

Reviewed Preprint

v1 • March 5, 2026

Not revised

✉ For correspondence:

marc777@bu.edu

Reviewing editor: Tatyana O

Sharpee, Salk Institute for Biological Studies, United States

© 2026, Sarkar et al. This article is distributed under the terms of the

[Creative Commons Attribution](#)

[License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

“What” × “When” working memory representations using Laplace Neural Manifolds

Aakash Sarkar¹, Chenyu Wang¹, Shangfu Zuo², Marc W Howard^{1,3} ✉

¹Department of Psychological and Brain Sciences, Boston University, Boston, United States • ²Department of Psychology, Rutgers University New Brunswick, United States • ³Department of Physics Boston, University Boston, United States

eLife Assessment

This manuscript provides **useful** insights into how the brain can simultaneously represent events and the times when they occurred. The results include a comparison between two different basis functions for temporal selectivity and how these generate different predictions for the dynamics of neural populations. The conclusions are partly **incomplete** because of questions such as the impact of the linear separability assumption and whether joint encodings of event type and time can be made without it.

<https://doi.org/10.7554/eLife.108804.1.sa1>

Abstract

Working memory—the ability to remember recent events as they recede continuously into the past—requires the ability to represent any stimulus at any time delay. This property requires neurons coding working memory to show mixed selectivity, with conjunctive receptive fields (RFs) for stimuli and time, forming a representation of ‘what’ × ‘when’ for the recent past. We study the properties of such a working memory in experiments where a single stimulus must be remembered for a short time. Conjunctive receptive fields allows the covariance matrix of the network to decouple neatly, allowing an understanding of the low-dimensional dynamics of the population. We study a specific choice—a Laplace space with exponential basis functions for time coupled to an “Inverse Laplace” space with circumscribed basis functions in time. We refer to this choice with basis functions that evenly tile log time as a Laplace Neural Manifold for time. Despite being related by a linear projection, the Laplace population shows a stable stimulus-specific subspace whereas the Inverse Laplace population shows rotational dynamics. The rank of the covariance matrix with time grows logarithmically in good agreement with data. We sketch a continuous attractor network that constructs a Laplace Neural Manifold for time. The attractor in the Laplace space appears as an edge; the attractor for the inverse space as a bump. This work provides a bridge between abstract cognitive models of WM and circuit-level continuous attractor neural networks.

Introduction

Consider a simple experiment where a single stimulus is presented for a brief moment followed by an unfilled delay interval. With the passage of time, the memory for the event is preserved. According to philosophers (James, 1890 [\[1\]](#); Husserl, 1966 [\[2\]](#); Bergson, 1910 [\[3\]](#)), the identity of the stimulus in memory is unchanged, but the memory takes on a new character with the passage of time. In the words of Husserl (1966) [\[2\]](#), with the passage of time, “points of temporal duration recede, as points of a stationary object in space recede when I ‘go away from the object’.”

Experimental data from cognitive psychology is consistent with this introspection; participants can *separately* judge the occurrence and relative time of different stimuli (Hacker, 1980 [↗](#); Hintzman, 2010 [↗](#)).

The great mathematical physicist Hermann Weyl introduced the argument above (with analogous requirements for space) as part of an axiomatic derivation of general relativity. Viewed in the light of contemporary discussions about artificial intelligence, we might say that Weyl required a compositional representation of empirical content and time. Given a way to describe empirical content—a what—it must be possible to describe every possible what at every possible when, *and* vice versa.

The way the early visual system conjunctively codes for what and *where* information provides a template for how the brain could code for what and *when* information. Through-out the early visual system there are neurons with similar sensitivity to patterns of light, but with receptive fields in different locations over retinal coordinates. For instance, simple cells respond best to patterns of light with a particular orientation—a what—at a particular location—a where. The pattern of activity of simple cells over all spatial locations gives a conjunctive representation of what is where in the visual field.

Like position in the visual field, physical time is also an ordered continuous dimension. In much the same we might notice that object A is located to the left of object B, we can also remember if event A happened before or after event B. This paper pursues the implications of assuming a conjunctive representation that places events a what on an internal timeline—a when—as a fundamental property of working memory for the recent past.

We show that neuronal populations with receptive fields of neurons as a a product of a stimulus term and a temporal term satisfies the requirement of compositionality of what and when (Machens, Romo, & Brody, 2010 [↗](#)). Such conjunctive codes of what × when make it straightforward to write out covariance matrices in simple experiments, enabling a description of population dynamics in low dimensional spaces like those widely used in neuroscience research. We then specify a particular choice for temporal basis functions inspired by work in theoretical neuroscience (Shankar & Howard, 2013 [↗](#)), cognitive psychology (Howard, Shankar, Aue, & Criss, 2015 [↗](#)) and deep networks (Jacques, Tiganj, Sarkar, Howard, & Sederberg, 2022 [↗](#)). When projected onto a linear space, these *Laplace Neural Manifolds* for time generate predictions that resemble empirical results from monkey cortex (Murray et al., 2017 [↗](#); Cueva et al., 2020 [↗](#)). Finally, to illustrate that these equations can be instantiated in biological networks, we sketch a continuous attractor neural network model that conjunctively codes for what × when, with temporal basis functions chosen as a Laplace Neural Manifold.

Theoretical Considerations

This paper looks into compositional representations and their implications and implementations in neuroscience. We show that a compositional working memory, coding for what happened when, requires that neurons must have conjunctive receptive fields, decomposable as functions of stimulus and time. Placing minimal restraints on the form of the receptive fields, we show that we can further write the population co-variance as a tensor product of the stimulus (Σ_{what}) and time (Σ_{when}) covariances, respectively.

We simulate these populations for simple working memory tasks, where a task variable, distributed either on a ring or on a line, must be remembered for a short amount of time. Depending on our choice of temporal basis set, we observe, through linear dimensionality reduction techniques like PCA, neurally observed traits like stable subspaces and rotational dynamics.

Finally, since the covariance matrix of the population decomposes into Σ_{what} and Σ_{when} - and the dimensionality of Σ_{what} is completely fixed by the task and choice of stimulus RFs, we observe that the subspace spanned by Σ_{when} controls the dimensionality of the covariance matrix. Measuring the dimensionality of neural trajectories as a function of time should reveal the density of basis

functions over the continuous dimension of time. As a natural consequence of smooth basis functions, we see that the dimensionality of Σ_{when} can grow without bound as a function of the elapsed time, but the growth can decelerate.

Conjunctive receptive fields can support a compositional working memory

Let us assume that we prepare an experiment in which one of several stimuli x, y, z is presented at $t = 0$ for many trials. We counterbalance the number of presentations of each stimulus and perform all other experimental controls. We record the firing rate over a population of neurons $\mathbf{m}$ that reflects the state of a memory with finite duration (Maass, Natschläger, & Markram, 2002). We choose the time between trials to be long enough that we can ignore any carryover from previous trials. Let us describe the population vector expected at a time t after presentation of a particular stimulus $\hat{x}$ as $\mathbf{m}(\hat{x}, t)$.

We operationalize Weyl's requirement that the empirical content of the memory—the what—be unchanged with the passage of time by requiring that x be linearly decodable without knowing t . It is acceptable that the *accuracy* of decoding changes with the passage of time. However, in order for the “empirical content” to remain fixed we require that the relationships between all pairs of stimuli in the decoding space are preserved at all time points (Fig. 1b).

Just like the relationships in the what representations should stay independent of the duration when they were experienced, the when representations should also be independent of the identity of the stimulus. The relationships between pairs of time intervals should thus be preserved independent of the stimuli used to mark those intervals (Fig. 1c). If three stimuli are presented in succession (x, y, z) such that the second one follows the first one by a delay of τ_1 and the third one follows the second one by a delay of τ_2 , the time interval between the first and third stimuli (τ_3) is just the sum of the first two delays ($\tau_3 = \tau_1 + \tau_2$). This relationship between the intervals would not change even if the order of the stimuli themselves are shuffled (y, z, x).

This requirement leads to the conclusion that neurons in $\mathbf{m}(\hat{x}, t)$ have conjunctive receptive fields as a function of what $\times$ when. To see this, let us refer to the column of a linear decoder pointing in the direction of x as $\mathbf{x}$ and define $m_x(\hat{x}, t) = \mathbf{x}'\mathbf{m}(\hat{x}, t)$ and $m_y(\hat{x}, t) = \mathbf{y}'\mathbf{m}(\hat{x}, t)$. It is acceptable that the magnitude of $m_x(\hat{x}, t)$ and $m_y(\hat{x}, t)$ change as a function of time as information recedes into the past. However for the “empirical content” to be the same, we require that the relationships between all stimuli be constant. This requires that $\frac{m_x(\hat{x}, t)}{m_y(\hat{x}, t)}$ be constant as a function of time for all pairs of stimuli x, y and all stimuli that could be presented $\hat{x}$. For each cell in $\mathbf{m}$ projecting into $\mathbf{x}$ with a particular time course, there must be another cell projecting into $\mathbf{y}$ with the same time course, up to multiplication by a constant.

We can thus subscript the cells in the population vector $\mathbf{m}$ with two indices i and j

$$m_{ij}(\hat{x}, t) = a_{ij} g_i(\hat{x}) h_j(t) \quad (1)$$

where a_{ij} is just a normalization factor that we will set to 1 for simplicity. Note that this decomposition of $\mathbf{m}(\hat{x}, t)$ also satisfies the other constraint to the effect that we can also linearly decode the when, t , without knowing the what, $\hat{x}$.

Conjunctive, mixed selective receptive fields can create a compositional representation of what and when. We can understand $g_i(\hat{x})$ as a set of receptive fields over whatever dimensions span the stimulus space and $h_j(t)$ as a set of temporal receptive fields.

This requirement would not be compatible with many forms of memory, some of which are widely used in neuroscience. For instance, suppose that $\mathbf{m}(\hat{x}, t)$ were maintained as a function of time with an RNN initialized as $\mathbf{m}(\hat{x}, t = 0) = \mathbf{I}(\hat{x})$ and that then evolved as $\dot{\mathbf{m}} = \mathbf{R}\mathbf{m}$. Being able to decompose $\mathbf{m}(\hat{x}, t)$ as Eq. 1 requires that $\mathbf{R}$ be decomposable as

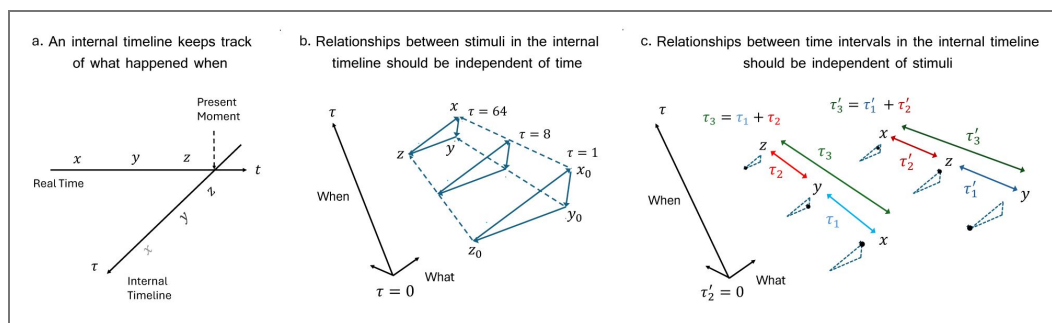


Figure 1. An internal timeline where stimulus ('What') representations change independently of temporal representations ('When'), and vice versa.

Left: An internal timeline creating a record of the past. We hypothesize that the brain forms a conjunctive representation that places events — a “what” — on an internal timeline — a “when”. Such a conjunctive representation of what happened when should have properties, shown in the middle and right. *Middle: The ‘What’ representations should behave independently from the elapsed time, or ‘When’.* Consider we saw a set of stimuli x , y , and z , at a delay of either $\tau = 1, 8$ or 64 seconds. The representation vectors for stimuli x , y and z in our internal timeline might shrink or expand as we move into the past (τ gets bigger), but the relationships between them, here represented as the angle between each pair of vectors, should remain the same for all three durations (τ), so that their meaning ('What') remains independent of the elapsed time ('When'). *Right: The ‘When’ representations should behave independently of the stimuli which are experienced.* We can represent different durations no matter what stimuli are used to define those intervals. For example, if three stimuli are presented in succession (x, y, z) such that the second one follows the first one by a delay of τ_1 and the third one follows the second one by a delay of τ_2 , we are aware that the time interval between the first and third stimuli (τ_3) is just the sum of the first two delays ($\tau_3 = \tau_1 + \tau_2$). This would hold true even if the stimuli ('What') were presented in a different order (y, z, x) - we might denote the time intervals differently (τ'_1, τ'_2 , and τ'_3), but the relationship between them ('When') would be unchanged ($\tau'_3 = \tau'_1 + \tau'_2$).

$$R = M_{\text{what}} \otimes M_{\text{when}} \quad (2)$$

and that $\mathbf{I}(x)$ provides the same input to each part of the space spanned by $\mathbf{M}_{\text{when}}$ for each stimulus. This is a very strong constraint on recurrent connectivity that reflects a specific structural inductive bias. This choice is not typically made in computational neuroscience network models.

Total covariance Σ decouples into Σ_{what} and Σ_{when}

To understand the behavior of neural data, linear dimensional reduction techniques is often used to visualize population vectors as neural trajectories in time. This requires knowledge of the population covariance matrix of the neural data. If the activity of populations of neurons can be decomposed into products of what $\times$ when receptive fields, as in Eq. 1, then this results in straightforward expressions for the population covariance matrices.

Lets start with a compositional representation of what happened when, with neurons obeying

$$\Phi_{ij}(\hat{x}, \tau) = g(\hat{x}, x_i) h(\tau, \tau_j) \quad (3)$$

where the activity Φ of the neuron (indexed by a *what* index i and a *when* index j) describes the receptive field when stimulus $\hat{x}$ is a time τ in the past. If a stimulus is presented at time $t = 0$, then at time t , the stimulus is $\tau = t$ seconds in the past. The function g describes the neuron's tuning curve over the stimulus dimension, with the preferred stimulus x_i , and the function h describes the temporal receptive field of the neuron, with a time sensitivity parameter τ_j . The stimulus and time preferences can be indexed separately, since the choice of one has no bearing on the other.

Let us calculate the general covariance matrix for populations that obey (3) for some generic set of stimuli indexed by x . We are not restrained to any specific choice of receptive fields for g and h , but only make the assumption that they are normalized over the stimulus and time space respectively

$$\frac{1}{T} \int_0^T h(\tau_j, \tau) d\tau = 1, \quad \frac{1}{N} \sum_x g(\hat{x}, x_i) = 1. \quad (4)$$

This allows us to write the expectation of the activity over time as a pure function of stimulus, $\langle \Phi_{ij} \rangle_t(\hat{x}) \equiv E_t[\Phi_{ij}](\hat{x}) = g(\hat{x}, x_i)$, and the expectation of the activity over stimulus as a pure function of experimental time $\langle \Phi_{ij} \rangle_{\hat{x}}(t) \equiv E_{\hat{x}}[\Phi_{ij}](t) = h(\tau_j, \tau = t)$.

For Σ_{when} , the covariance over time of $\langle \Phi_{ij} \rangle_x(t)$ and $\langle \Phi_{kl} \rangle_x(t)$, we get

$$[\Sigma_{\text{when}}]_{ijkl} = \frac{1}{T} \int_0^T h(\tau, \tau_j) h(\tau, \tau_l) d\tau - 1. \quad (5)$$

Note that the first term can be written as $1_{ik} H_{jl}$, where $\mathbf{H}$ is a symmetric matrix which encodes the expectation (over time) of the product of $h(\tau, \tau_j)$ and $h(\tau, \tau_l)$.

For Σ_{what} , the covariance over stimulus of $\langle \Phi_{ij} \rangle_t(\hat{x})$ and $\langle \Phi_{kl} \rangle_t(\hat{x})$, we get

$$[\Sigma_{\text{what}}]_{ijkl} = \frac{1}{N} \sum_x g(\hat{x}, x_i) g(\hat{x}, x_k) - 1. \quad (6)$$

Note that the first term can be written as $G_{ik} 1_{jl}$ where $\mathbf{G}$ is a symmetric matrix which encodes the expectation (over stimuli) of the product of $g(\hat{x}, x_i)$ and $g(\hat{x}, x_k)$.

The overall covariance Σ_{ijkl} is calculated over both stimuli and time and can be written as

$$\Sigma_{ijkl} + 1 = \left(\frac{1}{N} \sum_x g(\hat{x}, x_i) g(\hat{x}, x_k) \right) \times \left(\frac{1}{T} \int_0^T h(\tau, \tau_j) h(\tau, \tau_l) d\tau \right)$$

where the second term on the left appears as the expectation of the activity over both stimulus and time and is unity since g and h are normalized appropriately. For the first term, the sum and integral can be separated since g and h are pure functions of stimulus and time. Recognizing that the first term is just a product of the matrix elements G_{ik} and H_{jl} , we can generalize this to

$$\Sigma = G \otimes H - 1. \quad (7)$$

We can thus write the overall covariance matrix (Σ) as a Kronecker product

$$\Sigma + 1 = (\Sigma_{what} + 1) \otimes (\Sigma_{when} + 1). \quad (8)$$

Thus, we see that the overall covariance matrix Σ (computed over both time and stimulus dimensions) neatly decomposes into the tensor product of the covariance matrices for stimulus (Σ_{what}) and time (Σ_{when}) respectively, due to the conjunctive coding designed into the Laplace neural manifolds.

Low-dimensional projections of What × When conjunctive representations

A comparison with experimental data requires us to specify the form of the receptive fields for what and when information. For concreteness, we study reasonable choices of temporal receptive fields and examine working memory population dynamics during simple delay experiments that have been used to study cortical populations in monkeys. The choices of stimulus receptive fields are appropriate to the stimuli used in each of two tasks. The temporal receptive fields are chosen consistent with two broad classes - each choice forms a basis set over the time dimension, and they are provably related to one another by a linear operator. We will see that these two closely related choices of temporal basis functions lead to qualitatively different population dynamics as a function of time when projected onto a low-dimensional space.

Laplace Neural Manifolds: Two forms of temporal basis functions

We choose two forms of receptive fields, both of which form a basis set over the continuous dimension of time. One set of basis functions is chosen such that each cell fires in a circumscribed region of time. As a “what” recedes into the past, cells tiling the time axis fire sequentially, much like so-called time cells (Fig. 2f) that have been observed in the hippocampus and elsewhere (Pastalkova, Itskov, Amarasingham, & Buzsaki, 2008; Jin, Fujii, & Graybiel, 2009; Mac-Donald, Lepage, Eden, & Eichenbaum, 2011). Although cells that sequentially activate in time are widely observed in the brain (Tiganj et al., 2018; Akhlaghpour et al., 2016; Parker et al., 2022; Subramanian & Smith, 2024), these are not the only choice of basis functions.

In addition, we consider monotonic receptive fields in which individual neurons decay as a function of time, but with different time constants (Fig. 2e) (Tsao et al., 2018; Bright et al., 2020; Zuo et al., 2024; Cao et al., 2024; Atanas et al., 2023). Note that when projected into principal component space, decaying firing is indistinguishable from ramping firing.

Exponential temporal receptive fields for Laplace transform

For exponential receptive fields, we set the function h as

$$h_F(\tau, \tau_j) \propto e^{-\frac{\tau}{\tau_j}}, \quad (9)$$

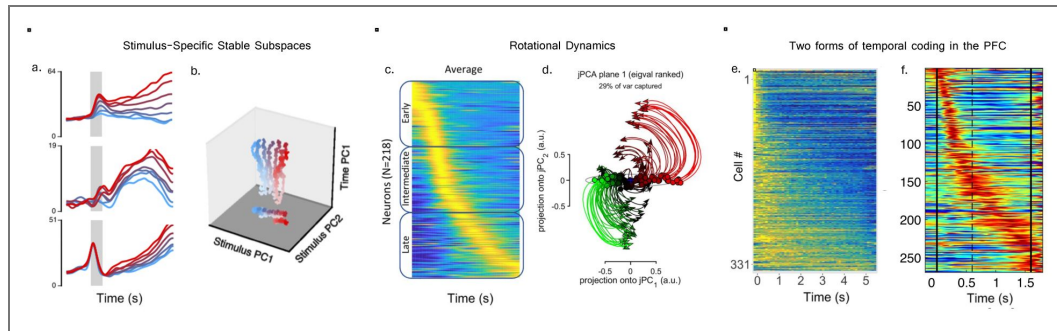


Figure 2. Low-dimensional dynamics in neural populations maintaining a stimulus representation show specific trends like stable subspaces and rotational dynamics, while timing information is maintained using two different temporal coding schemes.

Left: Stable subspaces exist simultaneously with diverse neural dynamics in Prefrontal Cortex (PFC) During Working Memory. While classical models posited stable, persistent firing for working memory, individual PFC neurons exhibit highly heterogeneous and dynamic activity during delay periods (a). However, population-level analysis can reveal stable representations. For instance, in a vibrotactile delayed discrimination task, where one has to keep track of a frequency, Principal Component Analysis (PCA) of PFC neuronal populations uncovers a stimulus-specific subspace (b). Neural trajectories within this subspace maintain a stable representation of the task variable in the shape of a line, with temporal variance relegated to an orthogonal subspace (perpendicular to the stimulus subspace), with different colored lines corresponding to different values of the stimulus frequency. Reproduced from Murray et al. (2017) [\[1\]](#). Middle: *Rotational Dynamics in Motor Cortex*. In contrast to the stable representations in working memory, studies on motor control have revealed different dynamics. Churchland et al. (2012) observed rotational dynamics in neuronal populations of the motor cortex during reaching tasks, where trajectories in an informative subspace show an angular rotation over time (d). Such oscillatory dynamics have been proposed as a fundamental form of neural processing. However, more recent work suggests these rotational dynamics could also be a signature of consistent sequential neuronal activity in the analyzed data (c). Reproduced from Lebedev et al. (2019) [\[2\]](#). Right: *Two complementary forms of temporal coding in the prefrontal cortex (PFC)*. Temporal context cells (e, reproduced from Cao et al., 2024 [\[3\]](#)) exhibit persistent activity throughout a delay period, with their firing rate modulated by the passage of time. This suggests they maintain a stable representation of the current temporal context. Time cells (f, reproduced from Tiganj et al., 2018 [\[4\]](#)) on the other hand, show transient bursts of activity at specific, successive moments within a trial, effectively tiling the delay period with their firing.

© 2017, PNAS. Panel b was reproduced from Murray et al., 2017 [\[1\]](#) with permission of PNAS. It is not covered by the CC-BY 4.0 licence and further reproduction of this panel would need permission from the copyright holder.

© 2024, Cao et al.. Panel e was reproduced from Cao et al., 2024 [\[3\]](#) (published under a CC BY-ND license [\[5\]](#)). Further reproductions must adhere to the terms of this license.

© 2018, Massachusetts Institute of Technology. Panel f was reproduced from Tiganj et al., 2018 [\[4\]](#) with permission of MIT Press. All rights reserved. It is not covered by the CC-BY 4.0 licence and further reproduction of this panel would need permission from the copyright holder.

where the constant of proportionality is set to ensure the normalization of h specified in Eq. 4. Comparison of these receptive fields with the definition of the Laplace transform shows that at time t this population is closely related to the Laplace transform with real coefficients, of a delta function a time τ in the past. This justifies the claim that h_F is a basis set if τ_j is effectively continuous. We refer to the population with temporal receptive fields $h(\tau_j, t) = e^{-t/\tau_j}$ as a Laplace population and label this population as F (Fig. 3c).

The inverse space: Circumscribed temporal receptive fields

To construct circumscribed temporal receptive fields, we use gamma functions (Tank & Hopfield, 1987; De Vries & Principe, 1992) parameterized by a variable k

$$h_{\tilde{f}}(\tau_j, \tau) = \left(\frac{\tau}{\tau_j} \right)^k e^{-k(\tau/\tau_j)}. \quad (10)$$

This receptive field is a product of a power law that grows with τ/τ_j and an exponential that decays with τ/τ_j . With this choice, $h(\tau, \tau_j)$ goes to zero as $\tau/\tau_j \rightarrow 0$ and also as $\tau/\tau_j \rightarrow \infty$. In between there is a single peak at a time that depends on the choice of τ_j . Choosing a new value of τ_j rescales the function. Thus a population with different values of τ_j will fire sequentially as a triggering stimulus enters the past (Fig. 3d).

It can be shown that the choice of temporal receptive fields in Eq. 10 leads to a deep relationship with neurons with receptive fields described by Eq. 9. Whereas a population of neurons with exponential receptive fields and a variety of time constants encode the Laplace transform with real coefficients, a population of neurons with receptive fields obeying Eq. 10 approximate the inverse Laplace transform, computed using the Post approximation with coefficient k (Shankar & Howard, 2012). This means that the two populations are related to one another via a linear transformation that can readily be computed (Shankar & Howard, 2013). For this reason, we refer to a population with receptive fields chosen as Eq. 10 as an Inverse Laplace Space and label the activity across the population as $\tilde{f}$.

Distributions of time constants for log time

Each Laplace neuron has a time constant τ_j that dictates how fast its activity decays, while each Inverse Laplace neuron has a time constant τ_j which dictates the position of its temporal receptive field. It remains to choose the distribution of time constants τ_j for the populations. Rather than a basis set uniformly tiling τ , we select time constants so that the basis functions tile $\log \tau$. This choice has theoretical advantages (Wei & Stocker, 2012; Piantadosi, 2016; Howard & Shankar, 2018) and is in agreement with some neural data (Cao, Bladon, Charczynski, Hasselmo, & Howard, 2022; Guo, Huson, Macosko, & Regehr, 2021).

The distribution of the time constants τ_j is instrumental in defining how the basis functions sample time. In this paper, the τ_j are distributed geometrically. The time constants for the temporal receptive fields are chosen such that the n th time constant is given by

$$\tau_n = (1 + c)^n \tau_0 \quad (11)$$

for some positive constant c . If we rearrange the terms a bit, we can express $n(\tau)$, the number of cells that tile the time axis from 0 to τ

$$n(\tau) = \log_{1+c} \frac{\tau}{\tau_0}. \quad (12)$$

The number of cells spanning an interval τ thus grows logarithmically with τ . This is also equivalent to saying that the time constants evenly tile log time

$$\log \tau_{n+1} - \log \tau_n = (1 + c). \quad (13)$$

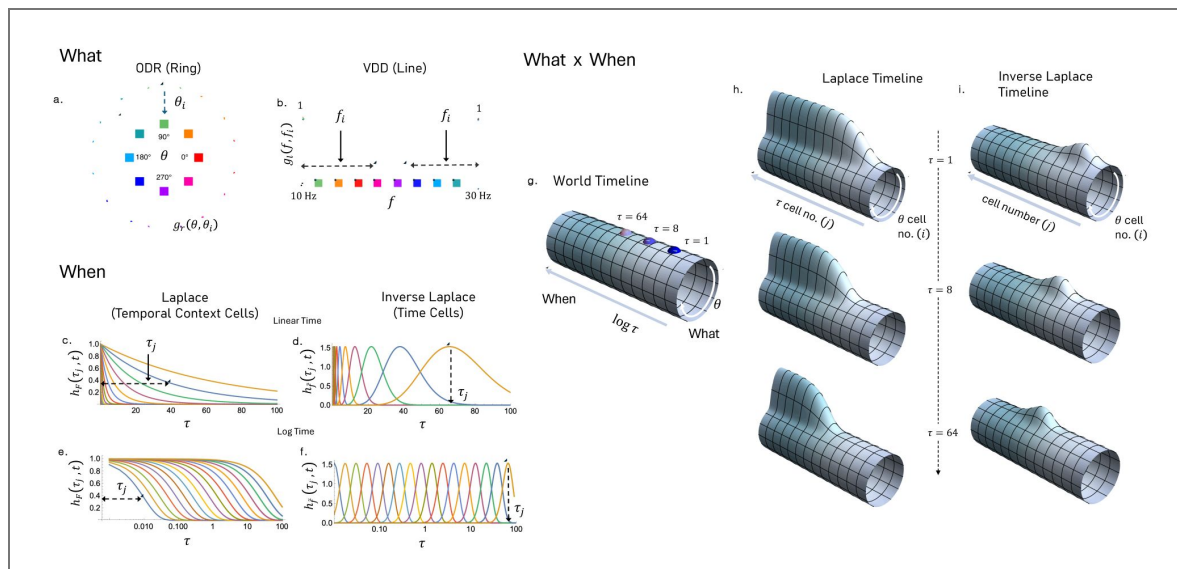


Figure 3. Conjunctive coding using Laplace Neural Manifolds.

LNM neurons have receptive fields (RFs) which are a product of a stimulus (*what*) and a temporal (*when*) term. *What*: The stimulus receptive fields tile the stimulus space, which is a ring in the case of ODR (Oculomotor Delayed Response, where an angular location θ has to be remembered), and a line in the case of VDD (Vibrotactile Delayed Discrimination task, where a frequency f has to be remembered). For the ODR task, neurons have bell-shaped tuning curves approximating a circular analogue of a normal distribution, preferentially encoding stimuli at different angles θ_i which evenly tile the ring (a). For the VDD task, the frequency range from 10 to 30 Hz is tiled by exponentially decaying and ramping cells with a spectrum of different ramp rates $1/f_j$. (b). *When*: Laplace cells (F) decay exponentially with different decay rates $1/\tau_j$, after the stimuli is presented, resembling temporal context cells (c), while inverse Laplace ($\tilde{f}$) cells fire sequentially, peaking at different times τ_j , resembling time cells (d). The receptive fields are chosen to evenly tile log time (d, e). When a stimuli is presented at a certain angle $\hat{\theta}$, the activity of the population as the time T after the presentation can be modeled on a cylinder (e). Laplace neurons (modeling temporal context cells) encode this history as a moving edge when seen in log time (f), while the inverse Laplace neurons encode it as a bump (g).

(Fig. 3b,c) shows the activity of cells with two kinds of temporal receptive fields firing as a function of experimental time, and as a function of log time (Fig. 3d,e). We will refer to these paired sets of basis functions with time constants chosen as in Eq. 11 as a *Laplace Neural Manifold* (Daniels & Howard, 2025; Howard, Esfahani, Le, & Sederberg, 2024).

‘What’ Representations

The experimental data reported in Murray et al. (2017) used simple memory tasks using two different types of to-be-remembered stimuli. The oculomotor delayed response (ODR) task requires a monkey to remember the angle of a visual stimulus for a later saccade. The vibrotactile delayed discrimination (VDD) task requires a monkey to remember the frequency of a vibrotactile stimulus for a short time and then compare it to the frequency of a test stimulus. We model the stimulus receptive fields $g(x_i, \hat{x})$ in these two experiments as a ring describing the possible angles and a line describing log frequency of the vibrations respectively (Fig. 3a,b).

For the ODR task (*Ring*), the stimulus specificity of each cell is specified by a symmetric tuning curve function $g(\theta_i, \hat{\theta})$, where θ_i is preferred angle of cell i and $\hat{\theta}$ is the stimulus presented in the experiment. We choose the tuning curves g to follow the von Mises distribution, which serves as a close approximation to the wrapped normal distribution over a circle, normalized approximately. Cells are chosen so that their preferred directions θ_i tile the stimulus space uniformly, and simulated for eight stimuli conditions $\hat{\theta}$ evenly distributed on the ring (Fig. 3a).

For the VDD task (*Line*), the stimulus dimension is along a line, and ten stimulus conditions $\hat{f}$ are chosen which tile this frequency space evenly from a minimum frequency of $f_{\min} = 10$ Hz to a maximum frequency $f_{\max} = 30$ Hz. The tuning curves coding this space are chosen to be decaying and ramping cells, clamped at these minimum and maximum frequencies, respectively, with rate constants $1/f_i$ distributed geometrically, such that f_i tile the frequencies between $f_{\min}$ and $f_{\max}$ logarithmically (Fig. 3b), akin to the choice of time constants for ‘When’ representations.

Visualizing conjunctive representations

Fig. 3g-i provides a way to visualize the activity over the network for stimuli chosen on a ring at different time points. Each stimulus presented at the beginning of a trial can be described by an angle. As the stimulus recedes into the past, the “true past” that would be recorded by a perfect observer with, say, a video camera is describable as a point on a cylinder. Fig. 3g provides a cartoon of this “true past” at three moments in external time as a function of $\log \tau$. In this depiction, the present is closest to the viewer.

Fig. 3h shows the pattern of activity over the conjunctive what $\times$ when representation with Laplace temporal receptive fields as a function of temporal index n . Fig. 3i shows the conjunctive representation with Inverse Laplace temporal receptive fields displayed in the same way. For the Laplace representation, the angle of the stimulus controls the angular location of the edge. As time passes and the stimulus recedes into the remembered past, the edge moves as a function of $\log \tau$. The Inverse Laplace space has the same properties, except that the remembered time is represented as a bump of activity. Memory for different kinds of stimuli would require a different structure along the stimulus directions.

Dynamics of conjunctive representations in low-dimensional linear projections

We study the population dynamics of conjunctive representations of what $\times$ when with Laplace and Inverse Laplace temporal receptive fields in both ODR and VDD tasks. Analyzing state-space population trajectories of primate PFC neurons, Murray et al. (2017) found stimulus-specific subspaces with persistent ‘What’ representations that held stable even as individual neurons exhibited heterogeneous temporal dynamics. We employ a similar methodology to look at the projections of population-level activity, using Principal Component Analysis (PCA) to account for variance, separately over the stimulus and time space, to generate neural trajectories $\phi(\hat{x}, t)$.

Projecting neural trajectories onto stimulus and time Principal Components

To calculate PCA over stimuli, we compute the stimulus covariance Σ_{what} using the time-averaged delay activity $\langle \Phi_{ij} \rangle_t(\hat{x})$, and use eigendecomposition to extract the first two principal axes $v_{\hat{x},1}$ and $v_{\hat{x},2}$. The projection of the mean-subtracted population activity onto the two principal axes gives us the stimulus principal components

$$\varphi_k(\hat{x}, t) = v_{\hat{x},k} \cdot [\Phi_{ij}(\hat{x}, t) - \langle \Phi_{ij} \rangle_{\hat{x},t}], \text{ for } k = 1, 2. \quad (14)$$

This gives us the first two axes of our neural trajectories. The z axis is constructed to be orthogonal to the stimulus subspace defined by the stimulus axes $v_{\hat{x},k}$, which captures the largest component of time variance in the neural activity. To define this, we calculate the time covariance Σ_{when} over the stimulus-averaged activity $\langle \Phi_{ij} \rangle_{\hat{x}}(t)$, and extract the first principal axis $v_{t,1}$. We then orthogonalize this axis to the stimulus subspace by subtracting the stimulus principal axes from it and normalizing

$$v'_{t,1} = \frac{v_{t,1} - v_{\hat{x},1} - v_{\hat{x},2}}{\|v_{t,1} - v_{\hat{x},1} - v_{\hat{x},2}\|}. \quad (15)$$

We now project the mean subtracted population activity onto the orthogonalized time axis to get the z axis of the neural trajectories

$$\varphi_3(\hat{x}, t) = v'_{t,1} \cdot [\Phi_{ij}(\hat{x}, t) - \langle \Phi_{ij} \rangle_{\hat{x},t}]. \quad (16)$$

There now exist neural trajectories ϕ corresponding to each stimulus condition $\hat{x}$, which are constructed to have two axes which captures stimulus variance, and a z axis that captures time variance (Fig. 4 [↗](#)).

Laplace cells, resembling temporal context cells, show stimulus-specific stable subspaces

Laplace cells, which have exponentially decaying temporal receptive fields, generate a covariance matrix with smoothly decaying activation along the diagonal (from bottom left to top right), which is the direction of maximum variance (Fig. 4b [↗](#)). The first principal component picks this up to show a monotonically varying trend as well (Suppl. Fig. S1 [↗](#), c2).

When the neural activity of Laplace cells is projected onto this smoothly and monotonically varying temporal principal component (z axis), the stimulus representation thus remains stable, with only a relative scaling of the relative distances between the trajectories for different stimuli. Thus, for both ODR (Fig. 4a [↗](#)) and VDD (Fig. 4c [↗](#)) tasks. Although the populations show strong temporal dynamics and show a heterogeneity of timescales, taken together, the population-wide activity seems to encode a stable representation of the stimulus space.

Inverse Laplace cells, resembling time cells, show rotational dynamics

Inverse Laplace neurons, which have sequential receptive fields tiling a one-dimensional timeline, generate a mostly diagonal covariance matrix (Fig. 4i [↗](#)), decaying monotonically from bottom left to top right. The biggest principal components pick this diagonal direction to explain the most variance, and to tile this diagonal line end up being oscillating sinusoids (Suppl. Fig. S1 [↗](#), d2) with a decaying envelope. When the neural activity is projected onto this oscillating component (z-axis), the trajectories end up rotating in the low-dimensional space.

The population trajectories for both ODR (Fig. 4h [↗](#)) and VDD (Fig. 4j [↗](#)) tasks thus show rotational dynamics, although the activity of the neurons themselves do not intrinsically have a oscillatory component. Sequential activity in the brain has also previously been documented to produce artefactual rotational dynamics in linearly projected low-dimensional neural trajectories (Lebedev et al., 2019 [↗](#); Shinn, 2023 [↗](#)).

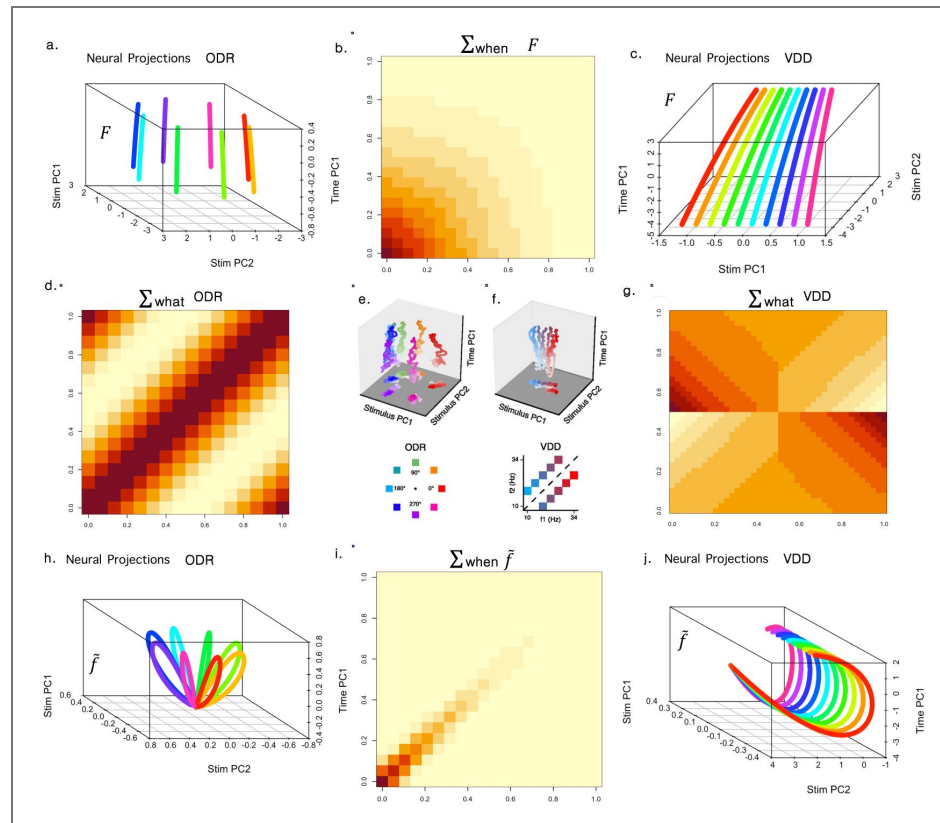


Figure 4. Neural trajectories from LNMs can show stable coding as well as temporal dynamics.

The figures show neural trajectories for ODR (Left) and VDD (right) tasks, computed with Laplace Neural Manifold populations simulating temporal context cells (Laplace, F , Top row, a and c) and time cells (inverse Laplace, $\tilde{f}$, Bottom row, h and j) respectively. The neural activity is projected, via eigendecomposition of the covariance matrices, onto the first two stimulus principal components (Stim PC1 and PC2, x and y axes) and the first temporal principal component (Time PC1, z axes), up to a sign factor. Neural trajectories corresponding to different initial stimulus conditions, the angle θ for the ODR task, shown on the left (a and h), and the frequency f for the VDD task (Right, c and j), are represented with different colors. The overall covariance matrix of the populations can be understood as a tensor product of covariance matrices across different 'whats' - stimulus representations for the ODR (d), and VDD tasks (g), and different 'whens' - temporal coding representations corresponding to the Laplace cells F (b), and the inverse Laplace cells $\tilde{f}$ (i). The corresponding primate PFC neural trajectories, reproduced from Murray et al. (2017), are shown for the ODR task (e) and VDD tasks (f) (Top), along with the stimulus spaces for both tasks (Bottom).

© 2017, PNAS. Parts of panel f are reproduced from Murray et al., 2017 with permission of PNAS. This content is not covered by the CC-BY 4.0 licence and further reproduction of this panel would need permission from the copyright holder.

Even though both populations share the same ‘what’ representation, their activity, when projected onto the stimulus and time principal components, paints very different pictures of their low-dimensional dynamics. This happens due to the choice of temporal basis functions, even though the Laplace and Inverse Laplace representations are simply related by a linear transformation. Both temporal representations are high-rank - as evidenced by the variance in their covariance matrix Σ_{when} , explained by their principal components falling off smoothly (as seen in Suppl. Fig. S1 [c3](#) and [d3](#)). However, the nature of their receptive fields (ramping vs. sequences) gives rise to qualitatively different principal components for time and thus different linear low-dimensional dynamics like stable subspaces or oscillatory dynamics.

Dimensionality of neural trajectories grows with the distribution of time constants

In this section, we examine the dimensionality of the neural space spanned by the population trajectories of neurons with these temporal basis functions. For the choices of temporal basis functions chosen here the dimensionality spanned by the network out to a particular time T depends only on the distribution of time constants. We compute the rank of the overall covariance matrix calculated using times from 0 to T as a measure of dimensionality of neural trajectories.

The rank of the total covariance grows with Σ_{when}

We have seen in [Eq. 8](#) that the total covariance can be expressed as a tensor product of Σ_{when} and Σ_{what} . Because Σ_{what} is not a function of time, all of the time dependence of the covariance matrix must be carried by Σ_{when} . This implies that the growth of linear dimensionality of the space should be the same for different kinds of stimuli maintained in working memory. Starting from [Eq. 8](#), using inequalities describing ranks of matrix sums and products, it is possible to establish lower bounds on the rank of the total covariance

$$\text{rank}(\Sigma) + 1 \geq [\text{rank}(\Sigma_{\text{what}}) - 1] [\text{rank}(\Sigma_{\text{when}}) - 1]. \quad (17)$$

We can see that the rank of the the total covariance goes up with the rank of Σ_{when} , as the rank of Σ_{what} remains fixed once we choose a particular experimental paradigm and the shape and number of receptive fields tiling the stimulus space. Thus, to determine the dimensionality of the neural representation as a function of the total recording time T , we only need to assess the rank of Σ_{when} estimated from an experiment with the first T seconds of the delay.

Before stating the result, let us provide an intuition about how the rank of Σ_{when} should change as a function of T . Consider a pair of Laplace cells ([Fig. 5a](#)) with different time constants. If their time constants are less than T , they will have some covariance and contribute to the rank of the covariance matrix. However, consider pairs of cells that both have time constants much greater than T . These cells would both be firing maximally throughout the interval $0 \leq t \leq T$. The covariance estimated up to time T between these cells is zero. Thus, the pair of cells with time constants much greater than T would not affect the rank of covariance matrix. A similar argument can be made for the Inverse Laplace cells ([Fig. 5b](#)).

Geometrically spaced time-constants result in logarithmic growth in dimensionality of Σ_{when} with recording time T .

As T changes, the number of cells that contribute to the rank of the covariance matrix goes up like the density of the basis $n(T)$, as defined in [Eq. 12](#). Thus, a logarithmic distribution of time constants, as in [Eq. 11](#), implies that the linear dimensionality of the population measured from 0 to T should grow like $\log T$. [Fig. 5c,d](#) show this argument corroborated numerically as the rank of Σ_{when} , computed for both Laplace and Inverse Laplace neurons grows logarithmically as a function of T . This matches the empirical trend of growth of the cumulative dimensionality with time ([Fig. 5e](#)) recorded from cortical regions when performing WM tasks ([Cueva et al., 2020](#)).

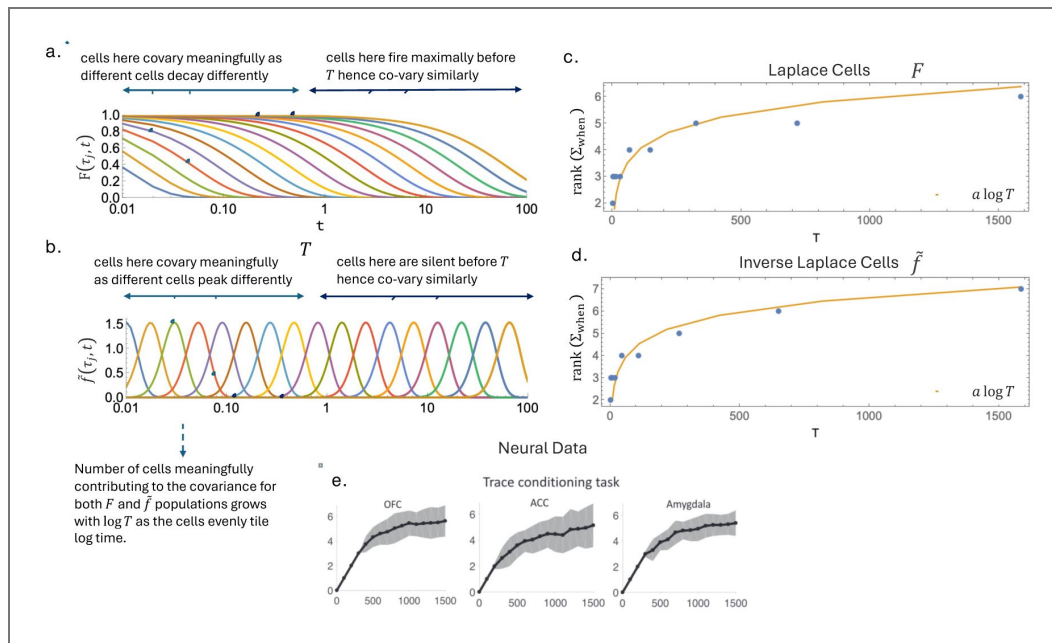


Figure 5. The cumulative dimensionality of neural trajectories in LNM cells grows logarithmically with elapsed time.

Both Laplace and Inverse Laplace cells tile log time evenly. When calculating covariance from 0 to T , cells with time constants much larger than T either both fire maximally (Laplace cells, *a*) or are both silent (Inverse Laplace cells, *b*). For both choices of temporal basis functions, only cells which have time constants less than T co-vary meaningfully and contribute to the total covariance, and the number of such cells grows as $\log T$. Explicitly calculating the rank of the covariance matrix of simulated Laplace and Inverse Laplace cells, for different values of T , shows this empirically (*c*, *d*), and gives us a measure of the dimensionality of their neural trajectories. This seems to follow the growth of cumulative dimensionality of actual neural data (*e*) collected during WM tasks, reproduced from Cueva et al. (2020) [\[1\]](#).

© 2020, PNAS. Panel e is reproduced from Cueva et al., 2020 [\[1\]](#) with permission of PNAS. It is not covered by the CC-BY 4.0 licence and further reproduction of this panel would need permission from the copyright holder.

For generally high recording times, the dimensionality of neural trajectories thus approaches full rank for both Laplace and Inverse Laplace representations. These neurons show nonlinear mixed selectivity, which has previously been shown to be a signature of high-dimensional representations and enables a wide combination of linear readouts (Fusi, Miller, & Rigotti, 2016). Just like neurons in the monkey PFC show high shattering dimensionality (Posani, Wang, Muscinelli, Paninski, & Fusi, 2024; Bernardi et al., 2020) which allows a linear readout to disambiguate different stimulus experimental conditions effectively, Laplace and Inverse Laplace neuronal populations have a high-rank temporal covariance matrix, which allows a wide range of subsets of the continuum of time intervals to be linearly decoded.

Neural Circuit Model for Compositional Working Memory of What $\times$ When

A compositional representation of items in time requires conjunctive representations of what $\times$ when (Eq. 1). With particular choices for temporal receptive fields (Eqs. 9 and 10) we find a good correspondence with a range of neural data (e.g., Figures 4 and 5). This raises the question of how neural circuits could come to have receptive fields that obey these high-level constraints. One possibility is to simply have an RNN that obeys Eq. 2, with temporal recurrent weights chosen to give appropriate temporal receptive fields (see Liu and Howard (2020)). However, a linear RNN would not be robust to perturbations. To address this question, we introduce a continuous attractor neural network (CANN) to implement the Laplace Neural Manifold. The neurons in this CANN will exhibit conjunctive what $\times$ when receptive fields as specified by Eq. 1, with temporal receptive fields given by Eqs. 9 and 10 and logarithmic distribution of time constants as given by Eq. 11.

If the temporal receptive fields across neurons differ by a single parameter that can be mapped onto time, then it is possible to describe the time of a past occurrence of a single stimulus by translating a pattern of activity along a population. As a thought experiment, imagine that we had built a CANN that maintains a memory of the time of past events by constructing a bump of activity that moves at constant velocity along the population as a function of time (W. Zhang, Wu, & Wu, 2022). In this case, the activity of individual neurons would rise and then fall as the bump approaches and then leaves their location along the population. The center of each neuron's temporal receptive field would depend only on its location along the population. If instead of a bump attractor, the network exhibited an edge across the population that moved at constant velocity, then instead of circumscribed temporal receptive fields, we would find monotonic temporal receptive fields.

The temporal receptive fields described by Eqs. 9 and 10 differ from the receptive fields in our thought experiment. Let us return to a bump attractor where the bump moves at a constant rate. In that case, the center of the temporal receptive fields would vary systematically across neurons, but their width in time would be unaffected. In contrast, in the temporal receptive fields used above (Eqs. 9 and 10) the shape of the temporal receptive fields depend only on τ/τ_j . The temporal receptive fields of different neurons are thus rescaled versions of one another, rather than translated versions of one another. Rescaling time translates log time for the simple reason that $\log ax = \log a + \log x$. By choosing logarithmic time constants, Eq. 11, we can exploit the fact that changing τ simply translates the representation of a single event over the population (see Fig. 3b,c, bottom). For Laplace neurons, the activity over the network as a function of n takes the form of an edge; for Inverse Laplace neurons, the activity takes the form of a bump. In both cases, local connections favor nearby neurons to be in the same state. The bump network implementing the Inverse Laplace temporal receptive fields works as a standard local excitation/global inhibition CANN. In the edge network, the two ends of the network are clamped to be in an up and down states, so that an edge appears in between. The edge can appear at any location such that the local connections are far from the clamped ends of the population. In contrast to the hypothetical CANN in which the bump moves at a constant rate, Laplace Neural Manifolds require the edge/bump to move at a velocity that decreases as $1/\tau$ (Daniels & Howard, 2025).

In this section, we flesh out this idea to build a CANN that implements a conjunctive working memory using a Laplace Neural Manifold focusing on the ODR task from the previous section. In this task, the monkey must remember the angle of a visual stimulus, allowing a straightforward model for the representation of “what” using a ring attractor CANN (Fig. 6a [↗](#)), a topic that has been extensively studied for decades (Amari, 1977 [↗](#); K. Zhang, 1996 [↗](#); Redish, Elga, & Touretzky, 1996 [↗](#); Kim, Rouault, Druckmann, & Jayaraman, 2017 [↗](#)). We couple this to an edge/bump attractor to implement temporal receptive fields (Fig. 6b [↗](#)). In this model, paired CANNs with edge solutions and bump solutions interact with one another. Feedback between the two networks causes the edge/bump complex to move at a velocity that goes down with the location of the edge/bump, thus implementing logarithmically compressed temporal receptive fields. Coupling these two ideas together (Fig. 6c [↗](#)) we can construct a single CANN that maintains a compositional representation of what happened when.

In this model a stimulus, characterized by an angle describing its location around the circle is presented at time zero. This forms a bump of activity along the ‘what’ dimension of the cylinder, with an edge/bump at one edge of the cylinder. Feedback from the bump moves the edge from one moment to the next. The strength of this connection decreases as a function of position along the network, causing the edge/bump to move more slowly as time progresses (Fig. 6d [↗](#), e [↗](#)).

Separate CANNs for what and when information

Before describing the combined network, for expository purposes, we first step through the properties of the CANNs for what and when in isolation.

Ring attractor tracks stimulus space ‘What’

To build a ring attractor to maintain information about the location of the visual stimulus, we assume that the neural interactions are Gaussian: $J(x, x') = \frac{A}{\sqrt{2\pi}a} \exp\left[-\frac{(x-x')^2}{2a^2}\right]$, i.e., we impose strong local excitation and weak global excitation over the ring attractor network. The input to each neuron saturates under global activity-dependent inhibition

$$\sigma(r(x); k, \rho) = \frac{r(x, t)^2}{1 + k\rho \|r(\cdot, t)\|^2} \quad (18)$$

given the inhibition strength κ and neural density ρ . The dynamics of the synaptic input $r(x, t)$ is governed by:

$$\frac{dr(x, t)}{dt} = -r(x, t) + \rho \sum_{x'} J(x, x') \sigma(r(x'); k, \rho). \quad (19)$$

We have chosen the time constant of the dynamics to be 1 and assume it is much faster than the functional time constants of the receptive fields. When the parameters are chosen properly the ring attractor network can maintain a stable “activity bump” centered around the initial input z : $\tilde{r}(x|z) = r_0 \exp\left[-\frac{(x-z)^2}{4a^2}\right]$, where a is the width of Gaussian interaction. The constant of proportionality r_0 depends on the parameters of the network, including the value of the inhibition parameter κ . This bump can shift around the ring in response to changes in stimulus identity. The stimulus identity is thus encoded by the manifold coordinate $\tilde{x} = z$.

Paired edge-bump attractors track temporal space ‘When’

Previous work has shown it is possible to build a continuous attractor model for Laplace/inverse representations of a delta function (Daniels & Howard, 2025 [↗](#)) under the assumption that the time constants $\{\tau_n\}$ form a geometric series. In this neural network the relative firing rate $r_E(\tau_i, t)$ of all neurons is governed by the dynamics:

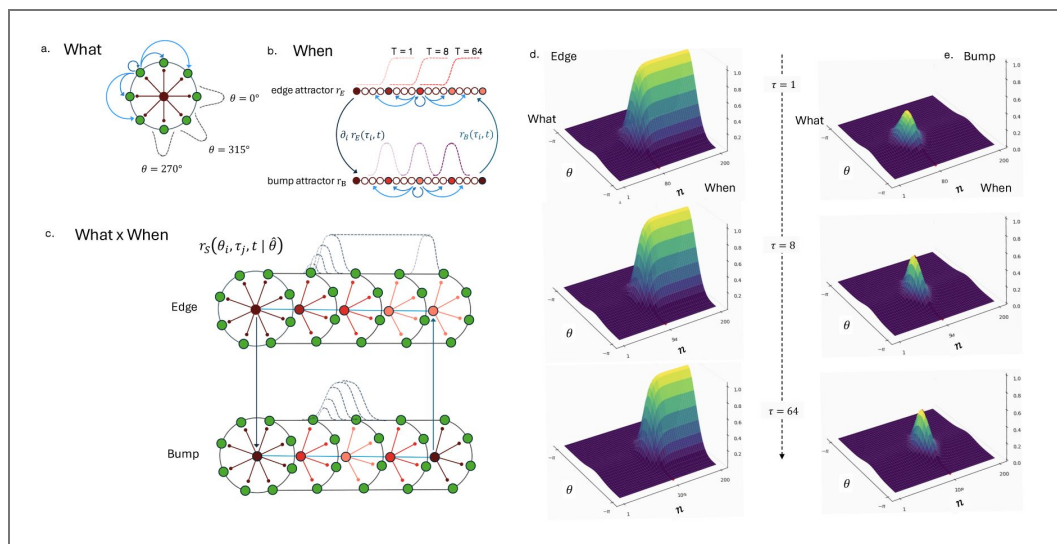


Figure 6. Continuous Attractor Neural Networks (CANNs) can be used to construct Laplace Neural Manifolds of what $\times$ when information.

a. What: A ring attractor maintains a persistent representation of the presented stimulus by sustaining a bump activation at that angle. Other kinds of stimulus information could be maintained with appropriate circuits. *b. When:* A specialized set of line attractors can maintain an edge (Laplace) and a bump (Inverse Laplace) to implement temporal receptive fields evenly spaced over log time. Recurrent connections maintain a particular shape of activity across the network; an edge for Laplace temporal receptive fields and a bump for Inverse Laplace temporal receptive fields. Connections between layers move the edge/bump at an appropriate speed, resulting in temporal receptive fields as a function of log time. *c. What $\times$ When:* A series of ring attractors as in *a*, crossed with lines, as in *b* create a cylinder. The edge/bump attractor for time supplies global inhibition to the ring attractors coding for stimulus identity. *d, e.* Paired cylinders with backbones emulating the edge and bump respectively, can maintain the activity of Laplace and Inverse Laplace representations which shift as we progress along the remembered timeline τ akin to Fig. 3, together forming a Laplace Neural Manifold for time.

$$\frac{dr_E(\tau_i, t)}{dt} = -r_E(\tau_i, t) + \sum_{j=1}^N W_{EE}(i, j) \phi(r_E; g) + I_{ext}^{BE}(\tau_i, t)$$

where W_{EE} is the recurrent matrix within the neural population, and Φ is the nonlinear transfer function $\Phi(x) = \tanh(gx)$ with gain g . Again, we assume the time constant of this dynamical system to be 1 and much less than the functional time constants τ_i . By carefully choosing the recurrent matrix W_{EE} and the gating g , the network shows monotonic exponential receptive fields if we provide external input to clamp two edges of the network such that at $i = 1$ and $i = N$ r_E are constrained to be -1 and 1 respectively. With less careful tuning, any solution that gives an edge that moves at the appropriate speed will give non-exponential monotonic receptive fields (Daniels & Howard, 2025 [DOI](#)).

To cause the edge to move at a decreasing rate as a function of time we provide a dynamical input I_{ext}^{BE} to the edge attractor. This input comes via another neuronal population that forms a bump attractor in register with the edge attractor:

$$\frac{dr_B(\tau_i, t)}{dt} = -r_B(\tau_i, t) + \sum_{j=1}^N W_{BB}(i, j) \phi(r_B; g) + I_{ext}^{EB}(\tau_i, t)$$

where $I_{ext}^{BE}(\tau_i, t) \equiv r_E(\tau_{i+1}, t) - r_E(\tau_i, t)$ and W_{BB} is chosen order for the neural network to maintain continuous bumpshaped states. There is a natural one-to-one correspondence between the edge attractor and the bump attractor as the corresponding units i are associated with the same time constant τ_i .

The bump attractor receives input from the edge attractor forming a bump at the position of the edge. Meanwhile, the bump network provides input to the edge attractor to stimulate it to move with a desired speed $I_{ext}^{BE}(\tau_i, t) = \nu(\tau_i) r_B(\tau_i, t)$. By customizing the speed function $\nu(t)$ the edge position grows logarithmically in time:

$$\bar{n}(t) = \bar{n}_0 + \log_a \frac{t}{t_0} \quad (20)$$

where $\bar{n}(t) = \arg \max_n \{r_E(\tau_{n+1}, t) - r_E(\tau_n, t)\}$ denotes the edge position at time t , and $\bar{n}_0, t_0$ are free parameters that determine the initial condition of the edge. The edge/bump attractors exhibit dynamics that correspond to the temporal receptive fields in [Eqs. 9](#) and [10](#) as follows:

$$\begin{aligned} \frac{1}{2}[r_E(\tau_j, \tau) + 1] &\propto e^{-\frac{t}{\tau_j}}, \\ \frac{1}{2}[r_B(\tau_j, \tau) + 1] &\propto \left(\frac{t}{\tau_j}\right)^k e^{-k(t/\tau_j)}. \end{aligned} \quad (21)$$

Thus this edge/bump CANN can implement the temporal receptive fields hypothesized for Laplace Neural Manifolds.

Ring × edge-bump attractors track What × When

The entire model is composed of a series of ring attractors that together form a 2-D manifold that can be visualized as a cylinder. Each neuron in this 2-D manifold is indexed by i and j , indicating its preferred orientation x_i on the ring and time constant τ_j . There are connections between the neurons in each ring of the cylinder, but in this simple model there are no connections between the neurons across different rings. The analytical solution of this continuous ring attractor shows that the height of the “activity bump” of a particular ring depends on the inhibition strength κ :

$$r(x_i|\hat{x}) = r_0(k) \exp \left[-\frac{(x_i - \bar{x})^2}{2a^2} \right]. \quad (22)$$

To generate conjunctive what × when receptive fields, each cylinder is controlled by an edge/bump “spine” that controls the magnitude of the inhibition κ of the ring attractor for each value of τ_j at each moment. To obtain conjunctive what × when receptive fields, the inhibition strength at the ring at τ_j depends on the activation of the spine edge-bump attractor, r_E/r_B . The simulations in Fig. 6d and 6e use inhibition chosen as:

$$k_S(\tau_j) = a_S r_E(\tau_j, \tau) + b_S \quad (23)$$

with $S \in \{E, B\}$. The parameter values in Fig. 6d and e are $a_E = -9.60$, $b_E = 10.71$, $a_B = -18.10$, $b_B = 20.31$. These values were chosen such that $r_0(k)$ lies within $[0, 1]$ for the edge attractor network as τ varies.

Encoding of What × When. Combining previous results, we are able to write out the dynamics of Laplace/Inverse Laplace neural manifolds:

$$\begin{aligned} \frac{dr_S(x_i, \tau_j, t|\hat{x})}{dt} = & -r_S(x_i, \tau_j, t|\hat{x}) \\ & + \rho \sum_k J(x_i, x_k) \sigma(r_S; \rho, k_S(\tau_j)). \end{aligned} \quad (24)$$

Fig. 6d and 6e show results from this network choosing a specific starting position $\hat{x}$. Neurons in our 2-D manifold simultaneously encode both the ‘what’ and ‘when’ information. The identity of the stimulus can be readily decoded by observing the center of the bump across the rings, while the timing of the stimulus can be inferred from the location of edge/bump along the ‘when’ axis. This simple circuit model is sufficient to generate conjunctive temporal receptive fields with properties consistent with the properties desired for h_F and $h_{\bar{F}}$.

Discussion

This paper explores the implications of a compositional neural representation of what happened when in working memory. Neurons in a population with such a compositional representation should exhibit conjunctive receptive fields for what and when, as has been reported in many brain regions (Tiganj et al., 2018; Terada, Sakurai, Nakahara, & Fujisawa, 2017; Taxidis et al., 2020). This property makes it straight-forward to write out closed form expressions for the covariance matrix of this population, which in turn allows us to work out the dynamics of the population when studied using standard linear dimensionality reduction techniques. The low-dimensional dynamics depends dramatically on the choice of temporal basis functions, even when the basis functions are related to one another by a simple linear transformation. With conjunctive receptive fields, the dimensionality of the space spanned by the the population is controlled by the rank of the temporal covariance matrix, allowing the density of basis functions to be directly assessed. We show that a logarithmic tiling of time, as proposed by work in cognitive psychology (Fechner, 1860/1912; Stevens, 1957; Murdock, 1960; Luce & Suppes, 2002; Brown, Neath, & Chater, 2007; Balsam & Gallistel, 2009) and supported by evidence from neuroscience (Cao et al., 2022; Guo et al., 2021) provides a reasonable approximation of empirical data. Finally, we sketch out a circuit model using continuous attractor neural networks that exhibits conjunctive what × when receptive fields when a single item is present in working memory.

Network Models in Neuroscience

Early attractor models used persistent activity of neurons to encode a task variable in working memory. The current paper builds on and extends network models that attempt to reconcile stable task representations with heterogeneous temporal dynamics (Druckmann & Chklovskii, 2012; Murray et al., 2017). In those papers, the requirement on temporal dynamics is to leave some stimulus-coding variable invariant to the passage of time. For instance, Druckmann and Chklovskii (2012) required that the summed activity over neurons coding for a particular stimulus is

constant. The inverse Laplace cells (Eq. 10) in the current paper can satisfy that requirement with appropriate normalization; because the temporal receptive fields decay monotonically, the Laplace cells (Eq. 9) do not satisfy this requirement.

More broadly, the strategy in those papers is to keep the stimulus representation within a “null space” of the network dynamics. As long as the changes in individual neuronal activity occur along directions in the high-dimensional activity space that do not affect the decoded stimulus representation, the stimulus can be linearly decoded. For instance Murray et al. (2017) used a circuit model that has a mnemonic coding subspace where the input stimulus activates a stable representation, and a non-coding subspace which exhibits temporal dynamics that are orthogonal to the coding subspace.

The major distinction between those previous models and the present paper is that our approach is designed to construct a particular temporal representation whereas previous papers did not place an emphasis on any particular form of temporal dynamics. For instance Cueva et al. (2020) used a low-rank RNN (Beiran, Meirhaeghe, Sohn, Jazayeri, & Ostojic, 2023) (see also Beiran et al. (2023)), to allow temporal dynamics along with a stable stimulus representation. The present approach starts from the prior that memory for the passage of time requires a continuous distribution of time constants, effectively tiling the time axis. If the time constants of a network must be trained by the requirements of a particular task, this leaves the memory less able to express temporal information in different tasks, or even in situations where there is no explicit demand for timing information. For instance, we are aware of the passage of time even when there is no explicit requirement to make a response (Husserl, 1966). Commitment to a timeline of what happened when leaves the memory able to effectively decode *any* what at *any* when.

Attractor circuits as a mechanism for slow functional time constants

The proposed circuit model using CANNs provides an existence proof that long functional time constants need not be a consequence of intrinsic time constants of individual neurons (Fransén, Alonso, & Hasselmo, 2002; Fransén, Tahvildari, Egorov, Hasselmo, & Alonso, 2006; Loewenstein & Sompolinsky, 2003; Tiganj, Hasselmo, & Howard, 2015; Saber Marouf et al., 2025) nor of statistical interactions between modes in a nearly-random RNN (Dahmen, Grün, Diesmann, & Helias, 2019; Helias & Dahmen, 2020). Ságodi, Martín-Sánchez, Sokó, and Park (2024) described conditions under which an RNN forms a line attractor (see also Can and Krishnamurthy (2024); Krishnamurthy, Can, and Schwab (2022)). Our results require further that line attractors should be formed for all ‘what’s that can exist in memory. In order to get basis functions that are evenly spaced over log time, the network must have degenerate eigenvalues in geometric series *and* eigenvectors that are translated versions of one another (Liu & Howard, 2020). Different temporal basis functions, which are apparently both observed in the mammalian brain, also require distinct forms of line attractors. Thus the cognitive requirement for compositional representations of what happened when provide strong high-level constraints on physical models of working memory maintenance.

Cognitive psychology of working memory

In this paper we considered working memory for retention of a single item in continuous time. Cognitive psychologists have studied considerably more complex working memory tasks involving multiple stimuli that are to be remembered. Time *per se* has little effect on either visual working memory performance (Shin, Zou, & Ma, 2017; Kahana & Sekuler, 2002), nor on verbal working memory performance (Baddeley & Hitch, 1977). However, there is widespread evidence that the amount of information that can be maintained in working memory in a highly veridical manner is finite. For many decades the dominant view has been that working memory maintains a discrete number of items in a discrete number of “slots” (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Luck & Vogel, 1997). However, more recently, cognitive psychologists’ understanding of the nature of working memory has become much more subtle (Ma, Husain, & Bays, 2014; Brady,

Störmer, & Alvarez, 2016 [↗](#); Schurgin, Wixted, & Brady, 2020 [↗](#)). In the words of Ma et al. (2014) [↗](#) “working memory might better be conceptualized as a limited resource that is distributed flexibly among all items to be maintained in memory.”

It has long been understood that recurrent attractor neural networks provide a natural means to understand capacity limitations (Grossberg, 1969 [↗](#), 1978 [↗](#)). Mutual inhibition controls the number of attractors, or bumps, that can be simultaneously sustained by recurrent activation. One can readily imagine extending the model described in Fig. 6 [↗](#) to allow multiple bumps to coexist. Given a particular structure of the “what” network, depending on how inhibition flows across the network, one can control capacity within a time point (as is typically done in visual working memory experiments) or across time points (as is typically done in verbal working memory experiments). In light of recent experimental work, much care should be taken in allowing stimulus features to cooperate and compete in working memory.

The wiring requirements necessary to extend this approach such that *all possible stimuli* can be maintained in working memory through continuous time seem daunting. For instance, what if the task also required memory for the color of the stimuli as well as their location? Or the orientation and color of a bar at an angular location? Or the identity of a letter at an angular location? The circuit would have to either be extremely complicated *a priori* or be able to be dynamically configured for task-relevant information.

Perhaps recent work on gated RNNs provides a way for these continuous attractor networks to dynamically form in response to task demands (Krishnamurthy et al., 2022 [↗](#); Can & Krishnamurthy, 2024 [↗](#)). In this way, gated input would dynamically specify the features that can be maintained in a continuous attractor network for what information.

Time and space and number

One could have easily generalized the compositionality argument for what × when to what × spatial position. Indeed, Weyl (1922) [↗](#) used precisely the same argument for space in developing an axiomatic development of space-time in general relativity. The requirement of compositional representations for what and when leads naturally to conjunctive neural codes.

In addition to physical space, one might have made analogous arguments for any number of continuous variables. Conjunctive codes for what × when is thus a special case of mixed selectivity, which has been proposed as a general property for neural codes (Rigotti et al., 2013 [↗](#)) and has been observed for many different variables in neural representations (Mante, Sussillo, Shenoy, & Newsome, 2013 [↗](#); Ward, Tan, & Grenfell-Essam, 2010 [↗](#); Johnston, Palmer, & Freedman, 2020 [↗](#); Dang, Jaffe, Qi, & Constantinidis, 2021 [↗](#); Wallach, Melanson, Longtin, & Maler, 2022 [↗](#)). Conjunctive receptive fields are also closely related to gain fields which were proposed for coordinate transformations and efficient function learning (Pouget & Sejnowski, 1997 [↗](#); Salinas & Abbott, 2001 [↗](#)).

Numerosity is a variable that can be composed (in principle) with any kind of object. Mammals (Gallistel & Gelman, 1992 [↗](#); Dehaene & Brannon, 2011 [↗](#)) and other animals (Kirschhock & Nieder, 2023 [↗](#)) are equipped with a number sense. This number sense appears to be computed over a neural scale that maps receptive fields onto the log of numerosity (Nieder & Miller, 2003 [↗](#); Dehaene, 2003 [↗](#); Nieder & Dehaene, 2009 [↗](#)), presumably using a set of basis functions distributed over a logarithmic number line.

Theoretically-driven neural data analysis techniques

The appearance of population dynamics projected onto a low-dimensional PCA space changed dramatically depending on how we chose temporal basis functions (Fig. 4 [↗](#)). Even though the temporal basis functions are simply related by a linear transformation, one might have concluded that they reflect very different coding schemes if one only looked at the PCA results. For instance, one may have concluded that the Laplace representation shows a stable subspace with persistent

firing (Murray et al., 2017; Constantinidis et al., 2018) whereas the inverse space shows sustained delay-period coding without persistent firing (King & Dehaene, 2014; Lundqvist, Herman, & Miller, 2018).

This is another example that extreme caution should be exercised in using linear dimensionality reduction techniques in neuroscience (Lebedev et al., 2019; De & Chaudhuri, 2023; Shinn, 2023). Jazayeri and Ostojic (2021) write that “... without concrete computational hypotheses, it could be extremely challenging to interpret measures of dimensionality.” If neural codes use basis functions tiling continuous variables out in the world, then the linear dimensionality of the neural code is in principle unbounded (Manley et al., 2024). Depending on the choice of basis functions, the principle components of the population may, or may not be readily interpretable, even if we know *a priori* the continuous variables that are being coded by the population.

A number of well-studied non-linear dimensionality reduction techniques exist (Belkin & Niyogi, 2003; Kohli, Cloninger, & Mishne, 2021; Tenenbaum, de Silva, & Langford, 2000; Roweis & Saul, 2000; McInnes, Healy, & Melville, 2018; Van der Maaten & Hinton, 2008). Although still vastly outnumbered by data analyses using linear dimensionality reduction techniques, some neuroscience work has made use of non-linear dimensionality reduction (Chaudhuri, Gerçek, Pandey, Peyrache, & Fiete, 2019; Nieh et al., 2021) (see Langdon, Genkin, & Engel, 2023 for a review). Resolving the empirical question of the generality of compositional codes using conjunctive basis functions will require careful experimentation and development of new data analysis tools.

Given a hypothesis about relevant variables, it should be possible to distinguish if the decomposition of the covariance matrix as asserted in Eq. 8 is valid. This would establish that a conjunctive code exists. Proximity of the parameters describing the receptive fields, e.g., s_i , τ_i^* , n , provide a natural metric for proximity along the population manifold. However, these parameters need to be recovered from the data, preferably without a strong prior on the form of the functions describing the receptive fields.

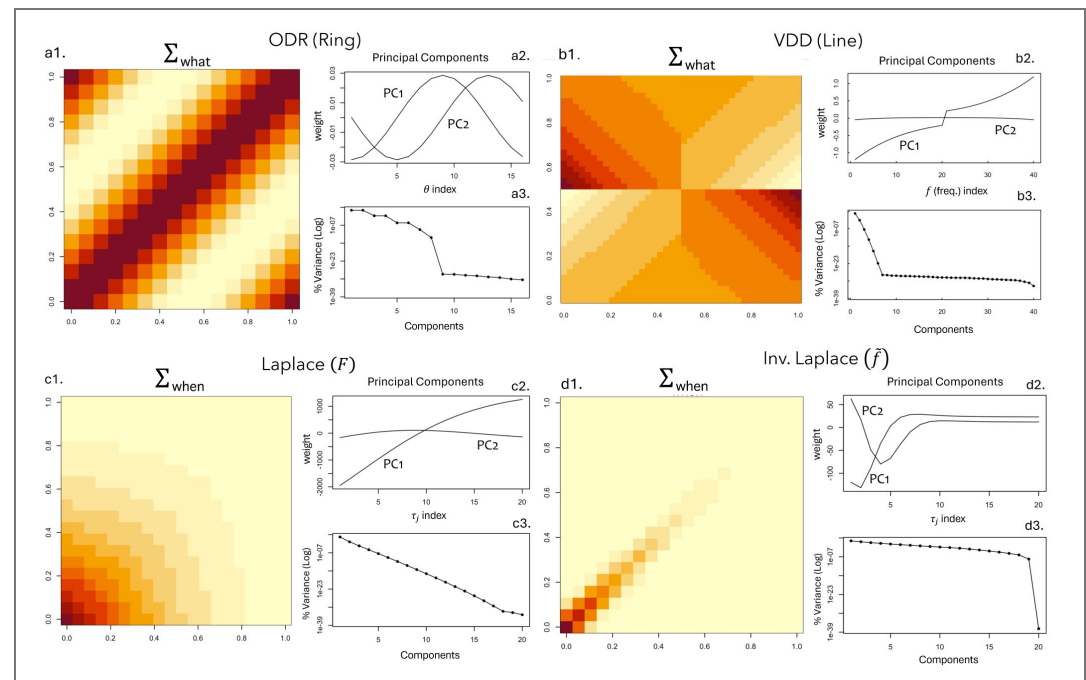


Fig. S1. Covariance Matrices and Principal Components for Σ_{what} (top) and Σ_{when} (bottom), with the variance explained by the components shown in a log scale. Top: The stimulus covariance matrix (Σ_{what}) shown for the ODR (a1, Left) and VDD (a2, Right) tasks, where the task variable lie on a ring and a line, respectively. The circular task variable is tiled with circular Gaussian (Von Mises) receptive fields with periodic boundary conditions, producing a periodic covariance matrix, which generates two sinusoidal principal components (a2) explaining the most variance (a3). The line variable is tiled with sets of exponentially decaying and ramping cells, which generates distinct quadrants in the covariance matrix (b1), with a smoothly-varying first principal

component (b_2) (with a discontinuity when switching from decaying to ramping cells) which explains the most variance (b_3), and a mostly flat second component. *Bottom*: The time covariance matrix (Σ_{when}) shown for Laplace (temporal context cells, F) and Inverse Laplace (time cells, $\tilde{f}$) populations. With exponentially decaying cells, the covariance matrix for Laplace neurons (c_1) has a smoothly decaying activation across the diagonal, which is the direction of maximum variance. The first principal component picks this up to show a *monotonically-varying* trend (c_2) and also explains the most variance (c_3). Time cells on the other hand are sequentially firing cells tiling a time-line, which shows up as a diagonally-concentrated covariance matrix which decays from bottom right to top left (d_1). The biggest principal components (d_2) of this covariance matrix also end up being oscillating sinusoids (similar to a_2) but with a decaying envelope as we go from left to right (smaller to larger τ_j), and explain the most variance (d_3). For *both* these temporal representations, the variance explained (set on a *logarithmic* scale) with number of principal components goes down smoothly, underscoring the *high-dimensional* nature of the covariance matrices (c_3 and d_3).

Note

This reviewed preprint has been updated to correct the author order.

References

- 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 <https://doi.org/10.7554/eLife.19507>
- Amari S.** (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics* **27**:77-87
- Atanas A. A.**, Kim J., Wang Z., Bueno E., Becker M., Kang D. (2023) Brain-wide representations of behavior spanning multiple timescales and states in *c. elegans*. *Cell* **186**:4134-4151
- Atkinson R. C.**, Shiffrin R. M. (1968) Human memory: A proposed system and its control processes. In: Spence K. W., Spence J. T. (Eds). *The psychology of learning and motivation* **2** New York: Academic Press. pp. 89-105
- Baddeley A. D.**, Hitch G. J. (1974) Working memory. In: Bower G. H. (Ed). *The psychology of learning and motivation: Advances in research and theory* **8** New York: Academic Press. pp. 47-90
- Baddeley A. D.**, Hitch G. J. (1977) Recency reexamined. In: Dornic S. (Ed). *Attention and performance VI* Hillsdale, NJ: Erlbaum. pp. 647-667
- Balsam P. D.**, Gallistel C. R. (2009) Temporal maps and informativeness in associative learning. *Trends in Neuroscience* **32**:73-78
- Beiran M.**, Meirhaeghe N., Sohn H., Jazayeri M., Ostojic S. (2023) Parametric control of flexible timing through low-dimensional neural manifolds. *Neuron*
- Belkin M.**, Niyogi P. (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* **15**:1373-1396
- Bergson H.** (1910) *Time and free will: An essay on the immediate data of consciousness* G. Allen & Unwin.
- Bernardi S.**, Benna M. K., Rigotti M., Munuera J., Fusi S., Salzman C. D. (2020) The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* **183**:954-967
- Brady T. F.**, Störmer V. S., Alvarez G. A. (2016) Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences* **113**:7459-7464
- 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
- Brown G. D. A.**, Neath I., Chater N. (2007) A temporal ratio model of memory. *Psychological Review* **114**:539-76
- Can T.**, Krishnamurthy K. (2024) Emergence of memory manifolds. arXiv preprint arXiv:2109.03879.
- Cao R.**, Bladon J. H., Charczynski S. J., Hasselmo M., Howard M. (2022) Internally generated time in the rodent hippocampus is logarithmically compressed. *eLife* **11**:e75353 <https://doi.org/10.7554/eLife.75353>
- Cao R.**, Bright I. M., Howard M. W. (2024) Ramping cells in the rodent medial prefrontal cortex encode time to past and future events via real laplace transform. *Proceedings of the National Academy of Sciences* **121**:e2404169121
- Chaudhuri R.**, Gerçek B., Pandey B., Peyrache A., Fiete I. (2019) The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience* **22**:1512-1520

- Constantinidis C., Funahashi S., Lee D., Murray J. D., Qi X.-L., Wang M., Arnsten A. F. T. (2018) Persistent spiking activity underlies working memory. *Journal of Neuroscience* **38**:7020-7028 <https://doi.org/10.1523/JNEUROSCI.2486-17.2018>
- Cueva C. J., Saez A., Marcos E., Genovesio A., Jazayeri M., Romo R., Fusi S. (2020) Low-dimensional dynamics for working memory and time encoding. *Proceedings of the National Academy of Sciences* **117**:23021-23032
- Dahmen D., Grün S., Diesmann M., Helias M. (2019) Second type of criticality in the brain uncovers rich multiple-neuron dynamics. *Proceedings of the National Academy of Sciences* **116**:13051-13060
- Dang W., Jaffe R. J., Qi X.-L., Constantinidis C. (2021) Emergence of nonlinear mixed selectivity in prefrontal cortex after training. *Journal of Neuroscience* **41**:7420-7434
- Daniels B. C., Howard M. W. (2025) Continuous attractor networks for laplace neural manifolds. *Computational Brain & Behavior* 1-18
- De A., Chaudhuri R. (2023) Common population codes produce extremely nonlinear neural manifolds. *Proceedings of the National Academy of Sciences* **120**:e2305853120
- Dehaene S. (2003) The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences* **7**:145-147
- Dehaene S., Brannon E. (2011) *Space, time and number in the brain: Searching for the foundations of mathematical thought* Academic Press.
- De Vries B., Principe J. C. (1992) The gamma model—a new neural model for temporal processing. *Neural networks* **5**:565-576
- Druckmann S., Chklovskii D. B. (2012) Neuronal circuits underlying persistent representations despite time varying activity. *Current Biology* **22**:2095-2103
- Fechner G. (1860) (1912) *Elements of psychophysics. Vol. I* Houghton Mifflin.
- Fransén E., Alonso A. A., Hasselmo M. E. (2002) Simulations of the role of the muscarinic-activated calcium-sensitive nonspecific cation current INCM in entorhinal neuronal activity during delayed matching tasks. *Journal of Neuroscience* **22**:1081-1097
- Fransén E., Tahvildari B., Egorov A. V., Hasselmo M. E., Alonso A. A. (2006) Mechanism of graded persistent cellular activity of entorhinal cortex layer V neurons. *Neuron* **49**:735-46
- Fusi S., Miller E. K., Rigotti M. (2016) Why neurons mix: high dimensionality for higher cognition. *Current opinion in neurobiology* **37**:66-74
- Gallistel C. R., Gelman R. (1992) Preverbal and verbal counting and computation. *Cognition* **44**:43-74
- Grossberg S. (1969) On the serial learning of lists. *Mathematical Biosciences* **4**:201-253
- Grossberg S. (1978) Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models?. *Journal of Mathematical Psychology* **17**:199-219
- 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**:1-12
- Hacker M. J. (1980) Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory* **15**:846-858
- Helias M., Dahmen D. (2020) *Statistical field theory for neural networks* **970** Springer.
- Hintzman D. L. (2010) How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition* **38**:102-15
- Howard M. W., Esfahani Z. G., Le B., Sederberg P. B. (2024) Learning temporal relationships between symbols with Laplace Neural Manifolds. *Computational Brain and Behavior* <https://doi.org/10.1007/s42113-024-00230-8>
- Howard M. W., Shankar K. H. (2018) Neural scaling laws for an uncertain world. *Psychological 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**:24-53
- Husserl E. (1966) *The phenomenology of internal time-consciousness* Bloomington, IN: Indiana University Press.
- Jacques B. G., Tiganj Z., Sarkar A., Howard M., Sederberg P. (2022) A deep convolutional neural network that is invariant to time rescaling. In: International conference on machine learning. pp. 9729-9738
- James W. (1890) *The principles of psychology* New York: Holt.
- Jazayeri M., Ostojic S. (2021) Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current opinion in neurobiology* **70**:113-120
- 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**:19156-19161
- Johnston W. J., Palmer S. E., Freedman D. J. (2020) Nonlinear mixed selectivity supports reliable neural computation. *PLoS computational biology* **16**:e1007544
- Kahana M. J., Sekuler R. (2002) Recognizing spatial patterns: a noisy exemplar approach. *Vision Research* **42**:2177-192
- Kim S. S., Rouault H., Druckmann S., Jayaraman V. (2017) Ring attractor dynamics in the drosophila central brain. *Science* **356**:849-853
- King J., Dehaene S. (2014) Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences* **18**:203-210
- Kirschhock M. E., Nieder A. (2023) Numerical representation for action in crows obeys the weber-fechner law. *Psychological Science* **34**:1322-1335
- Kohli D., Cloninger A., Mishne G. (2021) LDLE: Low distortion local eigenmaps. *Journal of machine learning research* **22**:1-64
- Krishnamurthy K., Can T., Schwab D. J. (2022) Theory of gating in recurrent neural networks. *Physical Review X* **12**:011011
- Langdon C., Genkin M., Engel T. A. (2023) A unifying perspective on neural manifolds and circuits for cognition. *Nature Reviews Neuroscience* 1-15
- Lebedev M. A., Ossadtchi A., Mill N. A., Urpí N. A., Cervera M. R., Nicolelis M. A. (2019) What, if anything, is the true neurophysiological significance of rotational dynamics?. *BioRxiv* 597419
- Liu Y., Howard M. W. (2020) Generation of scale-invariant sequential activity in linear recurrent networks. *Neural Computation* **32**:1379-1407
- Loewenstein Y., Sompolinsky H. (2003) Temporal integration by calcium dynamics in a model neuron. *Nature Neuroscience* **6**:961-7
- Luce R. D., Suppes P. (2002) Representational measurement theory. In: Wixted J., Pashler H. (Eds). *Stevens handbook of experimental psychology, 3rd edition* **4** Wiley Online Library. pp. 1-41
- Luck S. J., Vogel E. K. (1997) The capacity of visual working memory for features and conjunctions. *Nature* **390**:279-81
- Lundqvist M., Herman P., Miller E. K. (2018) Working memory: Delay activity, yes! persistent activity? maybe not. *Journal of Neuroscience* **38**:7013-7019
- Ma W. J., Husain M., Bays P. M. (2014) Changing concepts of working memory. *Nature neuroscience* **17**:347-356
- Maass W., Natschläger T., Markram H. (2002) Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation* **14**:2531-60
<https://doi.org/10.1162/089976602760407955>
- MacDonald C. J., Lepage K. Q., Eden U. T., Eichenbaum H. (2011) Hippocampal “time cells” bridge the gap in memory for discontinuous events. *Neuron* **71**:737-749

- Machens C. K.**, Romo R., Brody C. D. (2010) Functional, but not anatomical, separation of what and when in prefrontal cortex. *Journal of Neuroscience* **30**:350-360
- Manley J.**, Lu S., Barber K., Demas J., Kim H., Meyer D., Vaziri A. (2024) Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron* **112**:1694-1709
- Mante V.**, Sussillo D., Shenoy K. V., Newsome W. T. (2013) Context-dependent computation by recurrent dynamics in pre-frontal cortex. *Nature* **503**:78-84
- McInnes L.**, Healy J., Melville J. (2018) Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.
- Murdock B. B.** (1960) The distinctiveness of stimuli. *Psychological Review* **67**:16-31
- Murray J. D.**, Bernacchia A., Roy N. A., Constantinidis C., Romo R., Wang X.-J. (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences* **114**:394-399
<https://doi.org/10.1073/pnas.1619449114>
- Nieder A.**, Dehaene S. (2009) Representation of number in the brain. *Annual Review of Neuroscience* **32**:185-208 <https://doi.org/10.1146/annurev.neuro.051508.135550>
- Nieder A.**, Miller E. K. (2003) Coding of cognitive magnitude: compressed scaling of numerical information in the primate pre-frontal cortex. *Neuron* **37**:149-57
- Nieh E. H.**, Schottdorf M., Freeman N. W., Low R. J., Lewallen S., Koay S. A., Tank D. W. (2021) Geometry of abstract learned knowledge in the hippocampus. *Nature* **595**:80-84
- Parker N. F.**, Baidya A., Cox J., Haetzel L. M., Zhukovskaya A., Murugan M., Witten I. B. (2022) Choice-selective sequences dominate in cortical relative to thalamic inputs to nac to support reinforcement learning. *Cell Reports* **39**:110756
- Pastalkova E.**, Itskov V., Amarasingham A., Buzsaki G. (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**:1322-7
- Piantadosi S. T.** (2016) A rational analysis of the approximate number system. *Psychonomic Bulletin & Review* **1**-10
- Posani L.**, Wang S., Muscinelli S., Paninski L., Fusi S. (2024) Rarely categorical, always high-dimensional: how the neural code changes along the cortical hierarchy. *bioRxiv*
- Pouget A.**, Sejnowski T. J. (1997) Spatial transformations in the parietal cortex using basis functions. *Journal of cognitive neuroscience* **9**:222-237
- Redish A. D.**, Elga A. N., Touretzky D. S. (1996) A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems* **7**:671-685
- Rigotti M.**, Barak O., Warden M. R., Wang X.-J., Daw N. D., Miller E. K., Fusi S. (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**:585-90 <https://doi.org/10.1038/nature12160>
- Roweis S. T.**, Saul L. K. (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**:2323-6 <https://doi.org/10.1126/science.290.5500.2323>
- Saber Marouf B.**, Reboreda A., Theissen F., Kaushik R., Sauvage M., Dityatev A., Yoshida M. (2025) Single neurons act as a memory buffer for space. *bioRxiv*
- Ságodi Á.**, Martín-Sánchez G., Sokół P., Park I. M. (2024) Back to the continuous attractor. *ArXiv*
- Salinas E.**, Abbott L. (2001) Coordinate transformations in the visual system: how to generate gain fields and what to compute with them. *Progress in brain research* **130**:175-190
- Schurgin M. W.**, Wixted J. T., Brady T. F. (2020) Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour* **4**:1156-1172
- Shankar K. H.**, Howard M. W. (2012) A scale-invariant internal representation of time. *Neural Computation* **24**:134-193

- Shankar K. H., Howard M. W. (2013) Optimally fuzzy temporal memory. *Journal of Machine Learning Research* **14**:3753-3780
- Shin H., Zou Q., Ma W. J. (2017) The effects of delay duration on visual working memory for orientation. *Journal of vision* **17**:10-10
- Shinn M. (2023) Phantom oscillations in principal component analysis. *Proceedings of the National Academy of Sciences* **120**:e2311420120
- Stevens S. S. (1957) On the psychophysical law. *Psychological Review* **64**:153-81
- Subramanian D. L., Smith D. M. (2024) Time cells in the retrosplenial cortex. *bioRxiv*
- Tank D., Hopfield J. (1987) Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences* **84**:1896-1900
- 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**:984-998.e9
- Tenenbaum J. B., de Silva V., Langford J. C. (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* **290**:2319-23 <https://doi.org/10.1126/science.290.5500.2319>
- Terada S., Sakurai Y., Nakahara H., Fujisawa S. (2017) Temporal and rate coding for discrete event sequences in the hippocampus. *Neuron* **94**:1-15
- Tiganj Z., Cromer J. A., Roy J. E., Miller E. K., Howard M. W. (2018) Compressed timeline of recent experience in monkey IPFC. *Journal of Cognitive Neuroscience* **30**:935-950
- Tiganj Z., Hasselmo M. E., Howard M. W. (2015) A simple biophysically plausible model for long time constants in single neurons. *Hippocampus* **25**:27-37
- 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 Maaten L., Hinton G. (2008) Visualizing data using t-SNE. *Journal of machine learning research* **9**
- Wallach A., Melanson A., Longtin A., Maler L. (2022) Mixed selectivity coding of sensory and motor social signals in the thalamus of a weakly electric fish. *Current Biology* **32**:51-63
- Ward G., Tan L., Grenfell-Essam R. (2010) Examining the relationship between free recall and immediate serial recall: the effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory and Cognition* **36**:1207-41 <https://doi.org/10.1037/a0020122>
- Wei X.-X., Stocker A. A. (2012) Artificial neural networks and machine learning-icann 2012.
- Weyl H. (1922) *Space-time-matter* Dutton.
- Zhang K. (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience* **16**:2112-26
- Zhang W., Wu Y. N., Wu S. (2022) Translation-equivariant representation in recurrent networks with a continuous manifold of attractors. *Advances in Neural Information Processing Systems* **35**:15770-15783
- Zuo S., Wang C., Wang L., Jin Z., Kusunoki M., Kwok S. C. (2024) Neural signatures for temporal-order memory in the medial posterior parietal cortex. *bioRxiv* <https://doi.org/10.1101/2023.08.17.553665>

Peer reviews

Joint Public Review:

Quite obviously, the brain encodes "time", as we are able to tell if something happened before or after something else. How this is done, however, remains essentially not understood. In the context of Working Memory tasks, many experiments have shown that the neural activity during the retention period "encodes" time, besides the stimulus to be remembered; that is,

the time elapsed from stimulus presentation can be reliably inferred from the recordings, even if time per se is not important for the task. This implies 'mixed selectivity', in the weak sense of neural activity varying with both stimulus identity and time elapsed (since presentation).

In this paper, the authors investigate the implications of a specific form of such mixed selectivity, that is, conjunctive coding of what (stimulus) and when (time) at the single-neuron level, on the resulting dynamics of the population activity when 'viewed' through linear dimensionality-reduction techniques, essentially Principal Component Analysis (PCA). The theoretical/modeling results presented provide a useful guide to the interpretation of the experimental results; in particular, with respect to what can, or cannot, be rightfully inferred from those experimental results (using PCA-like techniques). The results are essentially theoretical in nature; there are, however, some conclusions that require a more precise justification, in my opinion. More generally, as the authors themselves discuss in the paper, it is not clear how to generalize this coding scheme to more complicated, but behaviorally and cognitively relevant, situations, such as multi-item WM or WM for sequences.

(1) It is unclear to me how the conjunctive code that the authors use (i.e., Equation (3)) is constrained by the theoretical desiderata (i.e., compositionality) they list, or whether it is simply an ansatz, partly motivated by theoretical considerations and experimental observations.

The "what" part: What the authors mean by "relationships" between stimuli is never clearly defined. From their argument (and from Figure 1b), it would seem that what they mean is "angles" between population vectors for all pairs of stimuli. If this is so, then the effect of the passing time can only amount to a uniform rescaling of the components of the population vector (i.e., it must be a similarity transformation; rotations are excluded, if the linear-decoder vectors are to be time-independent); the scaling factor, then, must be a strictly monotonous function of time (increasing or decreasing), if one is to decode time. In other words, the "when" receptive fields must be the same for all neurons.

The "when" part: The condition, $\tau_3 = \tau_1 + \tau_2$, does not appear to be used at all. In fact, it is unclear (to me at least) whether the model, as it is formulated, is able to represent time intervals between stimuli.

(2) For the specific case considered, i.e., conjunctive coding, it would seem that one should be able to analytically work out the demixed PCA (see Kobak et al., 2016). More generally, it seems interesting to compare the results of the PCA and the demixed PCA in this specific case, even just using synthetic data.

(3) In the Section "Dimensionality of neural trajectories...", there is some claim about how the dimensionality of the population activity goes up with the observation window T , backed up by numerical results that somehow mimic the results of Cueva et al. (2020) on experimental data. Is this a result that can be formally derived? Related to this point, it would be useful to provide a little more justification for Equation (17). Naively, one would think that the correlation matrix of the temporal component is always full-rank nominally, but that one can get excellent low-rank approximations (depending on T , following your argument).

<https://doi.org/10.7554/eLife.108804.1.sa0>