wagner & Luo, 2020), striatum (e.g., van der Meer & Redish, 2011), PFC (e.g., Ning, Bladon, & Hasselmo, 2022; Rainer, Rao, & Miller, 1999), OFC (e.g., Namboodiri et al., 2019; Schoenbaum, Chiba, & Gallagher, 1998; Young & Shapiro, 2011), hippocampus (Duvelle, Grieves, & van der Meer, 2023; Ferbinteanu & Shapiro, 2003) and thalamus (Komura et al., 2001) contain active representations that code for the future. One can find evidence of predictive signals extending over long periods of time that modulate firing in primary visual cortex (Gavornik & Bear, 2014; Homann, Koay,

Chen, Tank, & Berry, 2022; H. Kim, Homann, Tank, & Berry, 2019; Yu et al., 2022). Prediction apparently involves a substantial proportion of the brain. Coordinating activity and plasticity over such a wide region would require careful synchronization (Hamid, Frank, & Moore, 2021; Hasselmo, Bodelón, & Wyble, 2002). The timescale of this synchronization, presumably on the order of 100 ms, fixes $\Delta t$, places a bound on the fastest timescales $1/s$ that can be observed, and operationalizes the duration of the "present."

Given the widespread nature of predictive signal, we will not attempt to map specific equations onto specific brain circuits. Rather we will illustrate the observable properties implied by these equations with an eye towards facilitating future empirical work. The predictions fall into two categories. One set of predictions describes properties of ongoing firing of neurons participating in Laplace Neural Manifolds for past and future time. Another set of predictions are a direct consequence of the properties of learned weights. We also briefly discuss the model in this paper in the context of recent empirical work on the computational basis of the dopamine signal (Jeong et al., 2022).

## 3.1 Active firing neurons

This paper proposes the existence of Laplace Neural Manifolds to code for the identity and time of *future* events. This implies there should be two related manifolds, one implementing the Laplace space and one implementing the inverse space. Previous neuroscientific work has shown evidence for Laplace and inverse spaces for a timeline for the past. The properties of the proposed neural manifolds for future time can be understood by analogy to the neural manifolds for the past.

### 3.1.1 Single-cell properties of neurons coding for the past

So-called temporal context cells observed in the entorhinal cortex (Bright et al., 2020; Tsao et al., 2018) are triggered by a particular event and then relax exponentially back to baseline firing with a variety of time constants. The firing of temporal context cells is as one would expect for a population coding $F^-(s)$. So-called time cells observed in the hippocampus (MacDonald et al., 2011; Pastalkova et al., 2008; Schonhaut, Aghajan, Kahana, & Fried, 2022; Shahbaba et al., 2022; Shikano, Ikegaya, & Sasaki, 2021; Taxidis et al., 2020) and many other brain regions (e.g., Akhlaghpour et al., 2016; Bakhurin et al., 2017; Jin, Fujii, & Graybiel, 2009; Mello, Soares, & Paton, 2015; Subramanian & Smith, 2024; Tiganj et al., 2018; Tiganj, Kim, Jung, & Howard, 2017) fire sequentially as events recede into the past, as one would expect from neurons participating in $\tilde{f}(\overset{*}{\tau})$ for $\overset{*}{\tau} < 0$. Time cells are consistent with qualitative and quantitative predictions, including the conjecture that time constants are distributed along a logarithmic scale (Cao et al., 2022).

### 3.1.2 Single-cell and population-level properties of neurons coding for the past and the future

In situations where the future can be predicted, $F^+(s)$ and $\tilde{f}(\overset{*}{\tau} > 0)$ should behave as mirror images of the corresponding representations of the past. Figure 7A illustrates

the firing of cells coding for a stimulus remembered in the past (left) and predicted in the future (right). Neurons participating in the Laplace space, sorted on their values of $s$, are shown in the top; neurons participating in the inverse space, sorted on their values of $\overset{*}{\tau}$ are shown on the bottom.

The firing of neurons constituting the Laplace space shows a characteristic shape when plotted as a function of time in this simple experiment. Neurons coding for the past are triggered shortly after presentation of the stimulus and then relax exponentially with a variety of rates. Neurons coding for the future ramp up, peaking as the predicted time of occurrence grows closer. The ramps have different characteristic time constants. Different populations are triggered by the presentation of different symbols (not shown) so that the identity of the remembered and predicted symbols as well as their timing can be decoded from populations coding $\mathbf{F}^-(s)$ and $\mathbf{F}^+(s)$. The largest value of $1/s$ in the figure is chosen to be a bit longer than the delay in the experiment, resulting in a subset of neurons that appear to fire more or less constantly throughout the delay (Enel, Wallis, & Rich, 2020).

The firing of neurons constituting the inverse space also shows a characteristic shape when plotted as a function of time in this simple experiment. Neurons tile the delay, with more cells firing early in the interval with more narrow receptive fields. The logarithmic compression of $n$ results in a characteristic "backwards J" shape for the past and a mirror image "J" shape for the future. Again, different populations would code for different stimuli in the past and in the future (not shown) so that the identity of the remembered and predicted stimuli and their time from the present could be decoded from a population coding $\tilde{\mathbf{f}}(\overset{*}{\tau})$. Figure 7B shows firing that would be expected for a population that includes cells coding for the same stimulus, say Y, both in the past and the future around the time of a predicted occurrence of that symbol.

### 3.1.3 Plausible anatomical locations for an internal future timeline

This computational hypothesis should evaluated with carefully planned analyses. However, the published literature shows evidence that is at least roughly in line with the hypothesis of neural manifolds for future time. Firing that ramps systematically upward in anticipation of important outcomes including planned movements has been observed in (at least) mPFC (Henke et al., 2021), ALM (Inagaki, Inagaki, Romani, & Svoboda, 2018), cerebellum (Garcia-Garcia et al., 2024), and thalamus (Komura et al., 2001). Komura et al. (2001) showed evidence for ramping firing in the thalamus that codes for outcomes in a Pavlovian conditioning experiment. Two recent papers show evidence that ramping neurons during motor preparation in ALM (Affan et al., 2024; Inagaki et al., 2018) and interval timing in mPFC (Cao et al., in press; Henke et al., 2021) do so with a continuous spectrum of time constants.

For instance, Cao et al. (in press), reanalyzing data originally published by Henke et al. (2021) observed the firing of neurons during the reproduction phase of an interval reproduction task. On each trial, the animal is exposed to a delay of $T$ seconds, which must then be reproduced. Let us refer to the moment the reproduction phase begins as $t = 0$. Now at time $t < T$, the beginning of the interval is $\tau = t$ seconds in the past and the planned movement is a time $\tau = T - t$ seconds in the future. Figure 9A shows that some neurons in mPFC ramped down as $e^{-st}$ with a continuum of rate constants
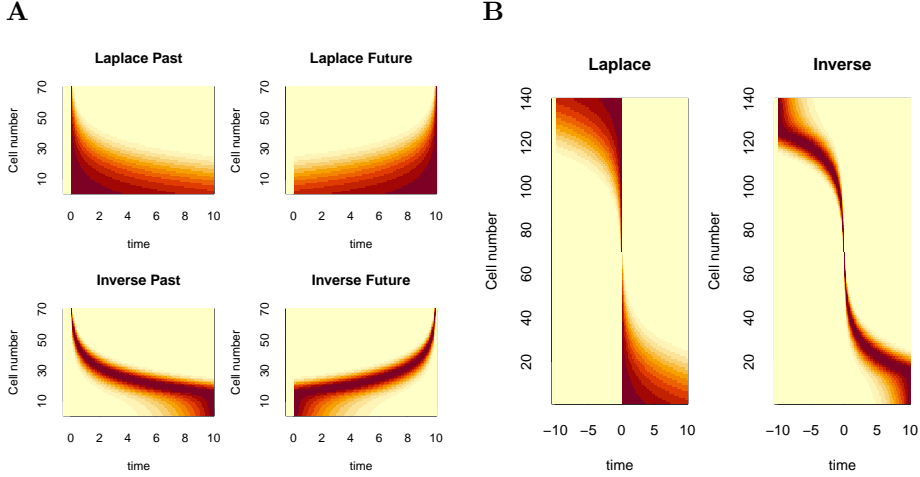
**Fig. 7** Predicted firing for Laplace and inverse spaces plotted as heatmaps. **A.** Consider an experiment in which X precedes Y separated by 10 s. The top row shows firing as a function of time for cells in the Laplace space for the past (left) and the future (right). Note that the cells in $F_t^-(s)$ peak at time zero and then decay exponentially. In contrast cells in $F_t^+(s)$ peak at 10 s and ramp up exponentially. The bottom row shows firing as a function of time for cells in the Inverse space. **B.** Consider an experiment in which Y is predicted to occur at time zero and then recedes into the past. Cells coding for both past and future are recorded together and sorted on the average time at which they fire. Left: For Laplace spaces, neurons in $F_t^+(s)$ are sorted to the top of the figure and neurons $F_t^-(s)$ are sorted to the bottom of the figure. Right: Inverse spaces show similar properties but give rise to a characteristic "pinwheel" shape.

$s$ and other neurons ramped up as $e^{-s(T-t)}$, again with a continuum of rate constants $s$. There are potentially important differences between the empirical results in the paper and the theoretical model presented here—for instance many of the neurons coding for the time of the future planned movements rescaled the timecourse of their firing depending on the value of $T$ on that trial—but the overall correspondence to the predictions described here is striking. In at least some regions, in some tasks, the ongoing firing of cortical neurons codes the time of planned future events *via* real Laplace transform of the future.

There is also circumstantial neurophysiological evidence for sequential firing leading to predicted future events as predicted by $\tilde{f}(\overset{*}{\tau})$ for $\overset{*}{\tau} > 0$ coding for future events. Granule cells in cerebellum appear to fire in sequence in the time leading up to an anticipated reward (Wagner, Kim, Savall, Schnitzer, & Luo, 2017; Wagner & Luo, 2020). During performance of a task in which monkeys must perform a sequence of movements, neurons firing in sequence that decoded the time of *future* movements were observed in PFC but not in posterior parietal cortex (Watanabe, Kadohisa, Kusunoki, Buckley, & Duncan, 2023). OFC may be another good candidate region to look for "future time cells." OFC has long been argued by many authors to code for the identity of predicted outcomes (Hikosaka & Watanabe, 2000; Mainen & Kepecs, 2009; Schoenbaum & Roesch, 2005). More recently Enel et al. (2020) showed sequential activation in OFC during a task in which it was possible to predict the value of a reward that was delayed for several seconds. Finally, it should be noted that the
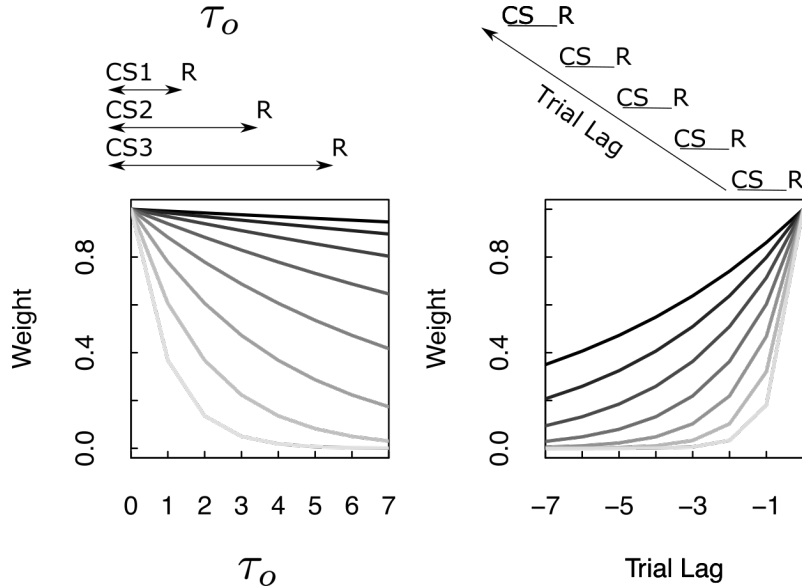
**Fig. 8** Neural predictions derived from properties of $\mathbf{M}(s)$. **Left.** Plot of the magnitude of the entry in $\mathbf{M}_r(\rho = 1, s)$ connecting each of the conditioned stimuli CS to the outcome R as a function of the $\tau_o$ corresponding to that CS. Different lines correspond to entries with different values of $s$. Weights corresponding to different values of $s$ show exponential discounting as $\tau_o$ is manipulated, with a variety of discount rates. **Right.** Plot of the magnitude of $\mathbf{M}(\rho, s = 0)$ associated with a single pairing of CS and R a certain number of trials in the past. Different lines show the results for different values of $\rho$. For clarity, these curves have been normalized such that they have the same value at trial lag zero.

properties of $\tilde{f}(\overset{*}{\tau})$ over the future are a temporal analog of spatial "distance-to-goal" cells that have been observed in spatial navigation studies (Gauthier & Tank, 2018; Sarel, Finkelstein, Las, & Ulanovsky, 2017).

## 3.2 Predictions from weight matrix M($\rho, s$)

### 3.2.1 Properties of weights due to $s$

Consider an experiment in which different symbols, denoted CS1, CS2, etc, precede an outcome R by a delay $\tau_o$. The value of $\tau_o$ changes across the different symbols (Figure 8A). Ignoring $\rho$ for the moment, the strength of the connections from each CS to R depend on the value of $\tau_o$ for that stimulus and the value of $s$ for each synapse: $e^{-s\tau_o}$. When a particular CS is presented at time $t$, the amount of information that flows along each synapse is $e^{-s\tau_o}$ and the pulse of input to $\mathbf{F}^+_{t+\Delta t}(s) - \mathbf{F}^+_t(s)$ corresponding to the outcome is $e^{-s\tau_o}$.

Thus, considering each connection as a function of $\tau_o$, firing should go down exponentially as a function of $\tau_o$ with a rate constant that depends on the value of $s$. This pattern of results aligns well with experimental results observed in mid-brain dopamine neurons (Masset et al., 2023; Sousa et al., 2023). It has long been known that firing of dopamine neurons, averaged over neurons, around the time of the conditioned stimulus goes down with delay (Fiorillo, Newsome, & Schultz, 2008). Masset et
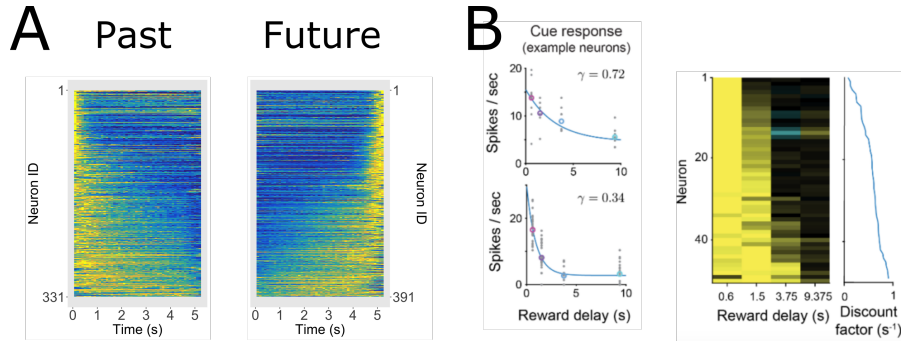
**Fig. 9** Recent observation of key neural predictions of this approach. **A.** In this interval reproduction experiment, rodents had to reproduce a delay of some duration $T$. Calling the start of the reproduction period $t = 0$, at time $t$, the beginning of the interval is $t$ seconds in past while the planned end of the reproduction period is $T - t$ seconds in the future. Firing of neurons in rodent mPFC have properties resembling those predicted for $F_t^-(s)$ and $F_t^+(s)$, showing exponential decay/ramping with a continuous spectrum of time constants. Compare to Figure 7A. Adapted from Cao, et al., (2024). **B.** In a classical conditioning experiment, different stimuli predicted a rewarding outcome at different delays. Firing of dopamine neurons was recorded following each of the stimuli. The change in firing as function of the time until the reward was fitted with an exponential curve (left), indexed by the discount rate $\gamma$. Across dopamine neurons, a wide range of time constants was observed (right). The results are as one would expect if the dopamine system projects information about the time of future events to the rest of the brain *via* $\mathbf{M}(s)$. Adapted from Masset, et al., (2023).

al. (2023) measured the firing of dopamine neurons to different stimuli that predicted reward delivery at different delays. This study showed that there was a heterogeneity of exponential decay rates in the firing of dopamine neurons in this paradigm (Fig. 9B), much as illustrated in Fig. 8A. In the context of TDRL, this finding is consistent with a continuous spectrum of exponential discount rates (Momennejad & Howard, 2018; Tano et al., 2020). In any event, these findings (Masset et al., 2023; Sousa et al., 2023) are clear evidence that the phasic firing of midbrain dopamine neurons at the time of a predictive stimulus codes for the Laplace transform of the time until future reward.

### 3.2.2 Properties of weights due to $\rho$

A continuum of forgetting rates $\rho$ predicts a range of trial history effects. Figure 8B shows the weights in $\mathbf{M}(\rho)$ over past trials that result from different values of $\rho$. This is simply $\rho^i$ where $i$ is the trial recency with values normalized such that the weight at the most recent trial is 1. The weights $\mathbf{M}(\rho)$ record the Z-transform of the trial history of reinforcement. Many papers show dependence on previous trial outcomes in response to a cue stimulus in learning and decision-making experiments (Akrami, Kopec, Diamond, & Brody, 2018; Bernacchia, Seo, Lee, & Wang, 2011; Hattori, Danskin, Babic, Mlynaryk, & Komiyama, 2019; Hattori & Komiyama, 2022; Morcos & Harvey, 2016; Scott et al., 2017). These studies show history-dependent effects in a wide range of brain regions and often show a continuous spectrum of decay rates within a brain region (see especially Bernacchia et al., 2011; Danskin et al., 2023). Notably, distributions of time constants for trial history effects cooccur

with distributions of ongoing activity in multiple brain regions (Spitmaan, Seo, Lee, & Soltani, 2020).

## 3.3 Dopamine and learning

The connection between TDRL and neuroscience related to dopamine has been one of the great triumphs of computational neuroscience (Schultz et al., 1997). The standard account is that the firing of dopamine neurons signals reward prediction error (RPE) which drives plasticity. Despite its remarkable success at predicting the findings of many behavioral and neurophysiological experiments, the RPE account has been under increasing strain over recent years. The standard account did not predict the existence of a number of striking effects, including increasing dopamine firing during delay under uncertainty (Fiorillo, Tobler, & Schultz, 2003), dopamine ramps in spatial experiments (Howe, Tierney, Sandberg, Phillips, & Graybiel, 2013), dopamine waves (Hamid et al., 2021), and heterogeneity of dopamine responses across neurons and brain regions (Dabney et al., 2020; Masset, Malik, Kim, Bech Vilaseca, & Uchida, 2022; W. Wei, Mohebi, & Berke, 2021), although many of these phenomena can be accommodated within the RPE framework with elaboration. (Gardner, Schoenbaum, & Gershman, 2018; Gershman, 2017; H.R. Kim et al., 2020; Lee, Engelhard, Witten, & Daw, 2022) Jeong et al. (2022) reported the results of several experiments that flatly contradict the standard model. These experiments were proposed to evaluate an alternative hypothesis for dopamine firing in the brain.

Jeong et al. (2022) propose that dopamine signals whether the current stimulus is a cause of reward. The model developed there, referred to as ANCCR, assesses the contingency between a stimulus and outcomes. $\mathbf{M}(\rho, s)$ and $\bar{\mathbf{M}}(s)$ contain information about the contingencies—temporal and otherwise—between a symbol and possible outcomes. Both ANCCR and the framework developed in this paper are inspired by a similar critique of Rescorla-Wagner theory and TDRL (C. Gallistel, 2021b). In order to make a complete account of the experiments in the (Jeong et al., 2022) paper, the current framework would have to be elaborated in several ways. However, the current framework does not require one to specify an intrinsic timescale of association *a priori*. Perhaps it is possible to develop a generalization of the current framework that does not rely on the simplifying assumption of discrete trials in order to yield readily interpretable measures of contingency.

# 4 Discussion

This paper takes a phenomenological approach to computational neuroscience. The strategy is to write down equations that, if the brain could somehow obey them, would be consistent with a wide range of observed cognitive and neural phenomena. The phenomenological equations make concrete predictions that can be evaluated with cognitive and neurophysiological experiments. To the extent the predictions hold, the question of how the brain manages to obey these phenomenological equations could then become a separate subject of inquiry. The phenomenological equations require a number of capabilities of neural circuits, both at the level of synapses and in terms of ongoing neural activity. We make those explicit here.

## 4.1 Circuit assumptions for synaptic weights

$\mathbf{M}(\rho, s)$ and $\bar{\mathbf{M}}(\rho, s)$ require that the brain uses continuous variables, $\rho$ and $s$, to organize connections between many neurons, most likely spanning multiple brain regions. For the phenomenological equations to be viable, these continuous variables should be deeply embedded in the functional architecture of the brain. For instance, in order to invert the integral transforms, it is necessary to compute a derivative over these continuous variables. This suggests a gradient in these continuous variables should be anatomically identifiable. Conceivably anatomical gradients in gene expression and functional architecture (e.g., Guo et al., 2021; Phillips et al., 2019; Roy, Zhang, Halassa, & Feng, 2022) could generate anatomical gradients in $s$ and/or $\rho$. Perhaps part of the function of traveling waves of activity such as theta oscillations (Lubenov & Siapas, 2009; Patel, Fujisawa, Berényi, Royer, & Buzsáki, 2012; Zhang & Jacobs, 2015) or dopamine waves (Hamid, Frank, & Moore, 2019) is to make anatomical gradients salient.

## 4.2 Circuit assumptions for ongoing activity

At the neural level, this framework assumes the existence of circuits that can maintain activity of a Laplace Neural Manifold over time. There is evidence that the brain has found some solution to this problem (Atanas et al., 2023; Bright et al., 2020; Cao et al., in press; Tsao et al., 2018; Zuo et al., 2023). Exponential growth of firing, as proposed by Eq. 19 seems on its face to be a computationally risky proposition (but see Daniels & Howard, submitted). However, this proposal does create testable predictions. Moreover, firing rates that increase monotonically as a function of one or another environmental variable are widely observed. For instance border cells as an animal approaches a barrier (Solstad, Boccara, Kropff, Moser, & Moser, 2008) and evidence accumulation cells (Roitman & Shadlen, 2002) both increase monotonically. If this monotonic increase in firing reflects movement of an edge along a Laplace Neural Manifold, the characteristic time scale of the increase should be heterogeneous across neurons. If the brain has access to a circuit with paired $\alpha$'s, it could reuse this circuit to construct cognitive models for spatial navigation (Howard et al., 2014), evidence accumulation (Howard et al., 2018), and perhaps cognitive computation more broadly (Howard & Hasselmo, 2020). Consistent with this hypothesis, monotonic cells in spatial navigation and evidence accumulation—border cells and evidence accumulation cells— have sequential analogues (Koay, Charles, Thiberge, Brody, & Tank, 2022; Morcos & Harvey, 2016; Wilson & McNaughton, 1993) as one would expect if they reflect a Laplace space that is coupled with an inverse space.

Perhaps part of the solution to implementing these equations in the brain is to restrict the kinds of functions that can be represented over the Laplace Neural Manifold. A continuous attractor network that can maintain and evolve the Laplace transform of a single delta function per basis vector can readily be constructed (Daniels & Howard, submitted). In this case, each component of $\mathbf{F}_t^-(s)$ and $\mathbf{F}_t^+(s)$ would be at any moment the Laplace transform of a delta function; $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$ would still be able to store distributions over multiple presentations. In this case when an item is

presented perhaps $\mathbf{F}^+_{t+\Delta t}(s)$ could update by sampling from the distribution expressed by $\mathbf{M}(s)\mathbf{f}_t$.

## 4.3 Generalizing beyond time

It should be possible to extend the current framework to multiple dimensions beyond time, including real space and abstract spaces (Howard et al., 2018, 2014). Properties of the Laplace domain enable data-independent operators that enable efficient computation (Howard & Hasselmo, 2020). For instance, given that a state of a Laplace neural manifold is the Laplace transform of a function, we can construct the Laplace transform of the translated function (Eq. 2, see also Shankar et al., 2016). Critically, the translation operator is independent of the function to be translated. Restricting our attention to Laplace transforms of delta functions, we can construct the sum or difference using convolution and cross correlation respectively (Howard & Hasselmo, 2020; Howard et al., 2015). The binary operators for addition and subtraction also do not need to be learned. Perhaps the control theory that governs behavior is analogous to generic spatial navigation in a continuous space.

## 4.4 Scale-covariance as a design goal

Because the $s$ values are sampled along a logarithmic scale, all of the quantities in this paper are scale-covariant. Rescaling time, taking $\tau_{xz} \to a\tau_{xz}$, $\tau_{xy} \to a\tau_{xy}$, etc, simply takes $s \to s/a$. Because the $s$ values are chosen in a geometric series, rescaling time simply translates along the $n$ axis. All the components of the model, $\mathbf{F}^-$, $\mathbf{F}^+$, $\mathbf{M}$, and $\bar{\mathbf{M}}$, all use the same kind of logarithmic scale for time. All of the components of the model are time scale-covariant, responding to rescaling time with a translation over cell number. Thus any measure that integrates over $n$ (and is not subject to edge effects) is scale-invariant.

Empirically, there is not a characteristic time scale to associative learning (Balsam & Gallistel, 2009; Burke et al., 2023; C.R. Gallistel & Shahan, 2024; Gershman, 2022); any model that requires choice of a time scale for learning to proceed is thus incorrect. Logarithmic time scales are observed neurally (Cao et al., 2022; Guo et al., 2021). Logarithmic time scales can be understood as a commitment to a world with power law statistics (Piantadosi, 2016; X.-X. Wei & Stocker, 2012) or as an attempt to function in many different environments without a strong prior on the time scales it will encounter (Howard & Shankar, 2018).

Recent work has shown that the use of logarithmic time scales also enables scale-invariant CNNs for vision (Jansson & Lindeberg, 2021) and audition (Jacques, Tiganj, Sarkar, Howard, & Sederberg, 2022). For instance, (Jacques et al., 2022) trained deep CNNs to categorize spoken digits. When tested on digits presented at very different speeds than the training examples (imagine someone saying the word "seven" stretched over four seconds), the deep CNN with a logarithmic time axis generalized perfectly. Rescaling time translates the neural representation at each layer; convolution is translation equivariant; including a maxpool operation over the convolutional layer renders the entire CNN translation-invariant. Time is not only important in speech perception

27

(e.g., Lerner, Honey, Katkov, & Hasson, 2014) but vision as well (Russ, Koyano, Day-Cooney, Perwez, & Leopold, 2022) suggesting that these ideas can be incorporated into a wide range of sensory systems.

## 4.5 Convolution, relational memory and cognitive graphs

There is a long-standing tension in psychology between accounts of learning based on simple associations and cognitive representations. For instance Tolman (1948) contrasted behaviorist accounts of stimulus-response associations with a "cognitive map" studying the behavior of rats in spatial mazes. This paper has already touched on this tension between association and temporal contingency—which requires metric temporal relationships between stimuli—in the study of Pavlovian learning and reward systems in the brain. (Fodor & Pylyshyn, 1988) used analogous arguments in an early critique of connectionism that echoes to the present day in contemporary debates about whether large language models "understand" language or not. Continuing interest in "neurosymbolic" artificial intelligence can be seen as an extension of this longstanding debate (Marcus, 2018).

For researchers studying episodic memory and neural representations in the hippocampus, cognitive maps rather than simple atomic associations have long been the dominant view (O'Keefe & Nadel, 1978). Cohen and Eichenbaum (1993) emphasized that cognitive maps are more general than spatial maps of the physical environment and can be used to describe other forms of relationships. In their view a *relational* representation "maintains the 'compositionality' of the items, that is, the encoding of items both as perceptually distinct 'objects' and as parts of larger scale 'scenes' and 'events' that capture the relevant relations between them." In the view of Cohen and Eichenbaum (1993), relational memory is critical for flexible, context-dependent expression of stored knowledge, in much the same way that rats can take a novel shortcut to a reward in a pre-learned maze (Tolman, 1948). These ideas about relational memory have led to "neurosymbolic" computational models developed with specific attention to hippocampal function (Whittington et al., 2020).

The convolutions stored in the Laplace domain in $\mathbf{M}(s)$ and $\bar{\mathbf{M}}(s)$ are precisely relational representations. The convolution of two functions $f * g$ is neither $f$, nor $g$, but is composed from them. The convolution between the function X WAS $\tau$ SECONDS IN THE PAST and the function Y IS IN THE REARWARD PORTION OF THE FUTURE describes an "event" including X and Y in a particular relationship. If we substitute functions of physical space rather than functions of time, it would be straightforward to understand this convolution as a "scene" as proposed by Cohen and Eichenbaum (1993). Because simple Hebbian association of Laplace representations is is sufficient to perform convolution, it is straightforward to at least write down neural models for relatively complex data structures in the Laplace domain.

Not only do convolutions provide a way to implement relational representations as envisioned by Cohen and Eichenbaum (1993), they also lend themselves to flexible *expression* of memory. Convolution has an inverse operation, cross-correlation. So, if $h = f * g$, then $h \# f \simeq g$, where $\#$ is the cross-correlation. This property enables symbolic computation (Gayler, 2004; Schlegel, Neubert, & Protzel, 2022). For instance, consider convolutions of delta functions. If $f$ is a delta function at $x_f$ and $g$ is a delta

function at $x_g$, then $(f * g)(x)$ is a delta function at $x_f + x_g$. Convolution of delta functions is thus mapped onto addition and cross-correlation—which is just convolution with reflection of one of the functions along the $x$ axis—maps onto subtraction. With a bit of creativity to deal with positive and negative numbers, not unlike the treatment of a timeline that continues from $-\infty$ to $\infty$ used here, one can build a computational system that implements the group describing the reals under addition, clearly meeting the requirements for a symbolic computer. Coming back to the hippocampus, navigation in a physical space requires vector subtraction. For instance to know how to get from physical location $x_f$ to physical location $x_g$, we must be able to compute $x_g - x_f$. One can also perform spatial navigation in abstract spaces using the same data-independent operators (C.R. Gallistel & King, 2011; Howard & Hasselmo, 2020).

Laplace Neural Manifolds are thus well-suited to not only learn, represent, and store relationships between stimuli but also to flexibly re-express relational information in a context-appropriate manner using data-independent operators. These two properties make Laplace Neural Manifolds ideal for cognitive maps of both real and abstract spaces.

### Conflict of interest statement

MWH and PBS are co-founders of Cognitive Scientific AI, Inc., which could benefit indirectly from publication of this manuscript.

# References

Affan, R.O., Bright, I.M., Pemberton, L., Cruzado, N.A., Scott, B.B., Howard, M. (2024). Ramping dynamics in the frontal cortex unfold over multiple timescales during motor planning. *bioRxiv*, 2024–02,

Aghajan, Z.M., Kreiman, G., Fried, I. (2023). Minute-scale periodicity of neuronal firing in the human entorhinal cortex. *Cell Reports*, *42*(11), ,

Akhlaghpour, H., Wiskerke, J., Choi, J.Y., Taliaferro, J.P., Au, J., Witten, I. (2016). Dissociated sequential activity and stimulus encoding in the dorsomedial striatum during spatial working memory. *eLife*, *5*, e19507,

Akrami, A., Kopec, C.D., Diamond, M.E., Brody, C.D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, *554*(7692), 368–372,

Arcediano, F., Escobar, M., Miller, R.R. (2005). Bidirectional associations in humans and rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(3), 301-18,

Atanas, A.A., Kim, J., Wang, Z., Bueno, E., Becker, M., Kang, D., ... others (2023). Brain-wide representations of behavior spanning multiple timescales and states in c. elegans. *Cell*, *186*(19), 4134–4151,

Bakhurin, K.I., Goudar, V., Shobe, J.L., Claar, L.D., Buonomano, D.V., Masmanidis, S.C. (2017). Differential encoding of time by prefrontal and striatal network dynamics. *Journal of Neuroscience*, *37*(4), 854–870,

Balsam, P.D., & Gallistel, C.R. (2009). Temporal maps and informativeness in associative learning. *Trends in Neuroscience*, *32*(2), 73–78,

Bernacchia, A., Seo, H., Lee, D., Wang, X.J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, *14*(3), 366-72,

Blouw, P., Solodkin, E., Thagard, P., Eliasmith, C. (2016). Concepts as semantic pointers: A framework and computational model. *Cognitive science*, *40*(5), 1128–1162,

Bright, I.M., Meister, M.L.R., Cruzado, N.A., Tiganj, Z., Buffalo, E.A., Howard, M.W. (2020). A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proceedings of the National Academy of Sciences*, *117*, 20274-20283,

Burke, D.A., Jeong, H., Wu, B., Lee, S.A., Floeder, J.R., K Namboodiri, V.M. (2023). Few-shot learning: temporal scaling in behavioral and dopaminergic learning. *BioRxiv*, 2023–03,

Cao, R., Bladon, J.H., Charczynski, S.J., Hasselmo, M., Howard, M. (2022). Internally generated time in the rodent hippocampus is logarithmically compressed. *eLife*, *https://doi.org/10.7554/eLife.75353*, ,

Cao, R., Bright, I.M., Howard, M.W. (in press). Ramping cells in rodent mpfc encode time to past and future events via real laplace transform. *Proceedings of the National Academy of Sciences*, ,

Carvalho, W., Tomov, M.S., de Cothi, W., Barry, C., Gershman, S.J. (2024). Predictive representations: building blocks of intelligence. *Neural Computation*, 1–74,

Chater, N., & Brown, G.D.A. (2008). From universal laws of cognition to specific cognitive models. *Cognitive Science*, *32*(1), 36-67, https://doi.org/10.1080/03640210701801941

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204,

Cohen, N.J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: The MIT Press.

Cole, R.P., Barnet, R.C., Miller, R.R. (1995). Temporal encoding in trace conditioning. *Animal Learning & Behavior*, *23*(2), 144-153,

Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C.K., Hassabis, D., Munos, R., Botvinick, M. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature*, 1–5,

Daniels, B.C., & Howard, M.W. (submitted). Continuous attractor networks for laplace neural manifold representations of sparse functions. *Computational Brain and Behavior*, *https://arxiv.org/abs/2406.04545*, ,

Danskin, B.P., Hattori, R., Zhang, Y.E., Babic, Z., Aoi, M., Komiyama, T. (2023). Exponential history integration with diverse temporal scales in retrosplenial cortex supports hyperbolic behavior. *Science Advances*, *9*(48), eadj4897, https://doi.org/10.1126/sciadv.adj4897 Retrieved from https://www.science.org/doi/abs/10.1126/sciadv.adj4897 https://www.science.org/doi/pdf/10.1126/sciadv.adj4897

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, *5*(4), 613–624,

De Zeeuw, C.I., Lisberger, S.G., Raymond, J.L. (2021). Diversity and dynamism in the cerebellum. *Nature neuroscience*, *24*(2), 160–167,

Duvelle, É., Grieves, R.M., van der Meer, M.A. (2023). Temporal context and latent state inference in the hippocampal splitter signal. *Elife*, *12*, e82357,

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.

Enel, P., Wallis, J.D., Rich, E.L. (2020). Stable and dynamic representations of value in the prefrontal cortex. *Elife*, *9*, e54313,

Ferbinteanu, J., & Shapiro, M.L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron*, *40*(6), 1227-39,

Fiorillo, C.D., Newsome, W.T., Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. *Nature Neuroscience*, , https://doi.org/10.1038/nn.2159

Fiorillo, C.D., Tobler, P.N., Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898-902, https://doi.org/10.1126/science.1077349

Floeder, J.R., Jeong, H., Mohebi, A., Namboodiri, V.M.K. (2024). Mesolimbic dopamine ramps reflect environmental timescales. *bioRxiv*, ,

Fodor, J.A., & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1), 3–71,

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127-138,

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211-21, https://doi.org/10.1098/rstb.2008.0300

Gallistel, C. (2021a). The physical basis of memory. *Cognition*, *213*, 104533,

Gallistel, C. (2021b). Robert Rescorla: Time, information and contingency. *Revista de Historia de la Psicología*, *42*(1), 7–21,

Gallistel, C., Craig, A.R., Shahan, T.A. (2019). Contingency, contiguity, and causality in conditioning: Applying information theory and weber's law to the assignment of credit problem. *Psychological review*, *126*(5), 761,

Gallistel, C.R., & King, A.P. (2011). *Memory and the computational brain: Why cognitive science will transform neuroscience* (Vol. 6). John Wiley &amp; Sons.

Gallistel, C.R., & Shahan, T.A. (2024). Time-scale invariant contingency yields one-shot reinforcement learning despite extremely long delays to reinforcement. *Proceedings of the National Academy of Sciences*, *121*(30), e2405451121,

Garcia-Garcia, M.G., Kapoor, A., Akinwale, O., Takemaru, L., Kim, T.H., Paton, C., ... Wagner, M.J. (2024). A cerebellar granule cell-climbing fiber computation to learn to track long time intervals. *Neuron*, ,

Gardner, M.P., Schoenbaum, G., Gershman, S.J. (2018). Rethinking dopamine as generalized prediction error. *Proceedings of the Royal Society B*, *285*(1891), 20181645,

Gauthier, J.L., & Tank, D.W. (2018). A dedicated population for reward coding in the hippocampus. *Neuron*, *99*, 179-193.e7,

Gavornik, J.P., & Bear, M.F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature neuroscience*, *17*(5), 732–737,

Gayler, R.W. (2004). Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. *arXiv preprint cs/0412059*, ,

Gershman, S.J. (2017). Dopamine, inference, and uncertainty. *Neural Computation*, *29*(12), 3311–3326,

Gershman, S.J. (2022). The molecular memory code and synaptic plasticity: a synthesis. *arXiv preprint arXiv:2209.04923*, ,

Gershman, S.J., Moore, C.D., Todd, M.T., Norman, K.A., Sederberg, P.B. (2012). The successor representation and temporal context. *Neural Computation*, *24*(6), 1553–1568,

Goh, W.Z., Ursekar, V., Howard, M.W. (2022). Predicting the future with a scale-invariant temporal memory for the past. *Neural Computation*, *34*(642-685), ,

Gosmann, J. (2018). *An integrated model of contex, short-term, and long-term memory* (Unpublished doctoral dissertation). University of Waterloo.

Guo, C., Huson, V., Macosko, E.Z., Regehr, W.G. (2021). Graded heterogeneity of metabotropic signaling underlies a continuum of cell-intrinsic temporal responses in unipolar brush cells. *Nature Communications*, *12*(1), 1–12,

Hamid, A.A., Frank, M.J., Moore, C.I. (2019). Dopamine waves as a mechanism for spatiotemporal credit assignment. *bioRxiv*, 729640,

Hamid, A.A., Frank, M.J., Moore, C.I. (2021). Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell*, *184*(10), 2733–2749,

Hasselmo, M.E., Bodelón, C., Wyble, B.P. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation*, *14*, 793-817,

Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N., Komiyama, T. (2019). Area-specificity and plasticity of history-dependent value coding during learning. *Cell*, *177*(7), 1858–1872,

Hattori, R., & Komiyama, T. (2022). Context-dependent persistency as a coding mechanism for robust and widely distributed value coding. *Neuron*, *110*(3), 502–515,

Henke, J., Bunk, D., von Werder, D., Häusler, S., Flanagin, V.L., Thurley, K. (2021). Distributed coding of duration in rodent prefrontal cortex during time

reproduction. *Elife*, *10*, e71612,

Hikosaka, K., & Watanabe, M. (2000). Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cerebral cortex*, *10*(3), 263–271,

Homann, J., Koay, S.A., Chen, K.S., Tank, D.W., Berry, M.J. (2022). Novel stimuli evoke excess activity in the mouse primary visual cortex. *Proceedings of the National Academy of Sciences*, *119*(5), e2108882119,

Howard, M.W., & Hasselmo, M.E. (2020). Cognitive computation using neural representations of time and space in the Laplace domain. *arXiv preprint arXiv:2003.11668*, ,

Howard, M.W., Luzardo, A., Tiganj, Z. (2018). Evidence accumulation in a Laplace decision space. *Computational Brain and Behavior*, *1*, 237-251,

Howard, M.W., MacDonald, C.J., Tiganj, Z., Shankar, K.H., Du, Q., Hasselmo, M.E., Eichenbaum, H. (2014). A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *Journal of Neuroscience*, *34*(13), 4692-707, https://doi.org/10.1523/JNEUROSCI.5808-12.2014

Howard, M.W., & Shankar, K.H. (2018). Neural scaling laws for an uncertain world. *Psychologial Review*, *125*, 47-58, https://doi.org/10.1037/rev0000081

Howard, M.W., Shankar, K.H., Aue, W., Criss, A.H. (2015). A distributed representation of internal time. *Psychological Review*, *122*(1), 24-53,

Howe, M.W., Tierney, P.L., Sandberg, S.G., Phillips, P.E., Graybiel, A.M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature*, *500*(7464), 575–579,

Husserl, E. (1966). *The phenomenology of internal time-consciousness*. Bloomington, IN: Indiana University Press.

Inagaki, H.K., Inagaki, M., Romani, S., Svoboda, K. (2018). Low-dimensional and monotonic preparatory activity in mouse anterior lateral motor cortex. *Journal of Neuroscience*, *38*(17), 4163–4185,

Jacques, B.G., Tiganj, Z., Howard, M.W., Sederberg, P.B. (2021). DeepSITH: Efficient learning via decomposition of what and when across time scales. M. Ranzato, A. Beygelzimer, P. Liang, J. Vaughan, & Y. Dauphin (Eds.), *35th conference on advances in neural information processing systems* (Vol. arXiv:2104.04646).

Jacques, B.G., Tiganj, Z., Sarkar, A., Howard, M., Sederberg, P. (2022). A deep convolutional neural network that is invariant to time rescaling. *International conference on machine learning* (pp. 9729–9738).

Jansson, Y., & Lindeberg, T. (2021). Scale-invariant scale-channel networks: Deep networks that generalise to previously unseen scales. *arXiv preprint arXiv:2106.06418*, ,

Jeong, H., Taylor, A., Floeder, J.R., Lohmann, M., Mihalas, S., Wu, B., . . . Namboodiri, V.M.K. (2022). Mesolimbic dopamine release conveys causal associations. *Science*, eabq6740,

Jin, D.Z., Fujii, N., Graybiel, A.M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proceedings of the National Academy of Sciences*, *106*(45), 19156–19161,

Jones, M.N., & Mewhort, D.J.K. (2007). Representing word meaning and order information composite holographic lexicon. *Psychological Review*, *114*, 1-32,

Kanter, B.R., Lykken, C.M., Moser, M.-B., Moser, E.I. (2024). Event structure sculpts neural population dynamics in the lateral entorhinal cortex. *bioRxiv*, 2024–06,

Kato, K., & Caplan, J.B. (2017). The brain's representations may be compatible with convolution-based memory models. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(4), 299,

Ke, N.R., Goyal, A., Bilaniuk, O., Binas, J., Mozer, M.C., Pal, C., Bengio, Y. (2018). Sparse attentive backtracking: Temporal credit assignment through reminding. *Advances in neural information processing systems* (pp. 7640–7651).

Khona, M., & Fiete, I.R. (2021). Attractor and integrator networks in the brain. *arXiv preprint arXiv:2112.03978*, ,

Kim, H., Homann, J., Tank, D.W., Berry, M.J. (2019). A long timescale stimulus history effect in the primary visual cortex. *bioRxiv*, 585539,

Kim, H.R., Malik, A.N., Mikhael, J.G., Bech, P., Tsutsui-Kimura, I., Sun, F., ... others (2020). A unified framework for dopamine signals across timescales. *Cell*, *183*(6), 1600–1616,

Koay, S.A., Charles, A.S., Thiberge, S.Y., Brody, C.D., Tank, D.W. (2022). Sequential and efficient neural-population coding of complex task information. *Neuron*, *110*(2), 328–349,

Komura, Y., Tamura, R., Uwano, T., Nishijo, H., Kaga, K., Ono, T. (2001). Retrospective and prospective coding for predicted reward in the sensory thalamus. *Nature*, *412*(6846), 546-9, https://doi.org/10.1038/35087595

Kurth-Nelson, Z., & Redish, A.D. (2009). Temporal-difference reinforcement learning with distributed representations. *PLoS One*, *4*(10), e7362,

Lee, R.S., Engelhard, B., Witten, I.B., Daw, N.D. (2022). A vector reward prediction error model explains dopaminergic heterogeneity. *bioRxiv*, 2022–02,

Lerner, Y., Honey, C.J., Katkov, M., Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of neurophysiology*, *111*(12), 2433–2444,

Lindeberg, T., & Fagerström, D. (1996). Scale-space with casual time direction. *European conference on computer vision* (pp. 229–240).

Lubenov, E.V., & Siapas, A.G. (2009). Hippocampal theta oscillations are travelling waves. *Nature*, *459*(7246), 534-9,

MacDonald, C.J., Lepage, K.Q., Eden, U.T., Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737-749,

Mainen, Z.F., & Kepecs, A. (2009). Neural representation of behavioral outcomes in the orbitofrontal cortex. *Curr Opin Neurobiol*, *19*(1), 84-91, https://doi.org/10.1016/j.conb.2009.03.010

Marcus, G.F. (2018). *The algebraic mind: Integrating connectionism and cognitive science*. MIT press.

Masset, P., Malik, A.N., Kim, H.R., Bech Vilaseca, P., Uchida, N. (2022). A distributional code for learning across timescales in dopamine-based reinforcement learning. *Society for neuroscience abstracts* (Vol. 234.27).

Masset, P., Tano, P., Kim, H.R., Malik, A.N., Pouget, A., Uchida, N. (2023). Multitimescale reinforcement learning in the brain. *bioRxiv*, 2023–11,

Mello, G.B., Soares, S., Paton, J.J. (2015). A scalable population code for time in the striatum. *Current Biology*, *25*(9), 1113–1122,

Momennejad, I., & Howard, M.W. (2018). Predicting the future with multi-scale successor representations. *bioRxiv*, 449470,

Momennejad, I., Russek, E.M., Cheong, J.H., Botvinick, M.M., Daw, N., Gershman, S.J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680,

Morcos, A.S., & Harvey, C.D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. *Nature Neuroscience*, *19*(12), 1672–1681,

Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609-626,

Namboodiri, V.M.K. (2021). What is the state space of the world for real animals? *BioRxiv*, ,

Namboodiri, V.M.K., Otis, J.M., van Heeswijk, K., Voets, E.S., Alghorazi, R.A., Rodriguez-Romaguera, J., ... Stuber, G.D. (2019). Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. *Nature neuroscience*, *22*(7), 1110–1121,

Namboodiri, V.M.K., & Stuber, G.D. (2021). The learning of prospective and retrospective cognitive maps within neural circuits. *Neuron*, *109*(22), 3552–3575,

Ning, W., Bladon, J.H., Hasselmo, M.E. (2022). Complementary representations of time in the prefrontal cortex and hippocampus. *Hippocampus*, ,

Ogata, K. (1970). Describing-function analysis of nonlinear control systems. *Modern Control Engineering*, *2*, 645–676,

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.

Palmer, S.E., Marre, O., Berry, M.J., 2nd, Bialek, W. (2015). Predictive information in a sensory population. *Proceedings of the National Academy of Sciences USA*, *112*(22), 6908-13, https://doi.org/10.1073/pnas.1506855112

Pastalkova, E., Itskov, V., Amarasingham, A., Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322-7,

Patel, J., Fujisawa, S., Berényi, A., Royer, S., Buzsáki, G. (2012). Traveling theta waves along the entire septotemporal axis of the hippocampus. *Neuron*, *75*(3), 410-7, https://doi.org/10.1016/j.neuron.2012.07.015

Phillips, J.W., Schulmann, A., Hara, E., Winnubst, J., Liu, C., Valakh, V., ... others (2019). A repeated molecular architecture across thalamic pathways. *Nature neuroscience*, *22*(11), 1925–1935,

Piantadosi, S.T. (2016). A rational analysis of the approximate number system. *Psychonomic Bulletin & Review*, 1–10,

Plate, T.A. (1995). Holographic reduced representations. *IEEE Transactions on Neural networks*, *6*(3), 623–641,

Rainer, G., Rao, S.C., Miller, E.K. (1999). Prospective coding for objects in primate prefrontal cortex. *Journal of Neuroscience*, *19*(13), 5493–5505,

Rao, R.P., & Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79-87, https://doi.org/10.1038/4580

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the efectivenesss of reinforcement and nonreinforcement. A.H. Black & W.F. Prokasy (Eds.), *Classical conditioning II: Current research and theory.* New York: Appleton-Century-Crofts.

Roitman, J.D., & Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience*, *22*(21), 9475–9489,

Roy, D.S., Zhang, Y., Halassa, M.M., Feng, G. (2022). Thalamic subnetworks as units of function. *Nature Neuroscience*, *25*(2), 140–153,

Russ, B.E., Koyano, K.W., Day-Cooney, J., Perwez, N., Leopold, D.A. (2022). Temporal continuity shapes visual responses of macaque face patch neurons. *bioRxiv*, ,

Salet, J.M., Kruijne, W., van Rijn, H., Los, S.A., Meeter, M. (2022). FMTP: A unifying computational framework of temporal preparation across time scales. *Psychological Review*, ,

Sarel, A., Finkelstein, A., Las, L., Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, *355*(6321), 176–180,

Schlegel, K., Neubert, P., Protzel, P. (2022). A comparison of vector symbolic architectures. *Artificial Intelligence Review*, *55*(6), 4523–4555,

Schoenbaum, G., Chiba, A.A., Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature neuroscience*, *1*(2), 155–159,

Schoenbaum, G., & Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron*, *47*(5), 633-6, https://doi.org/10.1016/j.neuron.2005.07.018

Schonhaut, D.R., Aghajan, Z.M., Kahana, M.J., Fried, I. (2022). A neural code for spatiotemporal context. *bioRxiv*, ,

Schonhaut, D.R., Aghajan, Z.M., Kahana, M.J., Fried, I. (2023). A neural code for time and space in the human brain. *Cell Reports*, *42*(11), ,

Schultz, W., Dayan, P., Montague, P.R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593-1599,

Scott, B.B., Constantinople, C.M., Akrami, A., Hanks, T.D., Brody, C.D., Tank, D.W. (2017). Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. *Neuron*, *95*(2), 385–398,

Shahbaba, B., Li, L., Agostinelli, F., Saraf, M., Cooper, K.W., Haghverdian, D., . . . Fortin, N.J. (2022). Hippocampal ensembles represent sequential relationships among an extended sequence of nonspatial events. *Nature communications*, *13*(1), 1–17,

Shankar, K.H., & Howard, M.W. (2013). Optimally fuzzy temporal memory. *Journal of Machine Learning Research*, *14*, 3753-3780,

Shankar, K.H., Singh, I., Howard, M.W. (2016). Neural mechanism to simulate a scale-invariant future. *Neural Computation*, *28*, 2594–2627,

Shikano, Y., Ikegaya, Y., Sasaki, T. (2021). Minute-encoding neurons in hippocampal-striatal circuits. *Current Biology*, *31*(7), 1438–1449,

Solstad, T., Boccara, C.N., Kropff, E., Moser, M.B., Moser, E.I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*(5909), 1865-8,

Sousa, M., Bujalski, P., Cruz, B.F., Louie, K., McNamee, D., Paton, J.J. (2023). Dopamine neurons encode a multidimensional probabilistic map of future reward. *bioRxiv*, 2023–11,

Spitmaan, M., Seo, H., Lee, D., Soltani, A. (2020). Multiple timescales of neural dynamics and integration of task-relevant signals across cortex. *Proceedings of the National Academy of Sciences*, *117*(36), 22522–22531,

Stachenfeld, K.L., Botvinick, M.M., Gershman, S.J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643-1653, https://doi.org/

10.1038/nn.4650

Steinberg, J., & Sompolinsky, H. (2022). Associative memory of structured knowledge. *Scientific Reports*, *12*(1), 21808,

Subramanian, D.L., & Smith, D.M. (2024). Time cells in the retrosplenial cortex. *bioRxiv*, ,

Sutton, R.S., & Barto, A.G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*, 135-171,

Tano, P., Dayan, P., Pouget, A. (2020). A local temporal difference code for distributional reinforcement learning. *Advances in Neural Information Processing Systems*, *33*, 13662–13673,

Taxidis, J., Pnevmatikakis, E.A., Dorian, C.C., Mylavarapu, A.L., Arora, J.S., Samadian, K.D., . . . Golshani, P. (2020). Differential emergence and stability of sensory and temporal representations in context-specific hippocampal sequences. *Neuron*, *108*(5), 984–998.e9,

Tiganj, Z., Cromer, J.A., Roy, J.E., Miller, E.K., Howard, M.W. (2018). Compressed timeline of recent experience in monkey lPFC. *Journal of Cognitive Neuroscience*, *30*, 935-950,

Tiganj, Z., Gershman, S.J., Sederberg, P.B., Howard, M.W. (2019). Estimating scale-invariant future in continuous time. *Neural Computation*, *31*(4), 681–709,

Tiganj, Z., Kim, J., Jung, M.W., Howard, M.W. (2017). Sequential firing codes for time in rodent mPFC. *Cerebral Cortex*, *27*, 5663-5671,

Tiganj, Z., Singh, I., Esfahani, Z.G., Howard, M.W. (2022). Scanning a compressed ordered representation of the future. *Journal of Experimental Psychology: General*, *151*(3082–3096), ,

Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189-208,

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J.J., Moser, M.-B., Moser, E.I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, *561*, 57-62,

van der Meer, M.A.A., & Redish, A.D. (2011). Theta phase precession in rat ventral striatum links place and reward information. *Journal of Neuroscience*, *31*(8), 2843-54, https://doi.org/10.1523/JNEUROSCI.4869-10.2011

Waelti, P., Dickinson, A., Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43-8,

Wagner, M.J., Kim, T.H., Savall, J., Schnitzer, M.J., Luo, L. (2017). Cerebellar granule cells encode the expectation of reward. *Nature*, *544*(7648), 96–100,

Wagner, M.J., & Luo, L. (2020). Neocortex–cerebellum circuits for cognitive processing. *Trends in neurosciences*, *43*(1), 42–54,

Watanabe, K., Kadohisa, M., Kusunoki, M., Buckley, M.J., Duncan, J. (2023). Cycles of goal silencing and reactivation underlie complex problem-solving in primate frontal and parietal cortex. *Nature Communications*, *14*(1), 5054,

Wei, W., Mohebi, A., Berke, J.D. (2021). A spectrum of time horizons for dopamine signals. *bioRxiv*, 2021–10,

Wei, X.-X., & Stocker, A.A. (2012). Efficient coding provides a direct link between prior and likelihood in perceptual bayesian inference. *Advances in neural information processing systems*, *25*, ,

Whittington, J.C., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N., Behrens, T.E. (2020). The tolman-eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell*, *183*(5), 1249–1263,

Wilson, M.A., & McNaughton, B.L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, *261*, 1055-8,

Young, J.J., & Shapiro, M.L. (2011). The orbitofrontal cortex and response selection. *Ann N Y Acad Sci*, *1239*, 25-32, https://doi.org/10.1111/j.1749-6632.2011.06279.x

Yu, Q., Bi, Z., Jiang, S., Yan, B., Chen, H., Wang, Y., ... Zhang, J. (2022). Visual cortex encodes timing information in humans and mice. *Neuron*, *110*, 4194-4211,

Zhang, H., & Jacobs, J. (2015). Traveling theta waves in the human hippocampus. *Journal of Neuroscience*, *35*(36), 12477-87, https://doi.org/10.1523/JNEUROSCI.5102-14.2015

Zuo, S., Wang, C., Wang, L., Jin, Z., Kusunoki, M., Kwok, S.C. (2023). Neural signatures for temporal-order memory in the medial posterior parietal cortex. *bioRxiv*, 2023–08,