PNAS RESEARCH ARTICLE | NEUROSCIENCE

Recall tempo of Hebbian sequences depends on the interplay of Hebbian kernel with tutor signal timing

Matthew Farrell[a,b,1] and Cengiz Pehlevan[a,b,c,2]

Edited by David Kleinfeld, University of California at San Diego, La Jolla, CA; received June 12, 2023; accepted June 4, 2024, by Editorial Board Member Peter L. Strick

Understanding how neural circuits generate sequential activity is a longstanding challenge. While foundational theoretical models have shown how sequences can be stored as memories in neural networks with Hebbian plasticity rules, these models considered only a narrow range of Hebbian rules. Here, we introduce a model for arbitrary Hebbian plasticity rules, capturing the diversity of spike-timing-dependent synaptic plasticity seen in experiments, and show how the choice of these rules and of neural activity patterns influences sequence memory formation and retrieval. In particular, we derive a general theory that predicts the tempo of sequence replay. This theory lays a foundation for explaining how cortical tutor signals might give rise to motor actions that eventually become "automatic." Our theory also captures the impact of changing the tempo of the tutor signal. Beyond shedding light on biological circuits, this theory has relevance in artificial intelligence by laying a foundation for frameworks whereby slow and computationally expensive deliberation can be stored as memories and eventually replaced by inexpensive recall.

Hebbian learning | sequences | spike-timing-dependent plasticity | motor learning

Significance

Sequences are a cornerstone of animal behavior. A population of neurons can store the memory of a sequence via Hebbian learning combined with a tutor signal. However, until now it is unknown how the diversity of Hebbian learning rules seen in the brain impacts the dynamics of the stored sequence. Here, we show how the shape of the Hebbian kernel interacts with the tempo of the tutor signal to form a sequence memory with its own distinct tempo and give conditions for noise robustness. This sheds light on how sequences of varying intrinsic tempo might be robustly stored in neural circuits with biologically realistic learning rules.

An important class of animal behaviors are those that are consolidated into "automatic," well-practiced sequential routines (1–9). An analogy is the process of learning a tennis serve, which starts as a slow and deliberate process, but eventually becomes nearly muscle memory. Importantly, the behavior can be thought of as a sequence that, once initialized in a starting state, progresses in a highly stereotyped, automatic fashion. From this perspective, the behavior can be thought of as a stored sequence memory that is recalled by an initial prompt.

For such sequence memories, the timing of the subelements of these sequences and overall tempo are essential components (10–16). For example, activities such as walking, swimming, and digestion follow a natural rhythm and tempo (17). Tempo-sensitive sequential activities also include those learned through experience, such as learning a sequence of precisely timed lever taps (18, 19). It is therefore important to understand how temporally specific sequences can be learned, memorized, and recalled.

To answer this question, we need to look at the mechanisms of learning and generation of sequential behavior. One common perspective holds that these behaviors are generated by stereotyped sequential neural activity (17, 20–23). Further, several experiments showed a tight correlation between neural activity and timing of behavior (13, 24–31). This suggests that automatic sequential behavior may at least in part be "stored" in the synapses of a neural circuit, such that the behavior can be generated by setting the network to a state that corresponds to the first point in the sequence.

Despite the importance of these topics to understanding brain function, it is unclear what actually determines the tempo of sequence generation in neural circuits, and how this is connected to the learning process. Here, we hypothesize that the tempo arises from an interaction between the temporal dependence of the synaptic learning rules and the temporal structure of the network activity when learning the sequential behavior. We refer to the mechanism that sets network activity during sequence learning as the "tutor signal;" this tutor signal may come from a higher-order brain area such as the motor cortex or from minimally processed sensory inputs. The idea of a tutor signal being the basis for forming long-term memories is a standard theory that has been posited by experimentalists studying birdsong learning and motor learning in rats (18, 32, 33).

This perspective is embodied by models featuring temporally asymmetric Hebbian (TAH) learning rules. Despite the long and illustrious legacy of this class of models (34–44), previous studies were seriously limited by the narrow class of learning rules that these preexisting models can describe, and the fact that these learning rules do not correspond with those observed in the brain. In this work, we extend the basic TAH

Author affiliations: [a]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; [b]Center for Brain Science, Harvard University, Cambridge, MA 02138; and [c]Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, MA 02138

Author contributions: M.F. and C.P. designed research; M.F. performed research; M.F. analyzed data; and M.F. and C.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. D.K. is a guest editor invited by the Editorial Board.

[1]Present address: Laboratory for Neural Computation and Adaptation, RIKEN Center for Brain Science, Wako, Saitama 351-0106, Japan.

[2]To whom correspondence may be addressed. Email: cpehlevan@seas.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2309876121/-/DCSupplemental.

Published July 30, 2024.

models previously considered to a much richer class of models. Not only do we derive a theory that describes the essential elements of sequence dynamics—including tempo and noise robustness—for learning rules currently seen in the brain; our theory describes a wide class of TAH learning rules that gives rise to sequential dynamics, covering the case of new learning rules that may be discovered in biology in the future.

TAH network models are built upon Hebbian learning rules, wherein the network state is set by the tutor to a sequence of activity patterns one after another. These activity patterns correspond to moments in behavioral sequences, and the Hebbian learning rules store the structure of these sequences in the network as a memory. The sequences can then be recalled by setting the network to the first state in a desired sequence, resulting in the execution of the now "automatic," well-practiced sequential behavior. These models can be seen as a generalization of Hopfield networks (34, 45) to storing sequences rather than static patterns; therefore, many of the motivations and rationale behind the Hopfield model are applicable here, such as the ability to "complete" a sequence based on partial inputs (see *Discussion* for more in-depth interpretations of our model). Our model is a straightforward extension of previous models [most closely, (40)] with the addition of general Hebbian learning rules and added white noise to the dynamics, which can be considered a minimalistic model to capture the phenomenon of sequence learning under general spike-timing-dependent plasticity (STDP) learning rules and noise perturbation.

While our theory is very general, we demonstrate its applicability and utility by focusing on two fundamental and biologically relevant cases: 1) TAH learning rules with fast timescales relative to the tutor signal and 2) double-sided exponential TAH learning rules as are commonly seen in biology (46, 47). In both cases, we find useful principles that are interpretable in the context of biology. For example, weight structures (and TAH learning rules) can be adjusted such that the tempo remains the same, but robustness of sequence recall to noise increases. We also show how networks with the same qualitative dynamical properties can arise from very different weight structures. For double-sided exponential TAH kernels, we show that the tempo of sequence recall depends linearly on the tempo of the tutor signal, with a slope that depends in a simple way on the parameters of the kernel. While we focus here on applications to neuroscience, considering its generality our analysis may also be of interest to theorists in other domains studying dynamical systems with asymmetric weight structures.

Results

Sequence Model Definition. The sequence is stored and produced by a recurrently connected network of N neurons. The connectivity is given by a recurrent weight matrix W, where W_{ij} is the connectivity strength of neuron j to neuron i. The dynamics of the network are then given by

$$\tau \dot{\boldsymbol{r}} = -\boldsymbol{r} + \boldsymbol{\phi}\left(W\boldsymbol{r} + \boldsymbol{\eta}(t)\right), \qquad [1]$$

where $\boldsymbol{\eta}(t)$ is white noise ($\langle \eta_i(t) \rangle = 0$ and $\langle \eta_i(t)\eta_i(t') \rangle = \tau\rho^2\delta(t-t')$) and $\boldsymbol{\phi}$ is a sigmoidal nonlinearity (see *SI Appendix*, 1 for more details). The dynamical variables r_i represent the firing rate of neuron i. Note that the effect of τ is to rescale the timescale of the dynamics, $\boldsymbol{r}(t;\tau = a) = \boldsymbol{r}(t/a;\tau = 1)$. Considering this, we set $\tau = 1$ without loss of generality. We will discuss the role of general τ again in the *Exponential kernels* subsection of *Results*. The noise term $\boldsymbol{\eta}(t)$ models the effects of noisy biological processes such as synaptic transmission failure or possibly from interference arising from the presence of multiple stored sequences (40).

The sequence is stored in this network by a tutor signal (Fig. 1*A*). Given a sequence of length P, this signal sets the state of the network sequentially to an ordered set of P patterns $(\boldsymbol{\xi}^\mu)_{\mu=1}^P$. These P patterns represent P neural states that correspond to the states making up the sequential behavior. The recurrent network memorizes these patterns by synaptic plasticity at the recurrent synapses.

To store a memory of this sequence, this synaptic plasticity can take the form of a TAH learning rule active during application of the tutor signal (34, 35, 40, 48). An example of a TAH learning rule commonly seen in the literature (i.e., the previous set of references) is

$$W_{ij} \leftarrow \frac{1}{N\Delta t}\int_0^{T-\Delta t} dt\, r_i(t+\Delta t) r_j(t), \qquad [2]$$

where Δt is an interval of time that spans the amount of time the network is placed in each pattern state and $T = P\Delta t$ is the total duration for the tutor signal associated with the sequence. With this learning rule, connections are strengthened according to the coincident firing of neurons for the current pattern with those of the next pattern; if weights are initially zero, this learning rule results in final weights

$$W = \frac{1}{N}\sum_{\mu=1}^{P-1} \boldsymbol{\xi}^{\mu+1}(\boldsymbol{\xi}^\mu)^\top. \qquad [3]$$

We refer to this weight structure as sequi-associative. Intuitively, this set of synaptic weights calculates the overlap between the current network state with the stored patterns, and if the overlap with one of the patterns dominates over others (say pattern ν), it steers the network activity to the pattern that is after the dominant pattern (pattern $\nu + 1$).

While sequi-associative weights capture some temporal aspects of synaptic plasticity, they fall short in their ability to express the full complexity of plasticity dynamics observed in experiments (46, 47, 49, 50) which can significantly affect synaptic structure and network dynamics (51–53). Mathematically, a more general Hebbian learning rule could be written as

$$W_{ij} \leftarrow \frac{1}{N}\int_0^T dt \int_0^T ds\, w(s-t) r_i(s) r_j(t), \qquad [4]$$

where w is a general kernel/filter that weights the Hebbian learning rule as a function of time offset. One sees that Eq. **2** corresponds to $w(s-t) = \delta(s-t-\Delta t)/\Delta t$. However, biologically observed kernels take much richer and different forms, including the famous double-sided exponential STDP curve of Bi and Poo (46, 47) (Fig. 1 *B*, *Right*). Note that the neuroscience literature often plots the kernel as a function of $-\delta t = -(s-t)$ instead of $\delta t = s - t$ (i.e., "pre–post" instead of the "post–pre" convention used here).

In this paper, we consider the learning of sequences under a tutor signal with the general Hebbian rule (Eq. **4**). This rule can give rise to a much richer set of weights defined by

$$W = \frac{1}{N}\sum_{\mu=1}^{P}\sum_{\nu=1}^{P} a_\mu^\nu \boldsymbol{\xi}^\nu(\boldsymbol{\xi}^\mu)^\top, \qquad [5]$$

where a_μ^ν are coefficients that satisfy

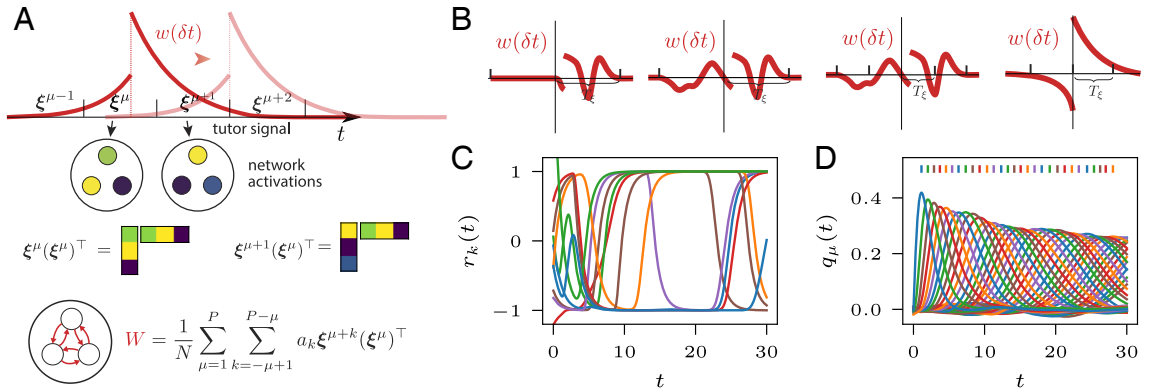

Fig. 1. Illustration of network model and stereotypical behavior. (*A*) The interaction of a tutor signal (the ξ^μ) with Hebbian temporal kernel $w(\delta t) = w(s-t)$ gives rise to weight matrix W. The kernel is convolved with the network state and the result is integrated, where the state of the network is set by the tutor signal. (*B*) Illustrations of different possible kernels, emphasizing the largely arbitrary shapes that they can take in our theory. *Left*: kernel that would give rise to two terms for each μ, $a_0\xi^\mu(\xi^\mu)^\top$ and $a_1\xi^{\mu+1}(\xi^\mu)^\top$. Second from *Left*: kernel that would give rise to three terms for each μ, $a_0\xi^\mu(\xi^\mu)^\top$, $a_1\xi^{\mu+1}(\xi^\mu)^\top$ and $a_{-1}\xi^{\mu-1}(\xi^\mu)^\top$. Third from *Left*: kernel that would give rise to five terms for each μ, corresponding to indices -2, -1, 0, 1, and 2. *Right*: double-sided exponential kernel. (*C*) Activity r_k of ten randomly selected units in the network for coefficient values $a_1 = 1$ and $a_k = 0$ for $k \neq 1$. Here, the network state is initialized to the first pattern in the sequence, $\boldsymbol{r}(0) = \boldsymbol{\xi}^1$, and $\rho = 0$. (*D*) Overlaps q_μ of the network activity with patterns $\boldsymbol{\xi}^\mu$ (coefficients a_k and ρ as in *C*). Color corresponds to pattern index μ and starts at $\mu = 2$. The vertical lines indicate the locations t_μ of the peaks. The first overlap q_1 is not included in the plot for ease of visualization (it is an exponential that decays from 1).

$$a_\mu^\nu = \int_0^{T_\xi^\mu} dt \int_{t_\xi^\nu - t_\xi^\mu}^{t_\xi^{\nu+1} - t_\xi^\mu} ds\, w(s-t). \qquad [6]$$

Here, t_ξ^ν is the time at which the tutor signal for pattern ν begins and $T_\xi^\mu = t_\xi^{\mu+1} - t_\xi^\mu$ is the duration for which pattern μ is shown to the network. Our goal is to understand the sequential dynamics and its tempo arising from such learning rules. Note that Eqs. **2** and **3** use kernels that are temporally asymmetric, which is essential for sequential dynamics to arise, whereas Eqs. **4** and **5** do not specify this explicitly. Indeed, not all kernels used in Eq. **4** give rise to forward-propagating, self-sustaining sequential dynamics. *SI Appendix*, 2 details certain conditions for this to occur.

While in general the relationship between the coefficients and the kernel is complex as indicated in Eq. **6**, in special cases the relationship can be described intuitively. One example is the case of double-sided exponential kernels. See the final subsection of *SI Appendix*, 2 for a description of this relationship (*SI Appendix*, Eqs. **A.30**–**A.32**).

To quantify the sequential behavior of the network dynamics, we define the overlaps

$$q_\mu(t) := \frac{1}{N}(\boldsymbol{\xi}^\mu)^\top \boldsymbol{r}(t), \qquad [7]$$

which measure the similarity of the recurrent network activity $\boldsymbol{r}$ at time t to stored patterns $\boldsymbol{\xi}^\mu$. For our initial condition, we will always initialize the network state to the first pattern $\boldsymbol{\xi}^1$. Note that the dynamics of the network in the basis of the neurons may not reveal the sequential structure (Fig. 1*C*), while the overlaps q_μ clearly display this structure (Fig. 1*D*). To measure sequence progression, we track the locations in time t_μ of the maxima of the overlaps $q_\mu(t)$; at time t_μ we say the sequence is in state μ (see vertical lines in Fig. 1*D*).* We also define $p_\mu := q_\mu(t_\mu)$ to be the height of the peaks. We are particularly interested in the tempo of the recalled sequences, which we measure by the peak difference $d_\mu := t_\mu - t_{\mu-1}$. Note that by this definition tempo is the inverse of sequence progression speed. As we will show, the tempo of the recalled sequence does not necessarily match the tempo of the tutor signal (54).

Note that the coefficients a_ν^μ can inhomogeneously depend on μ, resulting in a tempo of sequence recall that can be controlled based on the pattern state (i.e., the same sequence can have both fast and slow transitions between patterns, resulting in a d_μ that depends strongly on μ). If the tutor signal presents patterns at regular intervals, then the coefficients take the form $a_\mu^\nu = a_{\nu-\mu}$. In this case, the coefficient for the $\xi^\nu\xi^\mu$ term only depends on the difference between ν and μ rather than depending on μ independently of ν:

$$W = \frac{1}{N}\sum_{\mu=1}^{P}\sum_{\nu=1}^{P} a_{\nu-\mu}\boldsymbol{\xi}^\nu(\boldsymbol{\xi}^\mu)^\top = \frac{1}{N}\sum_{\mu=1}^{P}\sum_{k=-\mu+1}^{P-\mu} a_k\boldsymbol{\xi}^{\mu+k}(\boldsymbol{\xi}^\mu)^\top. \qquad [8]$$

This means that the tempo d_μ can no longer be controlled in a sequence-state-dependent (i.e., μ-dependent) manner. Even stronger, often the tempo is approximately independent of μ (it is a global property of the sequence), so that d_μ is approximately constant across μ.

Sequence Stability. Beside tempo, we are also interested in quantifying when sequential dynamics die out over time, or are instead (relatively) self-sustaining. Note that since we are using a saturating nonlinearity ϕ, each r_i is bounded which implies that the overlaps q_μ are also bounded. Hence we are primarily interested in how the solutions $q_\mu(t)$ decay with increasing μ. We observe that solutions generically decay to the zero solution $q_\mu(t) \equiv 0$ with increasing μ but there is a bifurcation point with decreasing magnitude of connection strength $\sum_k a_k$. Before the bifurcation point (large $\sum_k a_k$), the decay of $q_\mu(t)$ is subexponential, such as $\sim 1/\sqrt{2\pi\mu}$ (*SI Appendix*, 2). After the bifurcation point (small $\sum_k a_k$), the decay is exponential. Since the peak times t_μ are seen in our theory and simulations to be a linear function of time t, we can measure this decay of

*We can also define the sequence state represented at a time t by $\arg\max_\mu(q_\mu(t))$, but the t_μ are more mathematically convenient objects to consider.

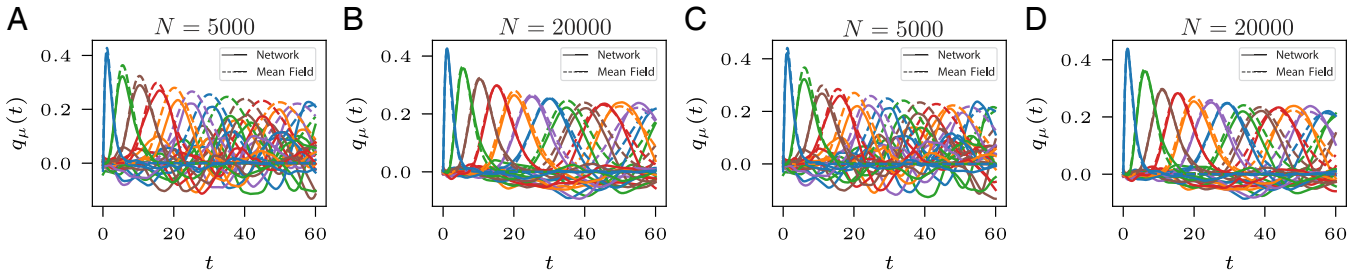

Fig. 2. Comparison of full network simulations Eq. **1** (solid lines) with the mean-field equations Eq. **9** (dashed lines). Throughout, every fifth overlap $q_\mu(t)$ is shown, starting with $\mu = 2$. For all plots, $\rho = 0$. (*A*) Overlaps $q_\mu(t)$ for network with $N = 5{,}000$ neurons. The only nonzero coefficients are $a_\mu^{\mu+1} = 1.5$. (*B*) As in (*A*), but for $N = 20{,}000$ neurons. (*C*) As in (*A*), but $a_\mu^\mu = 0.1$ and $a_\mu^{\mu+1} = 1.3$ for $\mu < 11$; and $a_\mu^\mu = -0.1$ and $a_\mu^{\mu+1} = 1.5$ for $\mu \geq 11$. All other coefficients are zero. (*D*) As in (*C*) but for $N = 20{,}000$.

the solutions $q_\mu(t)$ by measuring the "sup-norm" of the solution vector $\boldsymbol{q}(t) := [q_1(t), q_2(t), ..., q_P(t)]^\top$ while taking $P \to \infty$, where the sup-norm $\|\boldsymbol{v}\|_\infty$ of a vector $\boldsymbol{v}$ is $\max_i |v_i|$. Precisely, we say that the sequence of solutions $(q_1(t), q_2(t), ..., q_P(t))$ is unstable if the solution vector $\boldsymbol{q}(t)$ has the property that $\liminf_{P\to\infty} \|\boldsymbol{q}(t)\|_\infty$ decays to 0 with rate $O(t^b e^{at})$ for some $a < 0$ and some real number b. We say that the sequence of solutions is stable if it is not unstable.[†]

Related Works. Other works have also considered extensions to Eq. **3**. For example, some works (35–37, 54, 55) include an autoassociative component $\boldsymbol{\xi}^\mu(\boldsymbol{\xi}^\mu)^\top$ to the weights; Refs. 35–37 use discrete dynamics that selectively "delay" filter the sequi-associative component, and ref. 55 multiplies the sequi-associative term with random noise. The authors of ref. 43 conduct simulation studies of a model with an antisymmetric Hebbian kernel. Ref. 56 considers terms of the form $\xi^{\mu+k}(\xi^\mu)^\top$, but only for a specific kernel rather than the generality considered here.

Mean-Field Theory. To gain analytical insight about the behavior of our model, we extend the mean-field analysis of ref. 40 to the general weights described by Eq. **5**. We assume that the patterns are drawn identically and independently from a standard Gaussian distribution, $\xi_i^\mu \sim \mathcal{N}(0, 1)$, and take the network population size to infinity, $N \to \infty$.

In this mean-field limit, the overlaps evolve with dynamics given by (*SI Appendix*, 1)

$$\dot{q}_\mu = -q_\mu + g(t) \sum_{\nu=1}^{P} a_\nu^\mu q_\nu, \qquad [9]$$

where

$$g(t) = G\left(\|A\boldsymbol{q}(t)\|_2^2 + \rho^2\right),$$

for some function G and where A is the matrix $A = (a_\nu^\mu)_{1\leq\mu,\nu\leq P}$ (μ indexes the rows and ν indexes the columns). We give the full functional form of $G(x)$ in *SI Appendix*, 1 and Eq. **A.3**), which depends on the shape of ϕ. For the analysis here, the important aspect of G is the value of $G(\rho^2)$, which serves as an upper bound on $g(t)$ (*SI Appendix*, Fig. S1). In particular, for $\rho = 0$, $g(t)$ is bounded from above by $G(0)$, and $G(0) = 20/\sqrt{2\pi} \approx 8$ for our choice of parameters for ϕ (see *SI Appendix*, 1). The form of this expression comes from the Gaussian patterns and the error-function form of ϕ (*SI Appendix*, Eq. **A.1**).

The convergence with increasing N of the mean-field solutions to the full network solutions is illustrated in Fig. 2. This holds both in the case of uniform tutor signal intervals (Fig. 2 *A* and *B*) as well as nonuniform tutor intervals (Fig. 2 *C* and *D*). For the following, we assume that tutor signal intervals are uniform in temporal duration, so $a_\nu^\mu = a_{\mu-\nu}$. We will revisit the case of nonuniform intervals in the final subsection of *Results* (*Sequences with fast and slow parts*). In the uniform interval case, T_ξ^μ is constant across μ so we drop the μ dependence and write T_ξ as the tutor signal interval duration.

In *SI Appendix*, 2, we show under appropriate conditions that the solution to Eq. **9** is approximated by the linear equations

$$\dot{q}_\mu = -q_\mu + \sum_{\nu=1}^{P} \bar{a}_{\mu-\nu} q_\nu, \qquad [10]$$

where $\bar{a}_\mu := \bar{g} a_\mu$ and where we define $\bar{g} := 1/\sum_{k=-P+1}^{P-1} a_k$. This is a consequence of $g(t)$ asymptoting to a constant value as $t \to \infty$; while we do not prove that this occurs, we observe it happening in our simulations. In *SI Appendix*, 2, we argue that $\lim_{t\to\infty} g(t) = \bar{g}$ holds under certain conditions when solutions are stable, and that with additional conditions $\lim_{t\to\infty} g(t) = \bar{g}$ implies that solutions are stable. Conversely, one case where we can expect solutions to be unstable is when $g(t)$ is bounded away from $\bar{g}$. This occurs when connectivity is too weak, i.e., when $\bar{g} > G(\rho^2)$.

One Forward Term. We next turn to analyzing sequence progression in the recurrent neural network. We start by considering the case with one forward term where $a_k = 0$ for $k \notin \{0, 1\}$. This results in the weight matrix $W = \frac{1}{N}\sum_\mu (a_0 \boldsymbol{\xi}^\mu \boldsymbol{\xi}^\mu + a_1 \boldsymbol{\xi}^{\mu+1} \boldsymbol{\xi}^\mu)$ with autoassociative and sequi-associative terms $a_0 \boldsymbol{\xi}^\mu \boldsymbol{\xi}^\mu$ and $a_1 \boldsymbol{\xi}^{\mu+1} \boldsymbol{\xi}^\mu$, respectively. This case is relevant in the regime where the duration of sequence elements, T_ξ, is longer than the timescale of $w(t)$ (i.e. $w(\delta t) = 0$ for $\delta t > T_\xi$ and $w(\delta t) = 0$ for $\delta t < 0$). Fig. 3*A* illustrates two possible kernels w corresponding to this case, emphasizing the arbitrary shape that the kernel can take within its support.

In this case Eq. **10** becomes

$$\dot{q}_\mu = \begin{cases} -q_\mu + \bar{a}_0 q_\mu + \bar{a}_1 q_{\mu-1}, & \mu > 1, \\ -q_\mu + \bar{a}_0 q_\mu, & \mu = 1. \end{cases}$$

[†]Note the contrast to the dynamical systems literature, where stability typically means that the solutions remain finite.

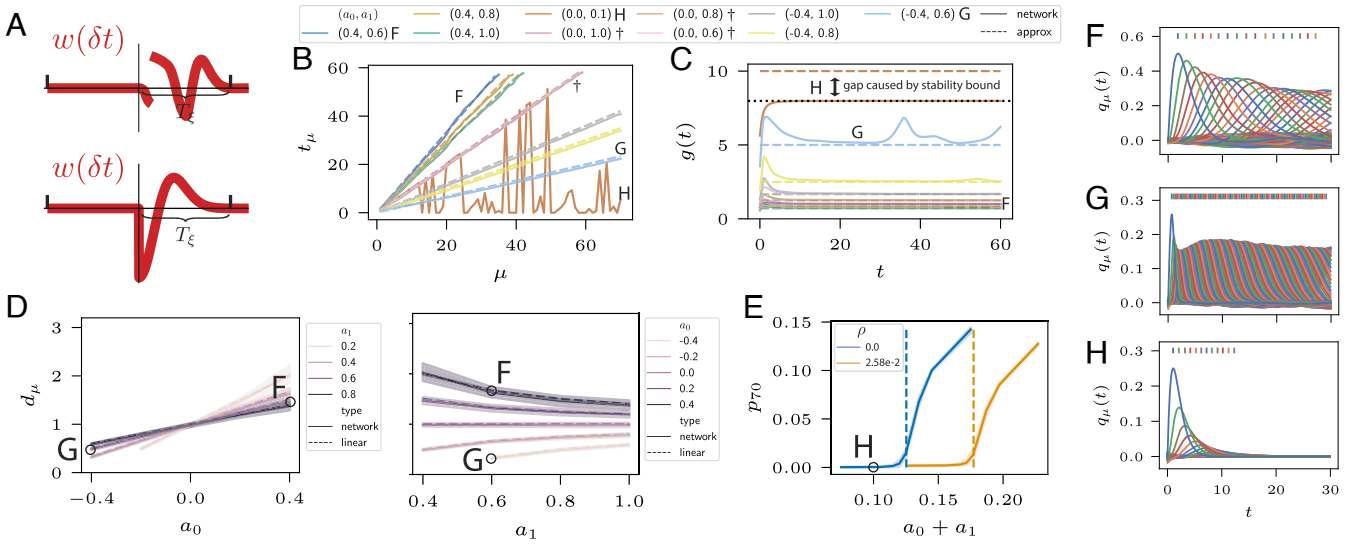

Fig. 3. Sequence progression for two-term connectivity corresponding to $a_k = 0$ for $k \notin \{0, 1\}$. Plots show the full network simulations Eq. **1** (solid lines) and linear approximations Eq. **11** (dashed lines). Throughout, letters F, G, and H correspond to coefficient values $(a_0, a_1) = (0.4, 0.6)$, $(-0.4, 0.6)$, and $(0.0, 0.1)$, respectively. For all plots except for panel (*E*), $\rho = 0$. (*A*) Illustrations of two TAH kernels w that would give rise to two terms. (*B*) Peak times t_μ. The † symbol denotes coefficient values that give rise to overlapping lines, and H denotes a coefficient values for which the network equations are unstable. (*C*) The value of $g(t)$ through time during the simulations in (*B*). Dashed lines denote $\bar{g}$. The function $g(t)$ is bounded from above by $G(0) \approx 8$, which is indicated by a dotted line. For the coefficient values corresponding to H, there is a mismatch between $\bar{g}$ and $\lim_{t\to\infty} g(t)$ (asymptote of solid line). (*D*) Plots of peak differences d_μ as a function of a_0 and a_1. Shaded regions are 95% CIs for $\mu \in \{3, \ldots, 72\}$, and lines are the means. *Left*: a_0 is plotted on the horizontal axis and a_1 is denoted by color. *Right*: a_1 is plotted on the horizontal axis and a_0 is denoted by color. (*E*) The peak height p_{70} corresponding to the maximum of $q_{70}(t)$ over t, in the full network simulation, as a function of the sum of coefficients $a_0 + a_1$. Color denotes the noise strength ρ and dashed vertical lines show the critical points $G(\rho^2)^{-1}$, where the network is predicted to pass from stable to unstable. Shaded regions are 95% CIs over $a_0 \in \{-0.2, 0, 0.1\}$. (F) Overlaps $q_\mu(t)$ corresponding to coefficient values $(0.4, 0.6)$. (G) Overlaps $q_\mu(t)$ corresponding to coefficient values $(-0.4, 0.6)$. (H) Overlaps $q_\mu(t)$ corresponding to coefficient values $(0.0, 0.1)$.

The solutions to these equations for $\bar{a}_1 > 0$ are (recall the initial conditions $q_1(0) = 1$ and $q_\mu(0) = 0$ for $\mu > 1$):

$$q_\mu(t) = \frac{(\bar{a}_1)^{\mu-1} t^{\mu-1}}{(\mu-1)!} e^{(\bar{a}_0 - 1)t},$$

which have maxima at

$$t_\mu = \frac{\mu - 1}{1 - \bar{a}_0}. \qquad [11]$$

Note that $\bar{a}_0 + \bar{a}_1 = 1$ by definition, which places some restrictions on the values that $\bar{a}_0$ and $\bar{a}_1$ can take. For instance, $\bar{a}_0 > 1$ implies $\bar{a}_1 < 1$. The tempo as measured by peak difference is

$$d_\mu = \frac{1}{1 - \bar{a}_0} = 1 + \frac{a_0}{a_1}. \qquad [12]$$

Peak magnitudes p_μ can also be computed explicitly (*SI Appendix*, 2 and Eqs. **A.20** and **A.21**) and are asymptotically

$$p_\mu \sim 1/\sqrt{2\pi\mu},$$

when solutions are stable.

Fig. 3 shows the simulation results for various values of a_0, a_1, and ρ. Fig. 3*B* plots peak times t_μ for $\rho = 0$. The full network simulations with $N = 35{,}000$ match the linear approximations Eq. **11** closely, except for coefficient values where the solutions for the full network decay to zero (denoted by H). This happens because $\bar{g}$ is larger than the critical value of $G(0) \approx 8$. Note the significant role of a_0 in determining the sequence tempo. In particular, for $a_0 = 0$ changing a_1 has no impact on the tempo († symbol in Fig. 3*B*). Note also that, unlike in standard Hopfield models, the autoassociative term a_0 can be negative. The *Bottom* panel of Fig. 3*A* shows an example of a kernel that gives rise to negative a_0 and positive a_1.

Fig. 3*C* plots the function $g(t)$. This plot supports our original assumption that $g(t)$ asymptotes to a constant value, and shows the close match between $\lim_{t\to\infty} g(t)$ and $\bar{g}$, except in the case where $\bar{g} > G(0) \approx 8$ (denoted by H), which results in instability.

To look more closely at the relative impacts of a_0 and a_1 on sequence tempo, we plot sequence tempo d_μ for different combinations of a_0 and a_1 in Fig. 3*D*. Fig. 3*D* shows that the dependence of d_μ on a_0 is roughly linear in the range plotted, with a slope that decreases with increasing a_1. Coefficient ranges were chosen in an attempt to capture the full behavior of the dynamics within the region of stable sequence generation; in particular, this requires that $a_1 > 0$ and $a_0 + a_1 > 1/G(0)$. Note that for $a_0 < 0$ sequence progression speed decreases, counterintuitively, with increasing a_1, while the opposite relationship holds for $a_0 > 0$ (recall that sequence progression speed is the inverse of the tempo d_μ). This can also be seen from Eq. **12**.

Fig. 3*E* quantifies the stability properties of sequences by plotting the peak magnitude p_μ for a pattern of large index (here, $\mu = 70$) as a function of $\bar{g}^{-1} = a_0 + a_1$ for two choices of noise level, $\rho = 0$ and $\rho = 2.58e - 2$. This figure indicates that increasing $\bar{g}^{-1}$ eventually transitions the network into a stable regime, while for a fixed $\bar{g}^{-1}$ increasing the noise ρ eventually destabilizes a stable network. Note that such noise could come from interference, such as that which arises if there are multiple stored sequences (40), or from biologically relevant sources such as synaptic transmission failure. The dashed vertical lines mark the critical values of stability where $\bar{g} = G(\rho^2)$. In summary, this plot illustrates how $\bar{g}$ determines the stability of sequences, and that smaller $\bar{g}$ (equivalently, larger $a_0 + a_1$) is required for stability in the presence of noise. The plot shows that, as predicted, peak

magnitudes p_μ have a sharp inflection point at $\bar{g} = G(\rho^2)$, where p_μ goes from being approximately zero to nonzero.

Note that a_0 and a_1 can be adjusted so that sequence tempo d_μ remains constant, but stability is improved by decreasing $\bar{g}$. This can most easily be seen by considering lines of constant d_μ in Fig. 3*D*. The most straightforward example is the line $d_\mu = 1$ in the *Right* panel of Fig. 3*D*, which clearly corresponds to many values of $a_0 + a_1$. This is particularly important in the presence of noise, as shown by Fig. 3*E*. This degree of freedom of the system makes some kernels $w(t)$ strictly better than others, even if they result in the same tempo.

General Approximate Solution. While we were able to explicitly solve Eq. **10** in the case of two nonzero terms in the preceding section, in general an explicit solution is not available and we need to resort to approximations. To approximately solve Eq. **10**, we replace the system of equations with one that is periodic in μ. A general approximate solution to the equations with these periodic boundary conditions can then be found (*SI Appendix*, 2):

$$t_\mu \approx \frac{\mu - 1}{\alpha} - \frac{\beta}{2\alpha^2}, \quad \mu > 1, \qquad [13]$$

where $\alpha = \sum_{k=1}^{\lceil P/2 \rceil - 1} (\bar{a}_k - \bar{a}_{-k})k$ and $\beta = \sum_{k=1}^{\lceil P/2 \rceil - 1} (\bar{a}_k + \bar{a}_{-k})k^2$. Sequence tempo as measured by peak difference is then

$$d_\mu \approx \frac{1}{\alpha}, \quad \mu > 2. \qquad [14]$$

When $\alpha > 0$, these equations describe forward propagation of the sequence.

In the following sections, we will look at special cases of interest that illustrate the theory.

Three Terms with Bidirectional Connectivity. Next we consider the case with three terms and bidirectional connectivity in Eq. **10**, where $a_k = 0$ for $k \notin \{-1, 0, 1\}$. This results in the weight matrix $W = \frac{1}{N} \sum_\mu \left(a_{-1} \boldsymbol{\xi}^{\mu-1} \boldsymbol{\xi}^\mu + a_0 \boldsymbol{\xi}^\mu \boldsymbol{\xi}^\mu + a_1 \boldsymbol{\xi}^{\mu+1} \boldsymbol{\xi}^\mu \right)$. Bidirectional connectivity occurs with TAH kernels commonly found in biology, such as double-sided exponential decay kernels (46, 47) (see Fig. 1 *B*, *Right* for an illustration). Three terms in particular are relevant when the TAH kernel's timescale is faster than the tutor signal's, for instance when the support of the kernel vanishes outside of time windows spanning more than two pattern presentations. Fig. 4*A* illustrates an arbitrary kernel that would give rise to three terms.

To understand sequence progression speed, we use Eq. **13**:

$$t_\mu \approx \frac{\mu - 1}{\bar{a}_1 - \bar{a}_{-1}} - \frac{\bar{a}_{-1} + \bar{a}_1}{2(\bar{a}_1 - \bar{a}_{-1})^2}, \quad \mu > 1. \qquad [15]$$

Sequence tempo as measured by peak difference is then

$$d_\mu \approx \frac{1}{\bar{a}_1 - \bar{a}_{-1}} = \frac{a_0 + a_1 + a_{-1}}{a_1 - a_{-1}}, \quad \mu > 2.$$

The effect of changing a_{-1} on sequence tempo is revealed by taking the derivative of d_μ with respect to a_{-1}:

$$\frac{d}{da_{-1}} d_\mu \approx \frac{a_0 + 2a_1}{(a_1 - a_{-1})^2}, \quad \mu > 2.$$

This shows that increasing a_{-1} increases d_μ provided $a_0 + 2a_1 > 0$. To compare the relative impact of changing a_{-1} and a_1 on sequence tempo, we can take the ratio of derivatives:

$$\frac{dd_\mu}{da_1} \Big/ \frac{dd_\mu}{da_{-1}} \approx -\frac{a_0 + 2a_{-1}}{a_0 + 2a_1}, \quad \mu > 2.$$

This equation reveals that changes in a_{-1}, counterintuitively, may impact sequence tempo more than changes in a_1 for certain coefficient configurations. Hence a_{-1} is also a natural parameter to use to control tempo.

Fig. 4 compares the full network simulations given by Eq. **1**, the simulations of the linear system Eq. **10**, and the approximation Eq. **15**. The peak times t_μ are plotted in Fig. 4*B*, showing that Eq. **15** is a good estimate of both the linear and full network system. Fig. 4*C* shows the relationship between d_μ, a_0, and a_1 for fixed $a_{-1} = -0.2$. The *Left* panel reveals

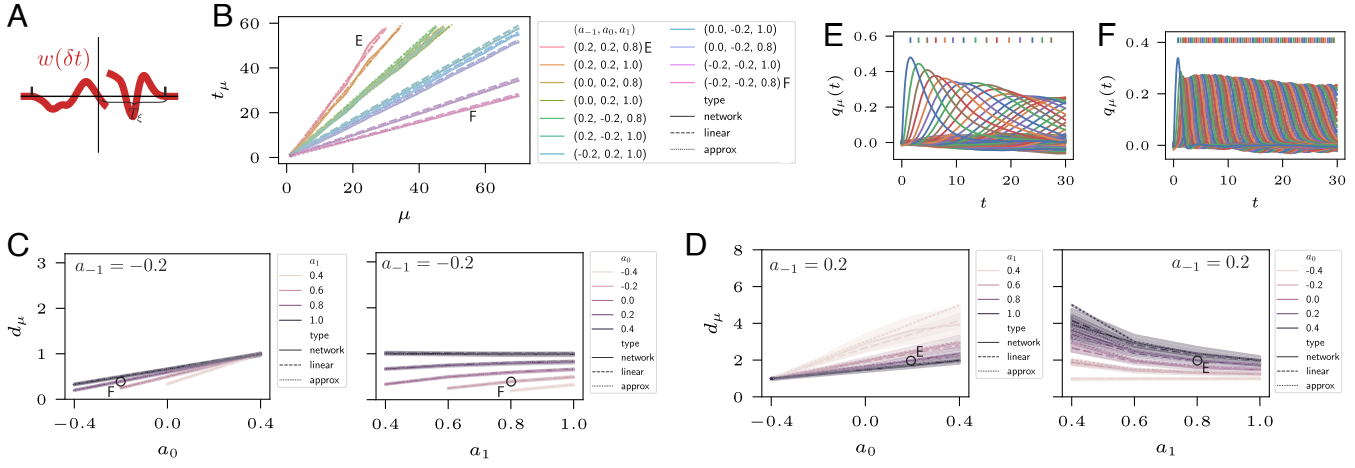

Fig. 4. Sequence progression for bidirectional terms. Plots show the full network simulations Eq. **1** (solid lines), linear approximations Eq. **10** (dashed lines), and approximation Eq. **15** (dotted lines). Throughout, letters E and F correspond to coefficient values $(a_{-1}, a_0, a_1) = (0.2, 0.2, 0.8)$ and $(-0.2, -0.2, 0.8)$, respectively. For all plots, $\rho = 0$. (*A*) Illustration of a TAH kernel w that would give rise to bidirectional terms. (*B*) Peak times t_μ plotted for a variety of coefficient combinations (a_{-1}, a_0, a_1), denoted by color. (*C*) Plots of peak differences d_μ as a function of a_0 and a_1. Here, $a_{-1} = -0.2$. Shaded regions are 95% CIs for $\mu \in \{3, \ldots, 72\}$, and lines are the means. *Left*: a_0 is plotted on the horizontal axis and a_1 is denoted by color. *Right*: a_1 is plotted on the horizontal axis and a_0 is denoted by color. (*D*) As in (*C*) but with $a_{-1} = 0.2$. (E) Overlaps $q_\mu(t)$ corresponding to coefficient values (0.2, 0.2, 0.8). (F) Overlaps $q_\mu(t)$ corresponding to coefficient values $(-0.2, -0.2, 0.8)$.

that d_μ increases approximately linearly with increasing a_0, with a slope that decreases with increasing a_1. Comparison with Fig. 3*D* shows the impact of introducing a negative a_{-1}, which can be mixed but is generally to decrease d_μ over the parameter regime considered.

Fig. 4*D* shows the relationship between d_μ, a_0, and a_1 for fixed positive $a_{-1} = 0.2$. Again, the *Left* panel reveals that d_μ increases approximately linearly with increasing a_0, with a slope that decreases with a_1. Comparing with Fig. 3*D* shows how $a_{-1} = 0.2$ typically increases d_μ relative to $a_{-1} = 0$.

As in the case with two terms, the tempo of the sequence can be held constant while the stability term $\bar{g}$ is decreased. Indeed, the introduction of a_{-1} introduces an additional degree of freedom. Examples of keeping d_μ fixed while varying $\bar{g}$ can be seen by following horizontal lines of fixed d_μ in Fig. 4 *C* and *D*.

Three Terms with Forward Connectivity. Now we consider the case with two forward terms, where $a_k = 0$ for $k \notin \{0, 1, 2\}$. This results in the weight matrix $W = \frac{1}{N}\sum_\mu \left(a_0 \boldsymbol{\xi}^\mu \boldsymbol{\xi}^\mu + a_1 \boldsymbol{\xi}^{\mu+1} \boldsymbol{\xi}^\mu + a_2 \boldsymbol{\xi}^{\mu+2} \boldsymbol{\xi}^\mu\right)$. This illustrates the effects of having TAH kernels with a slightly longer timescale than the tutor signal. Fig. 5*A* illustrates a TAH kernel that would give rise to two forward terms. Using Eq. **13**:

$$t_\mu \approx \frac{\mu - 1}{\bar{a}_1 + 2\bar{a}_2} - \frac{\bar{a}_1 + 4\bar{a}_2}{2(\bar{a}_1 + 2\bar{a}_2)^2}, \quad \mu > 1. \qquad [16]$$

Sequence tempo as measured by peak difference is then

$$d_\mu \approx \frac{1}{\bar{a}_1 + 2\bar{a}_2} = \frac{a_0 + a_1 + a_2}{a_1 + 2a_2}, \quad \mu > 2. \qquad [17]$$

Fig. 5 compares the full network simulations given by Eq. **1**, the simulations of the linear system Eq. **10**, and the asymptotic approximation of the peak times Eq. **16**, showing a close match for all three quantities. Similar to the case of the previous section, increasing a_2 has a qualitatively different effect on d_μ depending on the values of a_0 and a_1. Taking the derivative of d_μ with respect to a_2 in Eq. **17**

$$\frac{d}{da_2} d_\mu \approx \frac{a_1 + 2a_0}{(a_1 + 2a_2)^2}, \quad \mu > 2 \qquad [18]$$

indicates that increasing a_2 increases d_μ if and only if $a_1 + 2a_0 < 0$. Examples of the effect of increasing a_2 can be seen in Fig. 5*B*, which plots peak times t_μ for a range of coefficient combinations. The differing effect of increasing a_2 can be seen in this plot.

The relationship between d_μ and different coefficient combinations is further elucidated in Fig. 5*C*, which plots d_μ as a function of a_0 and a_1, where $a_2 = 0.2$. As in previous sections, the dependence of d_μ on a_0 is approximately linear over the coefficient values considered, with a slope that decreases with increasing a_1. Comparison with Fig. 3*D* shows the (multiplexed) effect of positive a_2.

Note that, as in the previous sections, the stability properties of the sequence (determined by the sum of coefficients) can be improved even as the timing is held constant († symbol in Fig. 5*B*); that is, $d_\mu \approx 1/(\bar{a}_1 + 2\bar{a}_2)$ can be fixed while $\bar{g}$ is decreased. This is most clearly illustrated by the horizontal line $d_\mu = 1$ in the *Right* panel of Fig. 5*C*.

Exponential Kernels. Here, we consider the special case of (double-sided) exponential kernels as commonly used to model STDP kernels seen in biology (46, 47). These have the form $w(t) = -m_1 e^{t/\tau_1}$ for $t < 0$ and $w(t) = m_2 e^{-t/\tau_2}$ for $t \geq 0$, where $\tau_1, \tau_2 > 0$ (see Fig. 6*A* for an illustration). We compute in *SI Appendix*, 2 that the approximate peak times are given by Eq. **13** where

$$\boldsymbol{\alpha} = \bar{g}(m_1\tau_1^2 + m_2\tau_2^2)$$

and

$$\bar{g} = (m_2\tau_2 - m_1\tau_1)^{-1}\, T_\xi^{-1}$$

in the $P \to \infty$ limit. Here, T_ξ is the length of the tutor signal interval. Note in particular that the sequence tempo as measured by

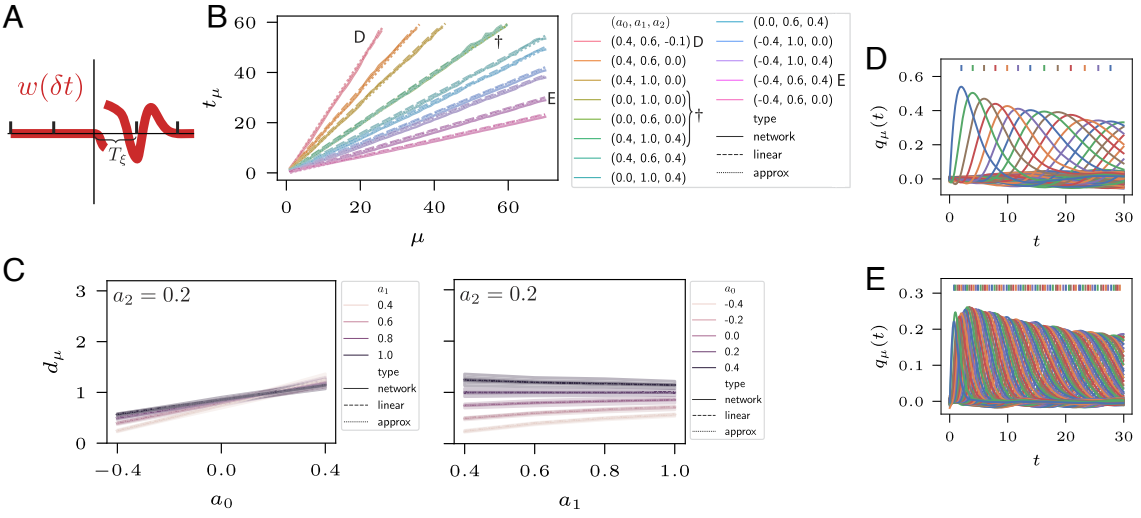

Fig. 5. Sequence progression for two forward terms. Plots show the full network simulations Eq. **1** (solid lines), linear approximations Eq. **10** (dashed lines), and approximation Eq. **16** (dotted lines). Throughout, letters D and E correspond to coefficient values $(a_0, a_1, a_2) = (0.4, 0.6, -0.1)$ and $(-0.4, 0.6, 0.4)$, respectively. For all plots, $\rho = 0$. (*A*) Illustration of a TAH kernel w that would give rise to two forward terms. (*B*) Peak times t_μ for a variety of coefficient combinations (a_0, a_1, a_2), denoted by color. The † symbol denotes coefficient values that give rise to overlapping lines. (*C*) Plots of peak differences d_μ as a function of a_0 and a_1. Here, $a_2 = 0.2$. Shaded regions are 95% CIs for $\mu \in \{3, \ldots, 72\}$, and lines are the means. *Left*: a_0 is plotted on the horizontal axis and a_1 is denoted by color. *Right*: a_1 is plotted on the horizontal axis and a_0 is denoted by color. (D) Overlaps $q_\mu(t)$ corresponding to coefficient values $(0.4, 0.6, -0.1)$. (E) Overlaps $q_\mu(t)$ corresponding to coefficient values $(-0.4, 0.6, 0.4)$.

$$d_\mu \approx \frac{1}{\alpha} = T_\xi \frac{m_2\tau_2 - m_1\tau_1}{m_1\tau_1^2 + m_2\tau_2^2}, \quad \mu > 2 \qquad [19]$$

scales linearly with the tutor signal interval T_ξ. The equation for β is given by *SI Appendix*, Eq. **A.33** in *SI Appendix*, 2.

Biology may use this relationship to dilate or contract the stored sequence tempo relative to the tutor signal by using particular STDP kernels. In particular, $\frac{m_2\tau_2 - m_1\tau_1}{m_1\tau_1^2 + m_2\tau_2^2} = 1$ indicates that the sequence will be stored faithfully (with the same tempo as the tutor signal). For general τ in Eq. **1**, the condition is $\frac{m_2\tau_2 - m_1\tau_1}{m_1\tau_1^2 + m_2\tau_2^2}\tau = 1$.

The quantity d_μ is plotted in Fig. 6 *B* and *C* for a variety of different kernels as a function of tutor interval T_ξ. These plots compare the full network simulations with the linear approximation Eq. **10** and the approximation Eq. **19**, showing a close match, especially for small T_ξ.

In Fig. 6*D*, we plot the dependence of sequence tempo d_μ on the parameters τ_1, τ_2, and m_2. The *Left* panel shows that d_μ typically decreases with increasing τ_2, but this can be reversed for large enough τ_1. The *Center* panel shows that, for the parameter values chosen, d_μ decreases with increasing τ_1. The *Right* panel shows the influence of changing m_2 on d_μ; while increasing m_2 increases d_μ, the effect is small. Parameter values were chosen to demonstrate the full spectrum of behaviors within the regime of stable, forward-propagating dynamics. *SI Appendix*, 2 describes simple relationships between the kernel parameters and coefficients a_k.

To look more closely at the interplay between the timescale of the TAH kernel and the timescale of the tutor signal, we look at the case where τ_1 is small relative to τ_2. Then $d_\mu \approx T_\xi/\tau_2$, indicating that in this case the ratio of the timescales sets the sequence tempo. Sequence stability is determined by $\bar{g} < G(\rho)$, where in this case $\bar{g} \approx m_2^{-1}\tau_2^{-1}T_\xi^{-1}$.

Sequences with Fast and Slow Parts. In naturalistic settings, different states of a sequence may have different durations. For instance, in playing a piece of music different notes are held with different durations. Our analysis indicates that changing these durations causes a change in the sequence progression speed, meaning that variable duration tutor signals T_ξ^μ will result in sequences with faster and slower parts (Fig. 7). Consider a sequence with sections that pass from one set of uniform durations $\{T_\xi^1\}_{\mu\in S_1}$ to another $\{T_\xi^2\}_{\mu\in S_2}$, illustrated by Fig. 7*A*. The overlaps q_μ for $\mu \in S_1$ will be governed by the first interval length, with some disturbance from q_ν for $\nu \in S_2$ mediated by the backward coefficients a_k for $k < 0$. For μ far from the boundary this disturbance should be small. On the other hand, q_ν far from the boundary will be governed by the second interval length with small disturbance from q_μ, with the caveat that the evolution up to time t has been influenced significantly by the dynamics of q_μ. With these caveats in mind, we should expect the tempo in each of these regions far from the boundary to progress at speeds determined by the corresponding tutor signal interval length, with more complex behavior occurring near the boundary of the transition. In Fig. 7, we consider a case with two transitions, where $T_\xi^1 = 0.6$ for $\mu \in [1, 20]$, $T_\xi^2 = 0.3$ for $\mu \in [21, 30]$, and $T_\xi^3 = 0.6$ for $\mu \in [31, 100]$ (slow to fast back to slow). Here, we use a double-sided exponential kernel with $\tau_1 = 0.25$, $m_1 = 2$, $\tau_2 = 1$, and $m_2 = 2$. This figure illustrates that after transitioning, the sequence speeds up or slows down to match the new duration.

However, these transitions have undesirable characteristics. For one, in the first transition, the sequence slows down just before entering the faster sequence section (Fig. 7 *B* and *C*). A second issue is that the peak differences become more variable after each transition (Fig. 7*C*). A third issue is that the transition from T_ξ^2 to T_ξ^3 is very gradual and takes a significant amount of time before recovering the tempo of the T_ξ^1 section (Fig. 7*C*). These issues, and particularly the third, are probably related to how overlaps q_μ generically "spread out" for increasing index μ (Fig. 7*D*); that is, the width of the bumps traced out by $q_\mu(t)$ increases with increasing μ. This suggests that an improved model represent overlaps in a way agnostic to the index of the pattern

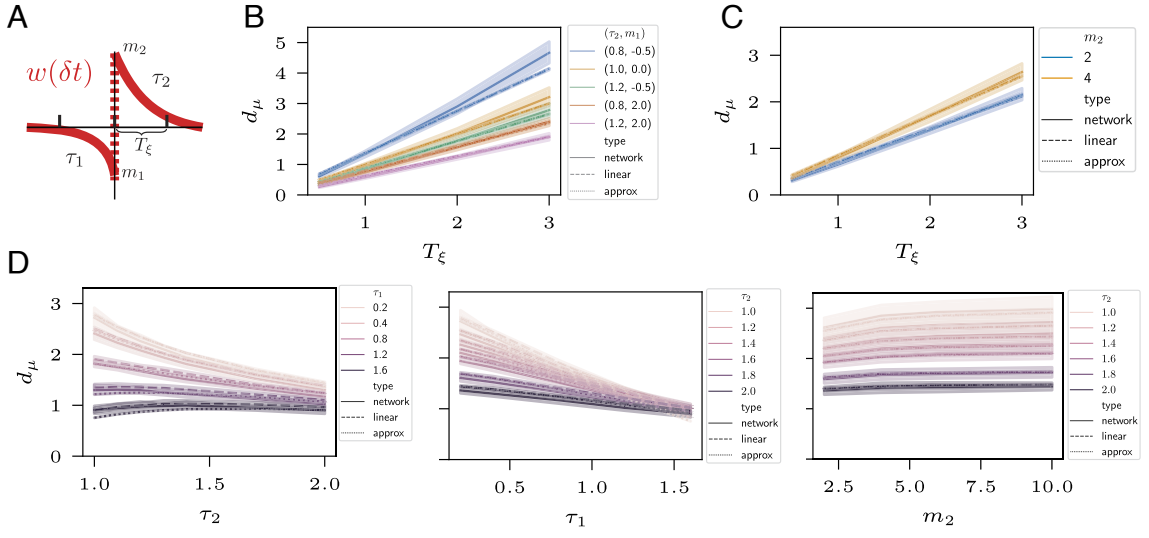

Fig. 6. Sequence progression for double-sided exponential kernels and varying T_ξ. Shaded regions are 95% CIs over $\mu \in \{3, \ldots, 12\}$, and lines are the means. Plots show peak differences d_μ derived from the full network simulations Eq. **1** (solid lines), linear approximations Eq. **10** (dashed lines), and approximation Eq. **19** (dotted lines). For all plots, $\rho = 0$. (*A*) Illustration of a double-sided exponential TAH kernel w. (*B*) Peak differences d_μ for a variety of double-sided exponential TAH kernel parameters τ_2 and m_1 (color). Here, $\tau_1 = 0.25$ and $m_2 = 2$. (*C*) As in (*B*), but with $m_1 = 2$ and varying m_2 (denoted by color). (*D*) Peak differences d_μ as a function of τ_1, τ_2, and m_2. Unless otherwise defined, parameter values are $\tau_1 = 0.25$, $m_1 = 1$, $m_2 = 3$, and $T_\xi = 3$. *Left*: d_μ as a function of τ_2 with varying τ_1 (color). *Center*: d_μ as a function of τ_1 with varying τ_2 (color). *Right*: d_μ as a function of m_2 with varying τ_2 (color).

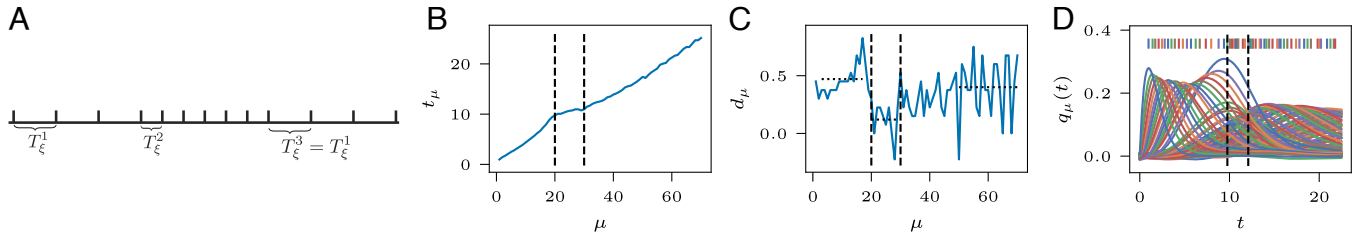

Fig. 7. Sequences with fast and slow parts. Sequence dynamics with three regions $T_\xi^1 = 0.6$, $T_\xi^2 = 0.3$, and $T_\xi^3 = 0.6$. The first transition occurs at $\mu = 20$ and the second at $\mu = 30$. Parameters are $\tau_1 = 0.25$, $m_1 = 2$, $\tau_2 = 1$, and $m_2 = 2$. For all plots, $\rho = 0$. (*A*) Illustration of the three regions (note that this schematic is not accurate with respect to the μ values at which transitions occur). (*B*) Peak times as a function of pattern index μ. Dashed vertical lines demarcate the three regions. (*C*) Peak difference d_μ as a function of pattern index μ. Dotted lines mark averages over the intervals [3, 18], [20, 28], and [50, 80]. (*D*) Plot of overlaps $q_\mu(t)$. Dashed vertical lines demarcate the three regions.

being represented, perhaps with the use of nonlinear learning rules as in refs. 40 and 54, "interaction modulation" as suggested by ref. 57, or "delay" filtering as suggested by refs. 35–37.

Discussion

The ability to store experiences as memories is fundamental to intelligent behavior. While memories are typically modeled as fixed points in recurrent networks, sequences constitute information that is also of primary importance. This is clear when considering the ubiquity of well-practiced, automatic motor behaviors.

Our model stores memories by assuming that the network is placed in states by a "tutor signal" during the learning process. There are multiple possible interpretations for this. The state of the network could be set by sensory inputs, so that TAH learning rules constitute an unsupervised method for learning sequential structure in the world. Another interpretation is that higher-level brain areas provide this tutor signal. For instance, the motor cortex could be directing motor outputs (a process involving slow, deliberate thought) while sending tutor signals to a subcortical network which makes a copy of the behavioral control that allows sequential behavior to be later recalled, thereby skipping the slow deliberation process and corresponding to automatic behavior (18, 58).

This work seeks to fill a significant gap in theory describing the ability to form sequence memories. Namely, it is not clear how general and commonly observed Hebbian learning rules could give rise to such sequence memories. Hebbian learning rules have the advantage of being simple, more biologically realistic alternatives to learning rules such as gradient descent, and have been shown to be implemented in neural circuits. The generality of our theory will allow it to be applicable as new STDP rules are discovered and will enable theorists and experimentalists to probe the functional role that this variety of STDP kernels might be playing (49). Our theory may also be useful for analysis of dynamical systems with sequence dynamics more generally (59) since the mean-field Eq. **9** is quite general.

Findings Relevant to Neuroscience. Our analysis reveals many interesting phenomena of potential consequence to neuroscience. Our first contribution is showing that the replay tempo of sequences stored in neural networks with Hebbian learning rules generically does not match the tempo of the tutor signal. This can be important in modeling observed neural behaviors such as rapid replay as observed in the hippocampus (60). In this case, the tutor signal would be provided by experience with the environment, and recall in the model would correspond to the replay events observed, for instance during sleep. In many cases, the tutor signal is viewed as a signal to be stored faithfully, so that recall has the same tempo as the stored signal. In these cases, our theory makes predictions for the STDP learning rules that will be observed in these neural circuits. We derive simple equations describing the relationship between tempo and the parameters of the (double-sided) exponential kernels commonly seen in biology, showing that recall tempo scales approximately linearly with the tempo of the tutor signal, where the slope of this linear relationship depends on the parameters of the kernel. In addition, our theory relates sequence stability to parameters of the kernel, which could provide predictions for how many applications of the tutor signal are needed for a sequence to be robustly encoded in the neural circuit, since repeat application of the learning rule is analogous to increasing the magnitude of the kernel. This also reveals that Hebbian kernels can be chosen that produce sequences with the same tempo, but that are more or less robust to noise. Our theory lays out a quantitative framework that allows such kernels to be delineated. Finally, we show the result of changing the tempo of the tutor signal throughout the sequence, so that the sequence has faster and slower parts. While the recall tempo roughly follows these changes, we find several undesirable characteristics of the recalled sequence and suggest possible goals for improving the model.

With all this said, it may be necessary to address limitations of our model as described below to acquire useful experimental predictions. Even if this is the case, our theory takes a significant step in the direction of making concrete experimental predictions.

Limitations and Future Work. A shortcoming of our model is that the possibility of correlated patterns is not addressed. It is reasonable to think that the sequence state patterns $\boldsymbol{\xi}^\mu$ could be correlated across neurons and through time in biological circuits, and addressing this may provide a more useful and biologically faithful model. An additional limitation is the interpretation of discrete sequence states, and on understanding what the biologically relevant timescales for tutor signals are. It is possible that neural control signals in higher-order brain areas are "chunked"—that is, a relatively constant control signal initiates another motor area that enacts a motor motif (1). In general, it may still be of interest to introduce tutor signals that are not constant over intervals $[t_\xi^\mu, t_\xi^{\mu+1}]$. Another limitation of our model is the assumption that during learning the network's activity is entirely governed by the tutor signal. In reality, there will likely be ongoing dynamics not related to the tutor signal present during the learning process. In addition, we consider only linear TAH learning rules, although we expect our analysis would extend in spirit to nonlinear learning rules such as those considered in ref. 40. Finally, our model suffers from the overlaps $d_\mu(t)$ "spreading out" with

increasing μ with peak heights that slowly decay, even in the stable regime. As mentioned in the *Sequences with Fast and Slow Parts* subsection of *Results*, adding mechanisms such as nonlinear learning rules (40, 54), "interaction modulation" as suggested by ref. 57, or "delay" filtering as suggested by refs. 35–37 could help mitigate this.

We further test our model's behavior in extreme cases of fast and slow tempos in *SI Appendix*, Fig. S3. This figure shows that there is a transient period where the tempo "settles in" to a limiting tempo. However, this period is relatively short over a wide range of limiting tempos, indicating that our model holds up well under wide choices of parameters. We also test our model's behavior in the biologically relevant parameter regime taken from ref. 47 in *SI Appendix*, Fig. S5. This shows that the model behaves well in this regime as well, though the exact kernel of ref. 47 creates periodic waves of activity due to strong backward connectivity terms (*SI Appendix*, Fig. S5*E*). To probe the dependence of sequence magnitude on parameter choices, we plot the sequence magnitudes p_μ in *SI Appendix*, Figs. S4 and S6, corresponding to *SI Appendix*, Figs. S3 and S5, respectively. These plots show that while sequence magnitudes can decrease, for instance as T_ξ is decreased, this can be counteracted by increasing the magnitude of the Hebbian kernel.

In the long term, memory models should be combined with other models, such as models of the motor cortex supporting flexible motor control. Tutor signals would then be generated by these other models, and gradually the memory model would take over control of motor actions as they become more practiced (61). In addition, mechanisms for refinement based on reward signals could be added to the memory modules. For instance, by gating TAH learning rules based on reward signals, this learning process can be turned into a reinforcement learning process.

Other biologically relevant details may also be useful to add, such as using a spiking neural network model. There is a straightforward extension of our model to spiking neurons, i.e., using the approach of ref. 40. While STDP learning rules are often probed in firing rate models (40, 62), spiking may in some cases be more appropriate (63).

Summary. Our theory shows that sequence memories can be formed by Hebbian learning rules that have been observed in biology, using a theory that is common in neuroscience: the presence of a tutor signal. Our analysis provides clear predictions for the STDP kernels that should be found in neural circuits that store sequence memories. Our work indicates areas of potential improvement and future development that we believe are important for advancing the field.

Data, Materials, and Software Availability. Code for reproducing the plots can be found at https://github.com/msf235/Recall-tempo-of-Hebbian-sequences (64). For details of the parameters used in the simulations for producing the plots, see *SI Appendix*, 3.

ACKNOWLEDGMENTS. M.F. thanks the Swartz Foundation and Harvard Data Science Initiative for its support. M.F. was also supported by the Special Postdoctoral Researcher (SPDR) program at RIKEN during part of this work. C.P. is supported by NSF Award DMS-2134157, NSF CAREER Award IIS-2239780, and a Sloan Research Fellowship. This work has been made possible in part by a gift from the Chan Zuckerberg Initiative Foundation to establish the Kempner Institute for the Study of Natural and Artificial Intelligence.

1. A. M. Graybiel, The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Memory* **70**, 119–136 (1998).
2. O. Hikosaka *et al*., Parallel neural networks for learning sequential procedures. *Trend. Neurosci.* **22**, 464–471 (1999).
3. P. Redgrave *et al*., Goal-directed and habitual control in the basal ganglia: Implications for Parkinson's disease. *Nat. Rev. Neurosci.* **11**, 760–772 (2010).
4. N. Kadmon Harpaz, K. Hardcastle, B. P. Ölveczky, Learning-induced changes in the neural circuits underlying motor sequence execution. *Curr. Opin. Neurobiol.* **76**, 102624 (2022).
5. A. Moors, J. De Houwer, Automaticity: A theoretical and conceptual analysis. *Psychol. Bull.* **132**, 297–326 (2006).
6. K. G. C. Mizes, J. Lindsey, G. S. Escola, B. P. Ölveczky, Dissociating the contributions of sensorimotor striatum to automatic and visually guided motor sequences. *Nat. Neurosci.* **26**, 1791–1804 (2023).
7. F. G. Ashby, B. O. Turner, J. C. Horvitz, Cortical and basal ganglia contributions to habit learning and automaticity. *Trend. Cogn. Sci.* **14**, 208–215 (2010).
8. A. M. Haith, J. W. Krakauer, The multiple effects of practice: Skill, habit and reduced cognitive load. *Curr. Opin. Behav. Sci.* **20**, 196–201 (2018).
9. S. Hélie, D. Cousineau, "The cognitive neuroscience of automaticity: Behavioral and brain signatures" in *Advances in Cognitive and Behavioral Sciences, Psychology Research Progress*, M.-K. Sun, Ed. (Nova Science Publishers, Hauppauge, NY, US, 2014), pp. 141–159.
10. M. Carter, D. Shapiro, Control of sequential movements: Evidence for generalized motor programs. *J. Neurophysiol.* **52**, 787–796 (1984).
11. D. R. Gentner, Timing of skilled motor performance: Tests of the proportional duration model. *Psychol. Rev.* **94**, 255 (1987).
12. S. Sternberg, R. L. Knoll, S. Monsell, C. E. Wright, Motor programs and hierarchical organization in the control of rapid speech. *Phonetica* **45**, 175–197 (1988).
13. M. A. Long, M. S. Fee, Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194 (2008).
14. M. A. Long, D. Z. Jin, M. S. Fee, Support for a synaptic chain model of neuronal sequence generation. *Nature* **468**, 394–399 (2010).
15. R. H. R. Hahnloser, A. A. Kozhevnikov, M. S. Fee, An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
16. C. Pehlevan, F. Ali, B. P. Ölveczky, Flexibility in motor timing constrains the topology and dynamics of pattern generator circuits. *Nat. Commun.* **9**, 977 (2018).
17. E. Marder, D. Bucher, Central pattern generators and the control of rhythmic movements. *Curr. Biol.* **11**, R986–R996 (2001).
18. R. Kawai *et al*., Motor cortex is required for learning but not for executing a motor skill. *Neuron* **86**, 800–812 (2015).
19. T. M. Otchy *et al*., Acute off-target effects of neural circuit manipulations. *Nature* **528**, 358–363 (2015).
20. P. Viviani, G. Laissard, Timing control in motor sequences in advances in psychology. *Elsevier* **81**, 1–36 (1991).
21. E. Marder, R. L. Calabrese, Principles of rhythmic motor pattern generation. *Physiol. Rev.* **76**, 687–717 (1996).
22. D. V. Buonomano, R. Laje, Population clocks: Motor timing with neural dynamics. *Trend. Cogn. Sci.* **14**, 520–527 (2010).
23. J. J. Paton, D. V. Buonomano, The neural basis of timing: Distributed mechanisms for diverse functions. *Neuron* **98**, 687–705 (2018).
24. A. K. Dhawale, S. B. E. Wolff, R. Ko, B. P. Ölveczky, The basal ganglia control the detailed kinematics of learned motor skills. *Nat. Neurosci.* **24**, 1256–1269 (2021).
25. J. Wang, D. Narain, E. A. Hosseini, M. Jazayeri, Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* **21**, 102 (2018).
26. K. H. Srivastava *et al*., Motor control by precisely timed spike patterns. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1171–1176 (2017).
27. C. Tang, D. Chehayeb, K. Srivastava, I. Nemenman, S. J. Sober, Millisecond-scale motor encoding in a cortical vocal area. *PLoS Biol.* **12**, 1–13 (2014).
28. G. B. M. Mello, S. Soares, J. J. Paton, A scalable population code for time in the striatum. *Curr. Biol.* **25**, 1113–1122 (2015).
29. E. D. Remington, D. Narain, E. A. Hosseini, M. Jazayeri, Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron* **98**, 1005–1019.e5 (2018).
30. S. J. Sober, S. Sponberg, I. Nemenman, L. H. Ting, Millisecond spike timing codes for motor control. *Trend. Neurosci.* **41**, 644–648 (2018).
31. D. Obeid, J. A. Zavatone-Veth, C. Pehlevan, Statistical structure of the trial-to-trial timing variability in synfire chains. *Phys. Rev. E* **102**, 052406 (2020).
32. T. Teşileanu, B. Ölveczky, V. Balasubramanian, Rules and mechanisms for efficient two-stage learning in neural circuits. *eLife* **6**, e20944 (2017).
33. T. L. Warren, E. C. Tumer, J. D. Charlesworth, M. S. Brainard, Mechanisms and time course of vocal learning and consolidation in the adult songbird. *J. Neurophysiol.* **106**, 1806–1821 (2011).
34. S. I. Amari, Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comput.* **C-21**, 1197–1206 (1972).
35. H. Sompolinsky, I. Kanter, Temporal association in asymmetric neural networks. *Phys. Rev. Lett.* **57**, 2861–2864 (1986).
36. D. Kleinfeld, Sequential state generation by model neural networks. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9469–9473 (1986).
37. D. Kleinfeld, H. Sompolinsky, Associative neural network model for the generation of temporal patterns. Theory and application to central pattern generators. *Biophys J.* **54**, 1039–1051 (1988).
38. A. Herz, B. Sulzer, R. Kühn, J. L. van Hemmen, The hebb rule: Storing static and dynamic objects in an associative neural network. *Europhys. Lett.* **7**, 663–669 (1988).
39. R. Kühn *et al*., "Temporal Association" in *Models of Neural Networks, Physics of Neural Networks*, E. Domany, Ed. (Springer, Berlin Heidelberg, 1991).
40. M. Gillett, U. Pereira, N. Brunel, Characteristics of sequential activity in networks with temporally asymmetric hebbian learning. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 29948–29958 (2020).
41. M. Rabinovich *et al*., Dynamical encoding by networks of competing neuron groups: Winnerless competition. *Phys. Rev. Lett.* **87**, 068102 (2001).

42. M. M. Churchland, K. V. Shenoy, Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *J. Neurophysiol.* **97**, 4235–4257 (2007).
43. I. R. Fiete, W. Senn, C. Z. H. Wang, R. H. R. Hahnloser, Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65**, 563–576 (2010).
44. U. Pereira, N. Brunel, Unsupervised learning of persistent and sequential activity. *Front. Comput. Neurosci.* **13**, 97 (2020).
45. J. J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3088–3092 (1984).
46. G. Bi, M. Poo, Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
47. G. Bi, M. Poo, Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* **24**, 139–166 (2001).
48. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558 (1982).
49. L. F. Abbott, S. B. Nelson, Synaptic plasticity: Taming the beast. *Nat. Neurosci.* **3**, 1178–1183 (2000).
50. G. M. Wittenberg, S. S. H. Wang, Malleability of spike-timing-dependent plasticity at the CA3–CA1 synapse. *J. Neurosci.* **26**, 6610–6617 (2006).
51. Y. Luz, M. Shamir, The effect of STDP temporal kernel structure on the learning dynamics of single excitatory and inhibitory synapses. *PLoS One* **9**, e101109 (2014).
52. B. Babadi, L. F. Abbott, Pairwise analysis can account for network structures arising from spike-timing dependent plasticity. *PLoS Comput. Biol.* **9**, 1–16 (2013).
53. E. T. Reifenstein, I. B. Khalid, R. Kempter, Synaptic learning rules for sequence learning. *eLife* **10**, e67171 (2021).
54. M. Gillett, N. Brunel, Dynamic control of sequential retrieval speed in networks with heterogeneous learning rules. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.03.22.533836 (Accessed 24 March 2023).
55. S. Recanatesi, U. Pereira-Obilinovic, M. Murakami, Z. Mainen, L. Mazzucato, Metastable attractors explain the variable timing of stable behavioral action sequences. *Neuron* **110**, 139–153.e9 (2022).
56. Z. Jiang, J. Zhou, T. Hou, K. Y. M. Wong, H. Huang, Associative memory model with arbitrary Hebbian length. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **104**, 064306 (2021).
57. L. Herron, P. Sartori, B. Xue, Robust retrieval of dynamic sequences through interaction modulation. *PRX Life* **1**, 023012 (2023).
58. J. M. Murray, G. S. Escola, Learning multiple variable-speed sequences in striatum via cortical tutoring. *eLife* **6**, e26084 (2017).
59. M. Asllani, R. Lambiotte, T. Carletti, Structure and dynamical behavior of non-normal networks. *Sci. Adv.* **4**, eaau9403 (2018).
60. W. E. Skaggs, B. L. McNaughton, Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* **271**, 1870–1873 (1996).
61. J. M. Murray, G. S. Escola, Remembrance of things practiced with fast and slow learning in cortical and subcortical pathways. *Nat. Commun.* **11**, 6441 (2020).
62. L. F. Abbott, K. I. Blum, Functional significance of long-term potentiation for sequence learning and prediction. *Cereb. Cortex* **6**, 406–416 (1996).
63. E. M. Izhikevich, Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb. Cortex* **17**, 2443–2452 (2007).
64. M. Farrell, Data from "Recall-tempo-of-Hebbian-sequences." Github. https://github.com/msf235/Recall-tempo-of-Hebbian-sequences. Deposited 9 June 2023.

Supporting Information for

Recall tempo of Hebbian sequences depends on the interplay of Hebbian kernel with tutor signal timing

Matthew Farrell and Cengiz Pehlevan

Cengiz Pehlevan
E-mail: cpehlevan@seas.harvard.edu

This PDF file includes:

Supporting Information Text

SI Appendix 1

Derivation of mean-field equations. Here we derive the mean-field equations Eq. (9). This is a fairly straightforward generalization of the derivation in (1). The full network equations are

$$\tau \dot{\boldsymbol{r}} = -\boldsymbol{r} + \phi(W\boldsymbol{r} + \boldsymbol{\eta})$$

where $\boldsymbol{\eta}(t)$ is a white noise vector ($\langle \eta_i(t) \rangle = 0$ and $\langle \eta_i(t)\eta_i(t') \rangle = \tau\rho^2\delta(t-t')$). To simulate these equations, we use forward Euler integration with timestep dt:

$$\boldsymbol{r}(t+dt) = \boldsymbol{r}(t) + dt\left(-\boldsymbol{r}(t) + \phi(W\boldsymbol{r}(t) + \boldsymbol{\eta}(t))\right)/\tau$$

where each $\eta_i(t)$ is drawn i.i.d. from a standard normal distribution with mean 0 and variance $\tau\rho^2/dt$.

The function ϕ is a sigmoidal nonlinearity:

$$\phi(x) = \frac{r_{\text{span}}}{2}\left(r_{\text{center}} + \text{erf}\left(\frac{x-\theta}{\sqrt{2}\sigma}\right)\right). \qquad [A.1]$$

Note that τ simply rescales the temporal timescale of the dynamics. Due to this simple behavior, we can simply take $\tau = 1$ for the following derivation without loss of generality. The weights W are

$$W = \frac{1}{N}\sum_{\mu=1}^{P}\sum_{\nu=1}^{P} a_\mu^\nu \boldsymbol{\xi}^\nu (\boldsymbol{\xi}^\mu)^\top$$

Here each pattern $\boldsymbol{\xi}^\mu$ is standard normal: $\xi_k^\mu \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. The mean-field equations are written in terms of the overlaps $q_\mu(t) = (\boldsymbol{\xi}^\mu)^\top \boldsymbol{r}(t)/N$. If we let $\boldsymbol{h} = W\boldsymbol{r} + \boldsymbol{\eta}$ we can then write

$$\boldsymbol{h} = \sum_{\mu=1}^{P}\sum_{\nu=1}^{P} a_\mu^\nu \boldsymbol{\xi}^\nu q_\mu + \boldsymbol{\eta}.$$

We now investigate the evolution of the overlaps q_μ:

$$\begin{aligned}\dot{q}_\mu &= (\boldsymbol{\xi}^\mu)^\top \dot{\boldsymbol{r}}/N \\ &= -q_\mu + (\boldsymbol{\xi}^\mu)^\top \phi(\boldsymbol{h})/N.\end{aligned}$$

Let's look more closely at the second term on the right hand side. As $N \to \infty$ this term approaches an average by the law of large numbers, yielding $(\boldsymbol{\xi}^\mu)^\top \phi(\boldsymbol{h})/N \to \langle \xi^\mu \phi(h) \rangle$ where $\xi^\nu \overset{i.i.d}{\sim} \mathcal{N}(0,1)$ and

$$h = \sum_{\mu=1}^{P}\sum_{\nu=1}^{P} a_\mu^\nu \xi^\nu q_\mu + \eta.$$

Here η is a scalar white noise term: $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \rho^2\delta(t-t')$. Hence,

$$\begin{aligned}\dot{q}_\mu &= -q_\mu + \left\langle \xi^\mu \phi\left(\sum_{\ell=1}^{P}\sum_{\nu=1}^{P} a_\ell^\nu \xi^\nu q_\ell + \eta\right)\right\rangle \\ &= -q_\mu + \left\langle \xi^\mu \phi\left(\xi^\mu \sum_{\ell=1}^{P} a_\ell^\mu q_\ell + \sum_{\nu\neq\mu}^{P} \xi^\nu \sum_{\ell=1}^{P} a_\ell^\nu q_\ell + \eta\right)\right\rangle\end{aligned}$$

We next note that each term of

$$\sum_{\nu\neq\mu}^{P} \xi^\nu \sum_{\ell=1}^{P} a_\ell^\nu q_\ell$$

is an independent normally distributed random variable with mean zero and variance $(\sum_{\ell=1}^{P} a_\ell^\nu q_\ell)^2$. Hence the sum $\sum_{\nu\neq\mu}^{P} \xi^\nu \sum_{\ell=1}^{P} a_\ell^\nu q_\ell + \eta$ is distributed like xR_μ where $R_\mu^2 = \sum_{\nu\neq\mu}^{P}(\sum_{\ell=1}^{P} a_\ell^\nu q_\ell)^2 + \rho^2$ and x is standard normal. We similarly define $S_\mu = \sum_{\ell=1}^{P} a_\ell^\nu q_\ell$ and write the average as a Gaussian integral:

$$\dot{q}_\mu = -q_\mu + \int \mathcal{D}\xi^\mu \mathcal{D}x\, \xi^\mu \phi\left(\xi^\mu S_\mu + xR_\mu\right) \qquad [A.2]$$

where $\mathcal{D}x = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.

The remainder of the calculation follows closely (1). We next define the change of coordinates

$$v = \frac{\xi^\mu S_\mu + x R_\mu}{\sqrt{S_\mu^2 + R_\mu^2}} = \frac{\xi^\mu S_\mu + x R_\mu}{c}$$
$$u = \frac{\xi^\mu R_\mu - x S_\mu}{\sqrt{S_\mu^2 + R_\mu^2}} = \frac{\xi^\mu R_\mu - x S_\mu}{c}$$

where v and u are uncorrelated standard normal random variables and where we have defined $c = \sqrt{S_\mu^2 + R_\mu^2}$. To perform this change of variables we first compute the determinant of the Jacobian:

$$\begin{vmatrix} \frac{dv}{d\xi^\ell} & \frac{dv}{dx} \\ \frac{du}{d\xi^\ell} & \frac{du}{dx} \end{vmatrix} = \frac{1}{c^2} \begin{vmatrix} S_\mu & R_\mu \\ R_\mu & -S_\mu \end{vmatrix} = \frac{1}{c^2} \left(S_\mu^2 + R_\mu^2 \right) = 1.$$

Hence we can use the substitution $dudv = d\xi^\mu dx$. Furthermore,

$$\begin{aligned} 2\pi \mathcal{D}u \mathcal{D}v &= e^{-u^2/2} e^{-v^2/2} dudv \\ &= \exp\left(-\frac{(\xi^\mu R_\mu - x S_\mu)^2}{2c^2} \right) \exp\left(-\frac{(\xi^\mu S_\mu + x R_\mu)^2}{2c^2} \right) dudv \\ &= \exp\left(-\frac{(\xi^\mu R_\mu)^2 + (\xi^\mu S_\mu)^2 + (x S_\mu)^2 + (x R_\mu)^2}{2c^2} \right) dudv \\ &= \exp\left[-\left((\xi^\mu)^2 + x^2 \right)/2 \right] d\xi^\mu dx \\ &= 2\pi \mathcal{D}\xi^\mu \mathcal{D}x. \end{aligned}$$

Inverting the change of coordinate equations above, we find that $\xi^\mu = (S_\mu v + R_\mu u)/c$. Using these substitutions, we can rewrite the integral

$$\begin{aligned} \int \mathcal{D}\xi^\mu \mathcal{D}x \, \xi^\mu \phi\left(\xi^\mu S_\mu + x R_\mu \right) &= \frac{1}{c} \int \mathcal{D}u \mathcal{D}v \, (v S_\mu + u R_\mu) \phi(cv) \\ &= \frac{S_\mu}{c} \int \mathcal{D}v \, v \phi(cv). \end{aligned}$$

We now define

$$G(x) = \frac{1}{\sqrt{x}} \int \mathcal{D} \, v v \phi(v\sqrt{x}).$$

Using the definition of ϕ, this integral can be evaluated with integration by parts and is

$$G(x) = \frac{r_{\text{span}}}{\sqrt{2\pi(\sigma^2 + x)}} \exp\left(-\frac{\theta^2}{2(\sigma^2 + x)} \right). \qquad [A.3]$$

Our equation becomes

$$\begin{aligned} \dot{q}_\mu &= -q_\mu + G(R_\mu^2 + S_\mu^2) S_\mu \\ &= -q_\mu + G\left(\sum_\nu \left(\sum_\ell a_\ell^\nu q_\ell \right)^2 + \rho^2 \right) \sum_{\ell=1}^P a_\ell^\mu q_\ell. \\ &= -q_\mu + G\left(\|A\boldsymbol{q}\|_2^2 + \rho^2 \right) \sum_{\ell=1}^P a_\ell^\mu q_\ell. \end{aligned}$$

where A is the matrix $A = (a_\ell^\nu)_{\nu,\ell}$ (ν indexes the rows and ℓ indexes the columns) and $\boldsymbol{q}$ is the vector $\boldsymbol{q} := (q_1, \ldots, q_P)^\top$. Finally, we define

$$g(t) = G\left(\|A\boldsymbol{q}(t)\|_2^2 + \rho^2 \right),$$

which results in Eq. (9). Note that $\lim_{N\to\infty} \|W\boldsymbol{r}\|_2^2/N = \|A\boldsymbol{q}(t)\|_2^2$: letting $\boldsymbol{u} = W\boldsymbol{r}$, the law of large numbers says that $\lim_{N\to\infty} \frac{1}{N} \sum_i u_i^2 = \langle u_i^2 \rangle = \sum_\nu \left(\sum_\ell a_\ell^\nu q_\ell \right)^2$ since $u_i = \sum_\nu \xi_i^\nu \sum_\ell a_\ell^\nu q_\ell$ is a Gaussian random variable with mean 0 and variance $\sum_\nu \left(\sum_\ell a_\ell^\nu q_\ell \right)^2$. Plots showing the convergence of the mean-field approximation to the network simulations for increasing N are shown in Fig. 2. Note that in the Appendices below, for theoretical analysis the number of patterns P will also be taken to infinity. In this case, in order for these mean-field dynamics to hold exactly, P/N must vanish as N and P are both taken large.

Note that stability is determined by the shape of the function G. For the values of ϕ considered in the main text ($r_{\text{span}} = 2$, $r_{\text{center}} = 0$, $\theta = 0$, and $\sigma = 0.1$), G has the shape plotted in Fig. S1. In particular, G is monotonically decreasing from its maximum at $x = 0$. Since $\|h(t)\|^2 \geq 0$, the maximal value that $g(t)$ can take during the ongoing dynamics is bounded from above by $G(\rho^2)$; see SI Appendix 2 below for more details. Note that the function G is the same as in (1), and the supplementary material of (1) contains plots of G for various values of r_{span}, θ, and σ.

Our choice of parameters for ϕ means that firing rates $\boldsymbol{r}$ are in the range $[-1, 1]$. For nonnegative firing rates, different parameter values can be chosen, as in (1). However, in this case it may be sensible to use nonnegative random patterns $\boldsymbol{\xi}$ rather than Gaussian. We do not investigate this possibility further here.

SI Appendix 2

In this appendix we show the full details of the mathematical derivations of the peak times t_μ and stability conditions as defined in the main text.

Stability conditions. Here we derive stability conditions, and for stable dynamics we show how the value of $g_\infty := \lim_{t\to\infty} g(t)$ depends on the coefficients a_k. We start with the mean-field equations Eq. (9) and assume that the coefficients are shift-invariant, so that $a^\mu_\nu = a_{\mu-\nu}$:

$$\dot{q}_\mu = -q_\mu + g(t) \sum_{\nu=1}^{P} a_{\mu-\nu} q_\nu. \tag{A.4}$$

Note that stability is defined only in the $P \to \infty$ limit, so the equation above and those below are defined also for $P = \infty$. As in SI Appendix 1, we define $\boldsymbol{q}$ to be the (possibly infinite length) column vector $\boldsymbol{q} := (q_\mu)_{\mu\in\{1,2,\ldots,P\}}$. To show the dependence of $\boldsymbol{q}$ on t, P, and initial conditions $\boldsymbol{q}(0)$ (which itself depends on P) we write $\boldsymbol{q}(t)$ or $\boldsymbol{q}(t; P)$ or $\boldsymbol{q}(t; P, \boldsymbol{q}(0))$. With this definition, we can write Eq. (A.4) as a matrix-vector equation

$$\dot{\boldsymbol{q}} = -\boldsymbol{q} + g(t) A \boldsymbol{q} \tag{A.5}$$

where A is a Toeplitz matrix that holds the coefficients a_k:

$$A = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-P+1} \\ a_1 & a_0 & a_{-1} & \ddots & \vdots \\ a_2 & a_1 & a_0 & \ddots & a_{-2} \\ \vdots & \ddots & \ddots & \ddots & a_{-1} \\ a_{P-1} & \cdots & a_2 & a_1 & a_0 \end{pmatrix}.$$

We make the following assumptions, which we refer to as our "standard list of assumptions":

Standard list of assumptions:

Positive pairs: $a_k + a_{-k} \geq 0$ for all $k > 0$.

Finite norm: $\|A\|_\infty < \infty$.

Continuity: $g(t)$ is continuous.

Finite $\bar{g}$: $\bar{g} < \infty$.

Rapid convergence: $\lim_{t\to\infty} \left| \int_0^t ds(g(s) - g_\infty) \right| < \infty$.

Our objective is to show that

$$\lim_{t\to\infty} g(t) =: g_\infty = \bar{g} := \left(\sum_{k=-P+1}^{P-1} a_k \right)^{-1} \tag{A.6}$$

is a critical relation for stability in the following senses:

Proposition 1 (Nonperiodic boundary conditions)**.** *Let $\boldsymbol{q}(t; P, \boldsymbol{q}(0; P))$ be a sequence of solution trajectories, indexed by P, for Eq. (A.5) under the standard list of assumptions above. In addition, let $\boldsymbol{q}(t; P = \infty)$ denote the solution trajectory for Eq. (A.5) with $P = \infty$ and initial condition $\boldsymbol{q}(0; P = \infty)$. Then the following hold:*

- *(exponential decay for $g_\infty < \bar{g}$): $\limsup_{P\to\infty} \|\boldsymbol{q}(t; P, \boldsymbol{q}(0; P))\|_p = O(e^{t(-1+g_\infty/\bar{g})})$ provided $\limsup_{P\to\infty} \|\boldsymbol{q}(0; P)\|_p < \infty$ for $p \geq 2$, including $p = \infty$. In addition, $\|\boldsymbol{q}(t; P = \infty, \boldsymbol{q}(0))\|_p = O(e^{t(-1+g_\infty/\bar{g})})$ for all initial conditions $\boldsymbol{q}(0)$ provided $\|\boldsymbol{q}(0)\|_p < \infty$.*

- *(unbounded growth for $g_\infty > \bar{g}$): If $g_\infty > \bar{g}$, then for each time point t, there exists a sequence of initial conditions $\boldsymbol{q}(0;P)$ indexed by P (and possibly depending on t) with $\limsup_{P\to\infty} \|\boldsymbol{q}(0;P)\|_2 < \infty$ such that $\lim_{t\to\infty} \limsup_{P\to\infty} \|\boldsymbol{q}(t;P,\boldsymbol{q}(0;P))\|_2 = \infty$. In addition, there exists a sequence of initial conditions $\boldsymbol{q}_k(0;P=\infty)$ (possibly depending on t) with $\limsup_{k\to\infty} \|\boldsymbol{q}_k(0;P=\infty)\|_2 < \infty$ such that $\lim_{t_0\to\infty} \sup_{0\le t\le t_0} \limsup_{k\to\infty} \|\boldsymbol{q}(t;\infty,\boldsymbol{q}_k(0;\infty))\|_2 = \infty$.*

To summarize the above results, as $P\to\infty$ and for $g_\infty < \bar{g}$ the solutions are unstable, while for $g_\infty > \bar{g}$ there are ways to choose initial conditions that make the solutions unbounded.

A proof of this proposition is given below. The result that we would ideally like to show, but that we were not able to at this point, is encoded in the following conjecture. First we need to define a new condition:

Forward-propagation-dominated: $a_k \ge |a_{-k}|$ for all $k > 0$.

This is a restriction of the positive pairs condition defined above.

Conjecture 1 (Nonperiodic boundary conditions)**.** *Let $\boldsymbol{q}$ be a solution trajectory for Eq. (A.5) under the standard list of assumptions as well as the forward-propagation-dominated condition with $P=\infty$ and $\boldsymbol{q}(0)=\boldsymbol{e}_1$ (the vector with 1 in the first coordinate and 0 elsewhere). Then $\boldsymbol{q}(t)$ is stable and bounded for all time t if and only if $g_\infty = \bar{g}$.*

The above results concern Eq. (A.4), which has nonperiodic boundary conditions in μ. Next we consider the analogous equation with periodic boundary conditions. In thise case, we will need to assume that P is finite. For ease of exposition, we will also assume that P is even; the odd case is similar. We will also need the following assumption:

Band-limited: $a_k = 0$ for $|k| > P/2-1$.

This is a stronger condition than the finite norm condition in our standard list of assumptions. Let $H = P/2-1$ and consider the equation

$$\dot{q}_\mu = -q_\mu + g(t)\sum_{\nu=-H}^{H} a_\nu q_{\mu-\nu \mod P+1}. \qquad \text{[A.7]}$$

(the modulo is in the subscript of q and is taken to be in the range $[1,P]$, so in particular $0 \mod P+1 = P$). This corresonds to Eq. (A.5) with A replaced by the circulant coefficient matrix

$$C = \begin{pmatrix} a_0 & a_{-1} & \cdots & a_{-H} & 0 & a_H & \cdots & a_2 & a_1 \\ a_1 & a_0 & a_{-1} & \cdots & a_{-H} & 0 & a_H & \cdots & a_2 \\ \vdots & a_1 & \ddots & \ddots & \vdots & & & & \\ a_H & \vdots & \ddots & \ddots & \vdots & & & & \\ 0 & \ddots & \ddots & \ddots & \vdots & & & & \\ a_{-H} & 0 & a_H & \vdots & \ddots & \ddots & & & \\ \vdots & \vdots & \ddots & \ddots & & & & & a_{-1} \\ a_{-1} & a_{-2} & a_{-3} & \cdots & & & & a_1 & a_0 \end{pmatrix}.$$

Matters simplify considerably in the periodic case:

Proposition 2 (Periodic boundary conditions)**.** *Let $\boldsymbol{q}(t;P)$ be a sequence of solution trajectories to Eq. (A.7) indexed by P under the standard list of assumptions with $\boldsymbol{q}(0;P)=\boldsymbol{e}_1$ (the vector with 1 in the first coordinate and 0 elsewhere). Then $\boldsymbol{q}(t;P)$ is stable and bounded if and only if $g_\infty = \bar{g}$. Futhermore, $\lim_{P\to\infty} \|\boldsymbol{q}(t)\|_2 \sim t^{-1/4} e^{t(-1+g_\infty/\bar{g})}$.*

A significant difference between the periodic and nonperiodic cases is that the direction of a propagating wave of activity matters in the latter but not in the former. In particular, "backward" propagating waves eventually hit the boundary condition at $\mu = 1$ in the nonperiodic case, which can cause activity to die out when it would not do so in the periodic case. This explains the presence of the forward-propagating-dominated condition in Conjecture 1. One approach to proving Conjecture 1 may be to formally connect the periodic and nonperiodic solutions in such a way that their stability properties are shown to be equivalent under the forward-propagating-dominated condition. The solutions in Proposition 1 avoid this issue by allowing for a sequence of initial conditions that move away from the $\mu = 1$ boundary; for instance, the initial conditions could be $\boldsymbol{q}(0;P) = \boldsymbol{e}_P$.

To start to prove these results, we need the following lemma which will allow us to link our nonlinear equation with a linear one:

Lemma 1 (Stability equivalence of linear and nonlinear equations)**.** *Suppose that $\boldsymbol{q}_1(t)$ is a solution trajectory to Eq. (A.5) for arbitrary A that satisfied the finite norm condition, and with $g(t)$ an arbitrary function satisfying the rapid convergence assumption, and $\boldsymbol{q}_2(t)$ is a solution trajectory to Eq. (A.5) with $g(t) = g_\infty$ under the finite norm condition above. Then $\boldsymbol{q}_1$ is stable if and only if $\boldsymbol{q}_2$ is stable, and $\boldsymbol{q}_1$ is bounded for all time if and only if $\boldsymbol{q}_2$ is bounded for all time.*

Proof. For now we will assume that $g_\infty \neq 0$. The solutions to equation Eq. (A.5) for general $g(t)$ (that satisfies the standard list of assumptions) and for $g(t) = g_\infty$ are

$$\boldsymbol{q}_1(t) = e^{-t} e^{g_\infty A \int_0^t ds\, g(s)/g_\infty} \boldsymbol{q}(0) \tag{A.8}$$

$$\boldsymbol{q}_2(t) = e^{-t} e^{g_\infty A t} \boldsymbol{q}(0), \tag{A.9}$$

respectively. Next define $f(t) := g(t)/g_\infty - 1$. The function f measures the deviation of g/g_∞ from its eventual limiting value of 1. Then

$$\int_0^t ds g(s)/g_\infty = \int_0^t ds(f(s)+1) = t + \int_0^t ds f(s).$$

It follows that

$$\boldsymbol{q}_1(t) = e^{-t} e^{g_\infty A t} e^{g_\infty A \int_0^t ds\, f(s)} \boldsymbol{q}(0) = Y(t)\boldsymbol{q}_2(t) \tag{A.10}$$

where we have defined $Y(t) := e^{g_\infty A \int_0^t ds\, f(s)}$. Suppose first that $\boldsymbol{q}_1$ is not stable, which means that $\|\boldsymbol{q}_1(t)\|_\infty = O(t^b e^{at})$ for some real number b and $a < 0$. Since

$$\|\boldsymbol{q}_2(t)\|_\infty = \|Y^{-1}(t)\boldsymbol{q}_1(t)\|_\infty \leq \|Y^{-1}(t)\|_\infty \|\boldsymbol{q}_1(t)\|_\infty \tag{A.11}$$

we have that

$$\|\boldsymbol{q}_2(t)\|_\infty = O\left(t^b e^{at} \|Y^{-1}(t)\|_\infty\right) = O\left(t^b e^{at}\right)$$

where the last equality follows from the finite norm and rapid convergence conditions as stated in the standard list of assumptions ($\|A\|_\infty < \infty$ and $\lim_{t\to\infty} |\int_0^t ds\, f(s)| < \infty$. Note that this second condition is equivalent to $\lim_{t\to\infty} |\int_0^t ds\, g(s) - g_\infty| < \infty$). Hence $\boldsymbol{q}_2$ is not stable.

Going the other way, suppose $\boldsymbol{q}_2$ is not stable, so $\|\boldsymbol{q}_2(t)\|_\infty = O(t^b e^{at})$ for some real number b and $a < 0$. Since

$$\|\boldsymbol{q}_1(t)\|_\infty = \|Y(t)\boldsymbol{q}_2(t)\|_\infty \leq \|Y(t)\|_\infty \|\boldsymbol{q}_2(t)\|_\infty \tag{A.12}$$

we have that

$$\|\boldsymbol{q}_1(t)\|_\infty = O\left(t^b e^{at} \|Y(t)\|_\infty\right) = O\left(t^b e^{at}\right)$$

where the last equality follows from the finite norm and rapid convergence conditions as stated in the standard list of assumptions. Hence under these conditions, we have completed the second direction of the equivalence.

Next we assume that $\sup_t \|q_1(t)\|_\infty < \infty$, which by Eq. (A.11) implies that $\sup_t \|q_2(t)\|_\infty < \infty$ provided $\sup_t \|Y^{-1}(t)\|_\infty < \infty$. To show that $\sup_t \|Y^{-1}(t)\|_\infty < \infty$, we first note that continuity of g implies that $\sup_t \int_0^t ds\, f(s) < \infty$ is equivalent to $\lim_{t\to\infty} \int_0^t ds\, f(s) < \infty$. Then the finite norm and rapid convergence assumptions imply that $\sup_t \|Y^{-1}(t)\|_\infty < \infty$ as above. Finally we assume that $\sup_t \|q_2(t)\|_\infty < \infty$, which by Eq. (A.12) implies that $\sup_t \|q_1(t)\|_\infty < \infty$ provided $\sup_t \|Y(t)\|_\infty < \infty$, which holds for the same reasons as for $\|Y^{-1}(t)\|_\infty$. This finishes the equivalence.

In the above, we assumed that $g_\infty \neq 0$. The case $g_\infty = 0$ follows a similar line of reasoning. We leave this simple case to the reader. □

Lemma 1 says that to understand the stability properties of the nonlinear Eq. (A.4), we can instead consider the stability properties of its linear analogue where we set $g(t) = g_\infty$. For the following proofs, $\boldsymbol{q}$ will refer to Eq. (A.9), the solution to the linear equation with $g(t) = g_\infty$. This equation is valid for P finite or $P = \infty$, in which case A is an infinite-dimensional matrix. In addition to the notation above, from here we let A refer to the infinite-dimensional matrix and let A_P refer to the $P \times P$ truncation of A. By this definition $A_\infty = A$.

We now need to incorporate machinery from the theory of Toeplitz matrices. This will allow us to define an analogue of the spectrum of the matrix in infinite dimensions.* Let $a : S^1 \to \mathbb{C}$ be the function mapping the unitary complex numbers $s(\theta) = e^{2\pi i\theta}$ to the complex plane defined by $a(s) = \sum_{k=-P+1}^{P-1} a_k s^k$. The function a is called the symbol of the Toeplitz matrix A. While the spectral absissa is the quantity of interest when investigating stability of finite-dimensional systems, it is actually the real part of the symbol which is the relevant quantity in infinite dimensions.

Lemma 2. *Suppose A_P (for P finite or infinite) satisfies the finite norm and positive pairs conditions listed above. Then*

$$\sup_s \Re a(s) = (\bar{g})^{-1} = \sum_{k=-P+1}^{P-1} a_k$$

*This is actually more closely related to the pseudospectrum of the finitely truncated matrices than it is to the spectrum of these truncations (2), and in general the eigenvalues are a misleading indication of stability. For instance, the infinite matrix A with $a_1 = 1$ and $a_k = 0$ for $k \neq 1$ gives rise to stable dynamics, even though the eigenvalues of every truncation A_P are 0, which together with the $-\boldsymbol{q}$ decay term should indicate that the solutions decay to 0 with asymptotic rate e^{-t}. Indeed, solutions do decay to 0, but only after the traveling wave of activity has hit the last pattern in the sequence. If there is no last pattern (i.e. infinite-dimensional A), then the activity never decays to 0.

Proof. The symbol of A is

$$a(s) = \sum_{k=-P+1}^{P-1} a_k s^k.$$

By the finite norm condition (see Section 1.6 of (2)), this sum is absolutely convergent and we can rearrange terms:

$$a(s) = a_0 + \sum_{k=1}^{P-1} (a_k s^k + a_{-k} s^{-k}).$$

Paramaterizing s as $s = e^{2\pi i\theta}$, we have that

$$\sup_{s\in S^1} \Re a(s) = a_0 + \sup_{\theta\in[0,2\pi)} \sum_{k=1}^{P-1} (a_k \cos(2\pi k\theta) + a_{-k}\cos(-2\pi k\theta)) = a_0 + \sup_{\theta\in[0,2\pi)} \sum_{k=1}^{P-1} (a_k + a_{-k})\cos(2\pi k\theta).$$

Using the positive pairs condition, we have that this sum is maximized when $\theta = 0$, yielding

$$\sup_{s\in S^1} \Re a(s) = \sum_{k=-P+1}^{P-1} a_k.$$

□

We are now ready to prove Propositions 1 and 2.

Proof of Proposition 1. Recall that $\boldsymbol{q}(t) = e^{tg_\infty A}\boldsymbol{q}(0)$. Our general strategy is to translate the results of Theorem 8.15 of (2) regarding the behavior of the norm of the transfer function $e^{tg_\infty A}$ to statements about the norm of the solution $e^{tg_\infty A}\boldsymbol{q}(0)$.

We first assume that $g_\infty < \bar{g}$. Then for $p \geq 2$ (including $p = \infty$)

$$\|\boldsymbol{q}(t; P, \boldsymbol{q}(0))\|_p \leq e^{-t}\|e^{g_\infty t A_P}\|_p \|\boldsymbol{q}(0)\|_p \leq e^{-t}\|e^{g_\infty t A_P}\|_2 \|\boldsymbol{q}(0)\|_p. \tag{A.13}$$

Using the finiteness of $\limsup_{P\to\infty} \|\boldsymbol{q}(0)\|_p$, we have that

$$\limsup_{P\to\infty} \|\boldsymbol{q}(t; P, \boldsymbol{q}(0))\|_p \leq \limsup_{P\to\infty} e^{-t}\|e^{tg_\infty A_P}\|_2 \|\boldsymbol{q}(0)\|_p = e^{-t}e^{g_\infty t \sup \Re a} \limsup_{P\to\infty} \|\boldsymbol{q}(0)\|_p$$

where the last equality uses the fact that $\lim_{P\to\infty} \|e^{tg_\infty A_P}\|_2 = e^{tg_\infty \sup \Re a}$ for each $t > 0$, which is Theorem 8.15 of (2). Using Lemma 2 it follows that $\limsup_{P\to\infty} \|\boldsymbol{q}(t; P, \boldsymbol{q}(0))\|_p = O(e^{t(-1+g_\infty/\bar{g})})$. Next we note that by the Banach-Steinhaus theorem $\|e^{tg_\infty A}\|_2 \leq \liminf_{P\to\infty} \|e^{tg_\infty A_P}\|_2$. As stated above, by Theorem 8.15 of (2) the limit exists and is equal to $e^{tg_\infty \sup \Re a}$. Hence

$$\|e^{tg_\infty A}\|_2 \leq \lim_{P\to\infty} \|e^{tg_\infty A_P}\|_2 = e^{tg_\infty \sup \Re a} = e^{tg_\infty/\bar{g}}. \tag{A.14}$$

It follows that

$$\boldsymbol{q}(t, P = \infty) \leq e^{-t}\|e^{tg_\infty A}\|_2 \|\boldsymbol{q}(0)\|_2 = O\left(e^{t(-1+g_\infty/\bar{g})}\right).$$

provided $\|\boldsymbol{q}(0)\|_2 < \infty$.

Next assume that $g_\infty > \bar{g}$ and fix a $t > 0$. By the definition of the matrix operator norm, for each time t and system size P we can choose a sequence of initial conditions $\boldsymbol{q}(0; P)$ (that may depend on t) such that $\|e^{tg_\infty A_P}\boldsymbol{q}(0; P)\|_2 \geq \|e^{tg_\infty A_P}\|_2 - c$ for some constant c which does not depend on t or P. Taking the limit yields

$$\liminf_{P\to\infty} \|e^{tg_\infty A_P}\boldsymbol{q}(0; P)\|_2 \geq e^{tg_\infty \sup \Re a} - c. \tag{A.15}$$

We have again by Theorem 8.15 of (2) that $\lim_{P\to\infty} \|e^{tg_\infty A_P}\|_2 = e^{tg_\infty \sup \Re a}$. Now by Lemma 2, $g_\infty \sup \Re a = g_\infty/\bar{g} > 1$. It follows that $\lim_{t\to\infty} \liminf_{P\to\infty} \|e^{g_\infty A_P t}\boldsymbol{q}(0; P)\|_2 = \infty$.

We now prove the final portion of the proposition, which again is under the case of $g_\infty > \bar{g}$. Following Theorem 8.15 of (2) is the following result: if the pseudospectrum $\mathrm{sp}_\varepsilon(A)$ of A contains points in the open right-half plane, then each point $\lambda \in \mathrm{sp}_\varepsilon(A)$ with $\beta := \Re\lambda > 0$ yields an estimate of the form

$$\sup_{0\leq t\leq t_0} \|e^{tA}\|_2 \geq e^{\beta t_0} / \left(1 + \varepsilon \frac{e^{\beta t_0} - 1}{\beta}\right). \tag{A.16}$$

We will not define the pseudospectrum $\mathrm{sp}_\varepsilon(A)$ here but instead note that Section 7.4 of (2) implies that if there is η in the spectrum of A with $\Re\eta > 0$, then for ε small enough there is also a point λ in the pseudospectrum of A with $\Re\lambda > 0$. It follows that if such an η exists, then $\lim_{t_0\to\infty} \sup_{0\leq t\leq t_0} \|e^{tA}\|_2 = \infty$. Finally, we need to connect $g_\infty \Re a$ with the spectrum of A. This connection comes in the form of Corollary 1.10 of (2), which implies that the symbol of A is contained in the spectrum of A. It follows that $g_\infty \Re a > 1$ implies that there is a point η in the spectrum of A with $\Re\eta > 0$. Finally, we construct a sequence of initial conditions $\boldsymbol{q}_k(0; P = \infty)$ (which may depend on t) that realizes the supremum of the operator norm. This concludes the proof. □

Next we can prove Proposition 2. The circulant (periodic) situation is much simpler because circulant matrices are diagonalizable by Fourier modes.

Proof of Proposition 2. This proof is similar to that of Lemma 2. Indeed, there is a close connection between the circulant matrix C and the Toeplitz matrix A; namely, the symbol of A is a continuous version of the eigenvalues of C. The solution to Eq. (A.7) is

$$\boldsymbol{q}(t) = e^{-t} e^{t g_\infty C} \boldsymbol{q}(0). \tag{A.17}$$

Since C is diagonalizable, $\|\boldsymbol{q}(t)\|_\infty$ is strictly bounded between 0 and ∞ as $t \to \infty$ if and only if the spectral abscissa of C is 1. The circulant matrix C has eigenvalues

$$\lambda_\mu = \sum_{k=-H}^{H} a_k e^{2\pi i k \mu / P}$$

with real parts

$$\Re\lambda_\mu = a_0 + \sum_{k=1}^{H} (a_k + a_{-k}) \cos\left(2\pi k \mu / P\right).$$

By the positive pairs condition, $\Re\lambda_\mu$ is maximized at $\mu = 0$ which yields $\max \Re\lambda_\mu = \sum_{-H}^{H} a_k$. Hence $q(t)$ is stable and bounded if and only if $g_\infty \max \Re\lambda_\mu = 1$ (assuming that $\sum_{-H}^{H} a_k \neq 0$, which is guaranteed by the finite $\bar{g}$ standard assumption).

For the second part, we note that since $\boldsymbol{q}$ is unitarily diagonalizable with first eigenvector equal to $\mathbf{1}/\sqrt{P}$, it follows that $\|\boldsymbol{q}(t)\|_2 = \frac{1}{P}\|(e^{-t+g_\infty \lambda_1 t}, ..., e^{-t+g_\infty \lambda_P t})^\top\|_2$. Taking P large, this becomes

$$\|\boldsymbol{q}(t)\|_2^2 = e^{-2t} \frac{1}{P} \sum_{\mu=1}^{P} |e^{2 t g_\infty \lambda_\mu}| \underset{P\to\infty}{\to} e^{-2t} \int_{-\pi}^{\pi} d\theta \, e^{2 t g_\infty \sum_k a_k \cos(k\theta)} \tag{A.18}$$

Using the positive pairs condition, we have that that the maximum of $\sum_k a_k \cos(k\theta)$ occurs at $\theta = 0$. Hence Laplace's method with large t yields

$$\int_{-\pi}^{\pi} d\theta \, e^{2 t g_\infty \sum_k a_k \cos(k\theta)} \sim \sqrt{\frac{\pi}{g_\infty t \sum_k a_k}} e^{2 t g_\infty \sum_k a_k}$$

It follows that

$$\|\boldsymbol{q}(t)\|_2 \sim t^{-1/4} e^{t(-1 + g_\infty/\bar{g})}. \tag{A.19}$$

□

Note that even in this simple case, it is not trivial to find the asymptotic behavior of $\|\boldsymbol{q}(t)\|_\infty$ (using the sup norm) since the sup norm and 2-norm are not equivalent in infinite dimensions.

We conjecture that stability of Eq. (A.4) is equivalent to Eq. (A.6) (under assumptions detailed in Conjecture 1 above). This does not, of course, automatically imply that Eq. (A.4) will be stable, as it is possible for Eq. (A.6) to not hold. Indeed, there is a clear situation that violates Eq. (A.6): since $g(t)$ is bounded from above by $\max_x G(x)$ (see SI Appendix 1), clearly Eq. (A.6) cannot be satisfied if $\bar{g} > \max_x G(x)$. In this case (and if our other assumptions hold), the nonlinear dynamics Eq. (A.4) are expected to not be stable (they will decay to 0 with (near) exponential rate).

However, if $\max_x G(x) \geq \bar{g}$, we can additionally argue that the nonlinear dynamics Eq. (A.4) will in fact be stable, under the additional assumption that $G(x)$ is monotonically decreasing from $x = 0$ (which is satisfied by a wide range of parameters for ϕ including those used in all of our simulations, see Fig. S1). This is because, if the network activity is vanishing, then the argument $\|A\boldsymbol{q}(t)\|^2 + \rho^2$ of G will approach ρ^2. This causes $g(t) = G(\|A\boldsymbol{q}(t)\|^2 + \rho^2)$ to grow. Once $g(t) \geq \bar{g}$, for each of these values of $g(t)$, say $g(t^*)$, the overlaps in the linear equation $\dot{q}_\mu = -q_\mu + g(t^*) \sum_{\nu=1}^{P} a_{\mu-\nu} q_\nu$ will grow. For large enough t, the linear equations closely approximate the nonlinear equations, so that the overlaps in the nonlinear equations will also grow. However, the overlaps for the nonlinear equations cannot grow indefinitely, since the nonlinearity ϕ bounds the activity of the network. Hence the overlaps will not decay to 0, nor will they grow without bound, so that Eq. (A.4) will be stable. Intuitively, the combatting forces of increased growth caused by decreasing $\|A\boldsymbol{q}(t)\|$ and bounded growth from the nonlinearity ϕ will in practice cause $\|A\boldsymbol{q}(t)\|^2 + \rho^2$ to settle at a stable value which coincides with $g(t) \to \bar{g}$.

Simple example: one forward connection. To demonstrate the theory derived above, we consider a simple example that is similar to Section 2.3 of the Supplementary Information of (1), and that is summarized in the One forward term subsection of the main text Results. This is the example with one sequi-associative and one autoassociative connection, i.e. where $a_k = 0$ for $k \notin \{0, 1\}$. For an even simpler, but slightly less informative example, one can take $a_0 = 0$ in the below calculations. The linear Eq. (A.4) with $g(t) = g_\infty$ for one sequi-associative and one autoassociative connection is

$$\begin{cases} \dot{q}_\mu = -q_\mu + g_\infty a_0 q_\mu + g_\infty a_1 q_{\mu-1}, & \mu > 1 \\ \dot{q}_\mu = -q_\mu + g_\infty a_0 q_\mu, & \mu = 1 \end{cases}$$

with initial conditions $q_1(0) = 1$ and $q_\mu(0) = 0$ for $\mu > 1$. The solutions to these equations for $\mu > 1$ are

$$q_\mu(t) = \frac{g_\infty^{\mu-1} a_1^{\mu-1} t^{\mu-1} e^{-t+g_\infty a_0 t}}{(\mu-1)!}.$$

Solving $\dot{q}_\mu(t) = 0$ for t yields peak times of $t_\mu = (\mu-1)/(1-g_\infty a_0)$. We can observe the growth properties of these solutions by evaluating q_μ at these peak values, yielding

$$q_{\mu+1}(t_{\mu+1}) = \frac{g_\infty^\mu a_1^\mu (\frac{\mu}{1-g_\infty a_0})^\mu e^{-\mu \frac{1-g_\infty a_0}{1-g_\infty a_0}}}{\mu!} = \frac{\mu^\mu (a_1 \frac{g_\infty}{1-g_\infty a_0})^\mu e^{-\mu}}{\mu!}. \quad [A.20]$$

where we offset the μ index by 1 to make the expression simpler. Sterling's approximation $\mu! \sim \sqrt{2\pi\mu}(\mu/e)^\mu$ yields

$$q_{\mu+1}(t_{\mu+1}) \sim \left(\frac{a_1 g_\infty}{1-g_\infty a_0}\right)^\mu (2\pi\mu)^{-1/2} = \frac{m^\mu}{\sqrt{2\pi\mu}} \quad [A.21]$$

where we have set $m := \frac{a_1 g_\infty}{1-g_\infty a_0}$. This experiences critical behavior at $m = 1$: for $m > 1$ the solutions grow without bound, while for $m < 1$ the solutions decay (nearly) exponentially quickly to 0. Solving $m = 1$ for g_∞ yields $g_\infty = 1/(a_0 + a_1)$, which is Eq. (A.6). At the critical value of $m = 1$ the solutions decay to zero, but do so with rate $\sim 1/\sqrt{\mu}$.

The periodic Eq. (A.7) with $g(t) = g_\infty$ and the same coefficient values as above has solution

$$\boldsymbol{q}(t) = \frac{1}{\sqrt{P}} W [e^{t\left(-1+g_\infty a_0 + g_\infty a_1 e^{2\mu\pi i/P}\right)}]_{\mu=0}^{P-1}$$

where W has columns $\boldsymbol{w}_\mu = [e^{-2\mu k\pi i/P}]_{k=0}^{P-1}$. The growth properties of this solution are revealed by the real part of $-1 + g_\infty a_0 + g_\infty a_1 e^{2\mu\pi i/P}$, which for $g_\infty a_1 > 0$ is maximized at $\mu = 0$ and realizes the maximum value $-1 + g_\infty a_0 + g_\infty a_1$. This is zero when $g_\infty = 1/(a_0 + a_1)$, so that this is a critical value for the solution similarly to the nonperiodic case (though in this case solutions neither decay to zero nor grow without bound when $g_\infty = 1/(a_0 + a_1)$).

Bidirectional connectivity – explicit solution. Here we derive an expression for a solution to the mean-field system with bidirectional connectivity. We include this derivation for completeness and to illustrate the difficulties that arise in deriving a simple exact expression for the solution. The reader may prefer to skip to the next sections showing our methods for finding approximate solutions. We start with the system of equations

$$\dot{q_\mu} = \begin{cases} -q_\mu + \bar{a}_0 q_\mu + \bar{a}_{-1} q_{\mu+1} + \bar{a}_1 q_{\mu-1}, & \mu > 1 \\ -q_\mu + \bar{a}_0 q_\mu + \bar{a}_{-1} q_{\mu+1}, & \mu = 1 \\ -q_\mu + \bar{a}_0 q_\mu + \bar{a}_1 q_{\mu-1}, & \mu = P. \end{cases} \quad [A.22]$$

In this case, the coefficient matrix A for the system written as $\dot{\boldsymbol{q}} = -\boldsymbol{q} + A\boldsymbol{q}$ has eigenvalues

$$\lambda_\mu = \bar{a}_0 + 2\sqrt{\bar{a}_{-1}\bar{a}_1}\cos\frac{\mu\pi}{P+1}, \quad \mu \in \{1, 2, \ldots, P\}$$

and eigenvectors $\boldsymbol{v}^\mu$:

$$v_k^\mu = \sqrt{\frac{2}{P+1}\left|\frac{\bar{a}_{-1}}{\bar{a}_1}\right|}(\bar{a}_1/\bar{a}_{-1})^{k/2}\sin\frac{\mu k\pi}{P+1}, \quad \mu, k \in \{1, 2, \ldots, P\} \quad [A.23]$$

The solution to Eq. (A.22) is then

$$\boldsymbol{q}(t) = VDV^{-1}\boldsymbol{q}(0). \quad [A.24]$$

where the columns of V are the eigenvectors $\boldsymbol{v}^\mu$ and where D is a diagonal matrix with $D_{\mu\mu} = e^{(-1+\lambda_\mu)t}$. Letting V^* denote the conjugate transpose of V, we see that VV^* is a diagonal matrix with the first entry being equal to 1. To see this, an easy calculation shows that

$$\sum_{\mu=1}^P v_1^\mu (v_1^\mu)^* = 1, \qquad \sum_{\mu=1}^P v_k^\mu (v_{k'}^\mu)^* = 0, \quad k \neq k'$$

where the asterisk $*$ denotes the complex conjugate. Hence, if we use the initial conditions $q_\mu(0) = \delta(\mu - 1)$, we have that $(V^{-1}\boldsymbol{q}(0))_k = (v_k^1)^*$. Hence we can write Eq. (A.24) as

$$\begin{aligned} q_\mu(t) &= \sum_j v_\mu^j \exp((-1+\lambda_j)t)(v_j^1)^* \\ &= \frac{2}{P+1}\left|\frac{\bar{a}_{-1}}{\bar{a}_1}\right|(\bar{a}_1/\bar{a}_{-1})^{\mu/2}\left(\sqrt{\bar{a}_1/\bar{a}_{-1}}\right)^* \sum_j \sin\frac{j\mu\pi}{P+1}\exp\left[\left(-1+\bar{a}_0+2\sqrt{a_{-1}a_1}\cos\frac{j\pi}{P+1}\right)t\right]\sin\frac{j\pi}{P+1} \end{aligned}$$

Taking $P \to \infty$ this becomes an integral

$$q_\mu(t) = 2\,|\bar{a}_{-1}/\bar{a}_1|\,(\bar{a}_1/\bar{a}_{-1})^{\mu/2}\left(\sqrt{\bar{a}_1/\bar{a}_{-1}}\right)^* e^{(-1+\bar{a}_0)t} K_\mu(2\sqrt{a_{-1}a_1}t)$$
$$= 2|\bar{a}_1/\bar{a}_{-1}|^{(\mu+1)/2-1} e^{(-1+\bar{a}_0)t} |K_\mu(2\sqrt{a_{-1}a_1}t)|$$

where

$$K_\mu(t) = \int_0^1 ds\, e^{t\cos s\pi} \sin \mu s\pi \sin s\pi.$$

Noting that

$$I_\mu(t) = K_\mu(t)t/\mu$$

is the modified Bessel function of the first kind, this equation can be written

$$q_\mu(t) = 2|\bar{a}_1/\bar{a}_{-1}|^{(\mu+1)/2-1} e^{(-1+\bar{a}_0)t} \mu t^{-1} |I_\mu(2\sqrt{a_{-1}a_1}t)|. \quad [A.25]$$

Instead of analyzing these equations further, we proceed with approximate methods that will generalize to more cases (i.e. more nonzero coefficients a_k).

Bidirectional connectivity – periodic boundary conditions. To further analyze the behavior, we convert the boundary conditions of Eq. (A.22) to periodic. This yields the following system of equations

$$\dot{q_\mu} = \begin{cases} -q_\mu + \bar{a}_0 q_\mu + \bar{a}_{-1} q_{\mu+1} + \bar{a}_1 q_{\mu-1}, & \mu > 1 \\ -q_\mu + \bar{a}_0 q_1 + \bar{a}_{-1} q_2 + \bar{a}_1 q_P, & \mu = 1. \\ -q_\mu + \bar{a}_0 q_P + \bar{a}_{-1} q_1 + \bar{a}_1 q_{P-1}, & \mu = P. \end{cases}$$

Note that the periodic equations may not match the solutions to the nonperiodic equations exactly, even as $P \to \infty$. This makes our simulations important for verifying that the approximations are useful.

The periodic equations can be written

$$\dot{\boldsymbol{q}} = -\boldsymbol{q} + A\boldsymbol{q}$$

where

$$A = \begin{pmatrix} \bar{a}_0 & \bar{a}_{-1} & 0 & 0 & \cdots & \bar{a}_1 \\ \bar{a}_1 & \bar{a}_0 & \bar{a}_{-1} & 0 & \cdots & 0 \\ 0 & \bar{a}_1 & \bar{a}_0 & \bar{a}_{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \bar{a}_1 & \bar{a}_0 & \bar{a}_{-1} \\ \bar{a}_{-1} & 0 & 0 & 0 & \bar{a}_1 & \bar{a}_0. \end{pmatrix}$$

Since A is circulant, it can be diagonalized by Fourier modes. Specifically, the μth eigenvalue and eigenvector of $-I + A$ can be written $\lambda_\mu = -1 + \bar{a}_0 + \bar{a}_{-1}e^{-2\mu\pi i/P} + \bar{a}_1 e^{2\mu\pi i/P}$ and $v_k^\mu = e^{-2\mu k\pi i/P}/\sqrt{P}$ for $\mu, k \in \{0, ..., P-1\}$. Using the initial condition $q_\mu(0) = \delta(\mu - 1)$, the solution can then be written

$$q_{\mu+1}(t) = \frac{1}{\sqrt{P}} \sum_{k=0}^{P-1} v_k^\mu \exp \lambda_k t \quad [A.26]$$

which becomes for $P \to \infty$

$$q_{\mu+1}(t) = \frac{e^{(-1+\bar{a}_0)t}}{2\pi} \int_0^{2\pi} e^{-i\mu s} e^{(\bar{a}_{-1}e^{-si} + \bar{a}_1 e^{si})t} ds.$$

Note that in taking $P \to \infty$, we still require that N is much larger than P so that the mean-field equations are valid. Simplifying the integrand results in

$$q_{\mu+1}(t) = \frac{e^{(-1+\bar{a}_0)t}}{2\pi} \int_0^{2\pi} e^{t(\bar{a}_1+\bar{a}_{-1})\cos s} e^{i(t(\bar{a}_1-\bar{a}_{-1})\sin s - \mu s)} ds.$$

and using the 2π-periodicity of the integrand

$$q_{\mu+1}(t) = \frac{e^{(-1+\bar{a}_0)t}}{2\pi} \int_{-\pi}^{\pi} e^{t(\bar{a}_1+\bar{a}_{-1})\cos s} e^{i(t(\bar{a}_1-\bar{a}_{-1})\sin s - \mu s)} ds. \quad [A.27]$$

We next have some choice of method for approximating this integral. A simple approach would involve taking a truncated Taylor expansion of the trigonometric functions in the integrand, with the hope that this resolves into an integral that can be evaluated. A more sophisticated and principled approach would be to use a saddle point approximation, where t is treated as a variable that is becoming large. We will start with saddle point approximations and show the limitations of this approach in this instance. We will then compare the result with the naïve Taylor expansion approach. To get the essential information, the reader may wish to skip to this Taylor expansion approach.

Bidirectional connectivity – saddle point approximation. We proceed by performing a saddle point approximation of the integral in Eq. (A.27). To do so, we write the integral as

$$I(t) = \int_{-\pi}^{\pi} e^{t\phi(s)} ds$$

where $\phi(s) = \beta \cos s + i\alpha \sin s - \gamma s$ and $\beta = \bar{a}_1 + \bar{a}_{-1}$ and $\alpha = \bar{a}_1 - \bar{a}_{-1}$ and $\gamma = \mu/t$. Note that we are treating μ/t as a finite scalar, so μ and t are growing large together.

For the saddle point approximation, we first extend ϕ to a function over the complex numbers and find the critical points of ϕ ($\Re\phi$ are saddle points at these critical points in the complex plane). When γ is large enough, there are two saddle points that both lie on the imaginary axis. To find these, we can take $s = iy$ and find the roots of $\phi'(iy)$. Letting $u = e^y$, $\phi'(iy)$ can be written $\phi'(iy) = -i\beta(u - 1/u)/2 + i\alpha(u + 1/u)/2 - i\gamma$. The roots of the resulting quadratic are

$$u_\pm = \left(\gamma \pm \sqrt{\beta^2 - \alpha^2 + \gamma^2}\right)/(\alpha - \beta).$$

Here we see that the condition that these roots exist is that $\gamma^2 \geq \beta^2 - \alpha^2$ since u must be real. We next need to calculate the angle of approach for these saddle points. This is given by $\arg \phi''$ evaluated at the saddle points. This shows that the angles are $-\pi/2$ and 0 for the saddle points in the $y > 0$ positive half-space and $y < 0$ negative half-space, respectively. The contribution of the saddle point in the positive half-space is pure imaginary, so we can focus on the saddle point in the negative half space, which we denote s^*. Given the negative root u_- above, it is straightforward to compute that

$$\phi(s^*) = \sqrt{\beta^2 - \alpha^2 + \gamma^2} + \log\left(\gamma - \frac{\sqrt{\beta^2 - \alpha^2 + \gamma^2}}{\alpha - \beta}\right), \qquad \phi''(s^*) = -\sqrt{\beta^2 - \alpha^2 + \gamma^2}.$$

The resulting saddle point approximation yields

$$I(t) \approx I_{\text{sp}}(t) := \sqrt{\frac{2\pi}{t|\phi''(s^*)|}} e^{t\phi(s^*)}$$

so that

$$q_{\mu+1}(t) \approx e^{(-1+\bar{a}_0)t} \sqrt{\frac{1}{2\pi t|\phi''(s^*)|}} e^{t\phi(s^*)}. \qquad \text{[A.28]}$$

To find the peak times, we seek the roots of $q'_\mu(t)$. Given the saddle point approximation above, it is straightforward to compute $q'_\mu(t)$. This expression, even in the asymptotic limit $t \to \infty$ (with μ/t constant), does not appear to have roots that can be expressed in an explicit equation for t.

Given that the roots of the saddle point approximation are unhelpfully complex expressions, we next try taking a saddle point approximation of $q'_{\mu+1}(t)$; i.e., instead of finding a saddle point approximation and then taking a derivative, we first take a derivative and then find the saddle point approximation. The derivative $q'_{\mu+1}(t)$ is

$$q'_{\mu+1}(t) = \frac{e^{(-1+\bar{a}_0)t}}{2\pi}\left((-1 + \bar{a}_0)I(t) + I'(t)\right)$$

where

$$I'(t) = \int_{-\pi}^{\pi} \psi(s) e^{t\phi(s)} ds, \qquad \psi(s) = \beta \cos s + i\alpha \sin s.$$

Taking a saddle point approximation of this involves taking a saddle point approximation of $I(t)$ and $I'(t)$; we have already done the former, and the latter is simply

$$I'(t) \approx \sqrt{\frac{2\pi}{t|\phi''(s^*)|}} \psi(s^*) e^{t\phi(s^*)}$$

where s^* is the critical point of ϕ as above. Hence

$$q'_{\mu+1}(t) \approx \sqrt{\frac{1}{2\pi t|\phi''(s^*)|}} e^{(-1+\bar{a}_0)t}\left((-1 + \bar{a}_0) + \psi(s^*)\right).$$

We again look for the peak times by finding the roots of this expression, which occur at $((-1 + \bar{a}_0) + \psi(s^*)) = 0$. Evaluating $\psi(s^*)$ is similar to evaluating $\phi''(s^*)$ and yields $\psi(s^*) = \sqrt{\beta^2 - \alpha^2 + \gamma^2}$. Substituting $\beta = \bar{a}_1 + \bar{a}_{-1}$, $\alpha = \bar{a}_1 - \bar{a}_{-1}$, $\bar{a}_0 = 1 - \bar{a}_1 - \bar{a}_{-1}$, and $\gamma = \mu/t$ into this expression and solving $(-1 + \bar{a}_0) + \sqrt{\beta^2 - \alpha^2 + \gamma^2} = 0$ for t yields

$$t_\mu \approx (\mu - 1)/(\bar{a}_1 - \bar{a}_{-1}). \qquad \text{[A.29]}$$

This is the same as Eq. (12) up to an added constant.

While this is a fair approximation, the differences between this estimate and the true peak times typically remain non-vanishing even as t and μ approach infinity; see the blue and orange lines of Fig. S2a and S2b. However, the differences between the true peak differences $d_\mu = t_\mu - t_{\mu-1}$ and this estimate do vanish with increasing t and μ; see the blue and orange lines of Fig. S2c and S2d.

Considering that the saddle point method yields either (1) an estimate for t_μ that doesn't appear to have a simple closed-form expression, or (2) an estimate that is not asymptotically precise, we seek a simpler approach. This simpler approach derived in the next section has the additional advantage of generalizing to the case of arbitrarily many nonzero coefficients a_k.

Bidirectional connectivity – Taylor expansion. In our next approach to approximating the integral Eq. (A.27), we expand the trigonometric functions around $s = 0$ to get

$$\begin{aligned} q_{\mu+1}(t) &\approx \frac{e^{(-1+\bar{a}_0)t}}{2\pi} \int_{-\pi}^{\pi} e^{t(\bar{a}_1+\bar{a}_{-1})(1-s^2/2)} e^{i(t(\bar{a}_1-\bar{a}_{-1})s-\mu s)} ds \\ &= \frac{e^{t(-1+\bar{a}_0+\bar{a}_1+\bar{a}_{-1})}}{2\pi} \int_{-\pi}^{\pi} e^{-s^2 t(\bar{a}_1+\bar{a}_{-1})/2} e^{is(t(\bar{a}_1-\bar{a}_{-1})-\mu)} ds \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-s^2 t(\bar{a}_1+\bar{a}_{-1})/2} e^{is(t(\bar{a}_1-\bar{a}_{-1})-\mu)} ds \end{aligned}$$

where in the last step we have used that $\sum_k \bar{a}_k = 1$. Letting $\alpha := \bar{a}_1 - \bar{a}_{-1}$ and $\beta := \bar{a}_1 + \bar{a}_{-1}$, we rewrite the integral as

$$q_{\mu+1}(t) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-(\beta t/2)s^2 + i(\alpha t - \mu)s} ds.$$

While we can write this integral in terms of error functions, a simpler expression results if we are able to integrate over the whole real line. We note that across a range of reasonable choices of the a_k, the integrand is vanishingly small outside of the interval $[-\pi, \pi]$. Hence we can indeed approximate the integral with one over the real line. This results in

$$\begin{aligned} q_{\mu+1}(t) &\approx \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(\beta t/2)s^2 + i(\alpha t - \mu)s} ds \\ &= \frac{1}{\sqrt{2\pi\beta t}} e^{-\frac{(\alpha t - \mu)^2}{2\beta t}} \end{aligned}$$

The locations t_μ of the peaks are given by the positive root of $\dot{q}_\mu(t) = 0$, which when using the approximation above yields

$$t_\mu \approx \frac{-\beta + \sqrt{\beta^2 + 4\alpha^2(\mu-1)^2}}{2\alpha^2}$$

Taking the asymptotic limit of large μ and taking the positive root yields Eq. (14):

$$t_\mu \approx \frac{\mu-1}{\alpha} - \frac{\beta}{2\alpha^2} = \frac{\mu-1}{\bar{a}_1 - \bar{a}_{-1}} - \frac{\bar{a}_{-1} + \bar{a}_1}{2(\bar{a}_1 - \bar{a}_{-1})^2}.$$

Similar to the estimate yielded by taking a saddle point approximation of $q'_\mu(t)$, this estimate is not asymptotically exact; see the green lines in Fig. S2a and S2b. Depending on the choice of a_k, the approximation can be better or worse than the saddle point approximation Eq. (A.29); compare Fig. S2a with Fig. S2b. However, as before the differences $d_\mu := t_\mu - t_{\mu-1}$ are asymptotically correct; see Fig. S2c and S2d. Hence we can consider this approximation to be practically at least as useful as that yielded by taking saddle point approximations. An extra benefit of this approach is that it generalizes; see the next section.

Generalizing the computation to general shift-invariant connectivity. The above technique of using Taylor expansions to find an approximate solution for q_μ extends to the general case of Eq. (10) with periodic boundary conditions in μ. As above, we assume that A is band-limited so that we can replace A with the circulant approximation of A. The eigenvalues of the circulant matrix $I - A$ are then $\lambda_\mu = -1 + \sum_{k=-\lceil P/2 \rceil + 1}^{\lceil P/2 \rceil - 1} \bar{a}_k e^{2\mu k \pi i / P}$. Grouping terms and taking the limit as $P \to \infty$ results in

$$e^{\lambda(s)t} = e^{(-1+\bar{a}_0)t} e^{t \sum_{k=1}^{\lceil P/2 \rceil - 1} c_k^+ \cos 2\pi k s + i c_k^- \sin 2\pi k s}$$

where $c_k^+ = \bar{a}_k + \bar{a}_{-k}$, $c_k^- = \bar{a}_k - \bar{a}_{-k}$, and $s \in [0, 1]$. Hence, using a diagonalization of circulant matrices as in Eq. (A.26) results in

$$q_\mu(t) \approx \frac{1}{P} \sum_{k=1}^{P} v_k^\mu \exp \lambda_k t \to \frac{1}{2\pi} \int_0^{2\pi} e^{t \sum_{k=1}^{\lceil P/2 \rceil - 1} c_k^+ \cos ks + i\left(t \sum_{k=1}^{\lceil P/2 \rceil - 1} c_k^- \sin ks - \mu s\right)} ds$$

as $P \to \infty$. Redefining the bounds of integration to $[-\pi, \pi]$ and expanding the trigonometric functions around $s = 0$ as before yields

$$q_\mu(t) \approx \frac{e^{(-1+\bar{a}_0)t}}{2\pi} \int_{-\pi}^{\pi} e^{t\sum_{k=1}^{\lceil P/2\rceil-1} c_k^+(1-(ks)^2/2)+i\left(t\sum_{k=1}^{\lceil P/2\rceil-1} c_k^- ks-\mu s\right)} ds = \frac{e^{t\left(-1+\sum_{k=-\lceil P/2\rceil+1}^{\lceil P/2\rceil-1} \bar{a}_k\right)}}{2\pi} \int_{-\pi}^{\pi} e^{-(\beta t/2)s^2+i(\alpha t-\mu)s} ds$$

where $\beta = \sum_{k=1}^{\lceil P/2\rceil-1} c_k^+ k^2$ and $\alpha = \sum_{k=1}^{\lceil P/2\rceil-1} c_k^- k$. Again extending this to an integral over the real line as before, as well as noting that $\sum_{k=-\lceil P/2\rceil+1}^{\lceil P/2\rceil-1} \bar{a}_k = 1$, yields the solution

$$q_{\mu+1}(t) \approx \frac{1}{\sqrt{2\pi\beta t}} e^{-\frac{(\alpha t-\mu)^2}{2\beta t}}.$$

As before, the positive root of $\dot{q}_\mu(t)$ is asymptotically

$$t_\mu \approx \frac{\mu-1}{\alpha} - \frac{\beta}{2\alpha^2}.$$

which is Eq. (13). This approximation will be used for the remainder of the Appendices.

Two forward terms. In the case of two forward terms, the system of equations is

$$\dot{q}_\mu = \begin{cases} -q_\mu + \bar{a}_0 q_\mu + \bar{a}_1 q_{\mu-1} + \bar{a}_2 q_{\mu-2}, & \mu > 2 \\ -q_\mu + \bar{a}_0 q_\mu + \bar{a}_1 q_{\mu-1}, & \mu = 2 \\ -q_\mu + \bar{a}_0 q_\mu, & \mu = 1. \end{cases}$$

Using Eq. (12) yields the approximate peak times of Eq. (15):

$$t_\mu \approx \frac{\mu-1}{\bar{a}_1 + 2\bar{a}_2} - \frac{\bar{a}_1 + 4\bar{a}_2}{2(\bar{a}_1 + 2\bar{a}_2)^2}.$$

Exponential kernels. Here we derive the peak times t_μ for exponential kernels of the form $w(t) = -m_1 e^{t/\tau_1}$ for $t < 0$ and $w(t) = m_2 e^{-t/\tau_2}$ otherwise. Given the approximation Eq. (12), we need only compute $\alpha = \sum_{k=1}^{P'} (\bar{a}_k - \bar{a}_{-k})k$ and $\beta = \sum_{k=1}^{P'} (\bar{a}_k + \bar{a}_{-k})k^2$. To start we adapt Eq. (6) to the setting of equal tutor intervals ($T_\xi^\mu = T_\xi$) so that the relationship between the coefficients a_k and the kernel w takes the form

$$a_k = \int_0^{T_\xi} dt \int_{kT_\xi}^{(k+1)T_\xi} ds\, w(s-t).$$

Then, for $k > 0$,

$$a_k = m_2 \int_0^{T_\xi} \int_{kT_\xi}^{(k+1)T_\xi} e^{-(s-t)/\tau_2}\, dsdt = m_2\tau_2^2 \left(e^{T_\xi/\tau_2} - 1\right)^2 e^{-(k+1)T_\xi/\tau_2} \quad [A.30]$$

$$a_{-k} = -m_1 \int_0^{T_\xi} \int_{-kT_\xi}^{(-k+1)T_\xi} e^{(s-t)/\tau_1}\, dsdt = -m_1\tau_1^2 \left(e^{T_\xi/\tau_1} - 1\right)^2 e^{-(k+1)T_\xi/\tau_1} \quad [A.31]$$

These equations afford an intuitive interpretation of the relationship between the kernel and the coefficients. For positive k, a_k increases with m_2 and doesn't depend on m_1 and τ_1. The coefficient a_k also increases with τ_2: after some algebra we find that the derivative of a_k with respect to τ_2 has the same sign as $e^{T_\xi/\tau_2} - 1 - bT_\xi/\tau_2$ where $b = \frac{(k+1)-e^{T_\xi/\tau_2}(k+1-2)}{2}$, which is positive provided $\tau_2 \le T_\xi$. An analogous relation holds for a_{-k}. From this we compute

$$\lim_{P\to\infty} \sum_{k=1}^{P-1} a_k = m_2\tau_2^2 \left(1 - e^{-T_\xi/\tau_2}\right)$$

$$\lim_{P\to\infty} \sum_{k=1}^{P-1} a_{-k} = m_1\tau_1^2 \left(1 - e^{-T_\xi/\tau_1}\right)$$

After computing

$$a_0 = -m_1 \int_0^{T_\xi} dt \int_0^t ds\, e^{t/\tau_1} + m_2 \int_0^{T_\xi} dt \int_t^{T_\xi} ds\, e^{-t/\tau_2} = T_\xi(m_2\tau_2 - m_1\tau_1) - m_2\tau_2^2 \left(1 - e^{-T_\xi/\tau_2}\right) - m_1\tau_1^2 \left(1 - e^{-T_\xi/\tau_1}\right) \quad [A.32]$$

we see that, for large P,

$$\bar{g} = \left(a_0 + \sum_{k=-P+1, k\neq 0}^{P-1} a_k \right)^{-1} \quad \rightarrow \quad T_\xi^{-1} \left(m_2 \tau_2 - m_1 \tau_1 \right)^{-1}.$$

Similarly,

$$\lim_{P\to\infty} \sum_{k=1}^{P-1} k a_k = m_2 \tau_2^2$$

$$\lim_{P\to\infty} \sum_{k=1}^{P-1} -k a_{-k} = -m_1 \tau_1^2.$$

Hence, for large P,

$$\alpha = \bar{g} \sum_{k=-\infty}^{\infty} k a_k = \bar{g} \left(m_1 \tau_1^2 + m_2 \tau_2^2 \right).$$

A similar calculation yields

$$\beta = \bar{g} \left(\frac{\tau_2^2 m_2 \left(e^{T_\xi/\tau_2} + 1 \right)}{\left(e^{T_\xi/\tau_2} - 1 \right)} - \frac{\tau_1^2 m_1 \left(e^{T_\xi/\tau_1} + 1 \right)}{\left(e^{T_\xi/\tau_1} - 1 \right)} \right). \qquad [A.33]$$

SI Appendix 3

Here we give the details of the network simulations. Differential equations are solved with forward Euler integration with a timestep of $dt = 0.075$, except for Fig. 1 which uses $dt = .0375$, Fig. 3e, which uses $dt = 1/15$, Fig. S2 which uses $dt = .001$, Fig. S3e-f and S4c-d which use $dt = 1/120$, and Figs. S5-S6 which use a different timestep for each line color (see Github repository linked in the main text for the precise values). Recall that the length of the input patterns is denoted by N and the number of patterns in the sequence is P. In Figs. 1 and 3 to 5 and Fig. S2, $N = 35000$. In Figs. 6 and 7, $N = 40000$. In Fig. 1, $P = 40$. In Figs. 3 to 5, 6b, 6c and 7 and Fig. S2, $P = 100$. In 6d, $P = 60$. In Figs. 2 to 6 and Fig. S2, $P = 100$.

Throughout, we use the nonlinearity Eq. (A.1) with $r_{\text{span}} = 2$, $r_{\text{center}} = 0$, $\theta = 0$, and $\sigma = 0.1$.

Error bars (shaded regions) are computed automatically by the Seaborn plotting library, which uses bootstrapping to compute 95% confidence intervals.

References

1. M Gillett, U Pereira, N Brunel, Characteristics of sequential activity in networks with temporally asymmetric hebbian learning. *Proc. Natl. Acad. Sci.* **117**, 29948–29958 (2020).
2. A Böttcher, SM Grudsky, *Spectral Properties of Banded Toeplitz Matrices.* (Society for Industrial and Applied Mathematics), (2005).
3. Gq Bi, Mm Poo, Synaptic Modification by Correlated Activity: Hebb's Postulate Revisited. *Annu. Rev. Neurosci.* **24**, 139–166 (2001).

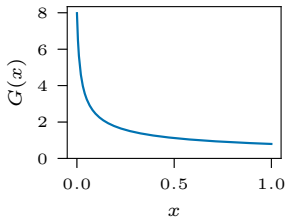

Fig. S1. The function $G(x)$ for $r_{\text{span}} = 2$, $\theta = 0$, and $\sigma = 0.1$. The function achieves a maximum of 8 at $x = 0$ ($x = 0$ is not a vertical asymptote).

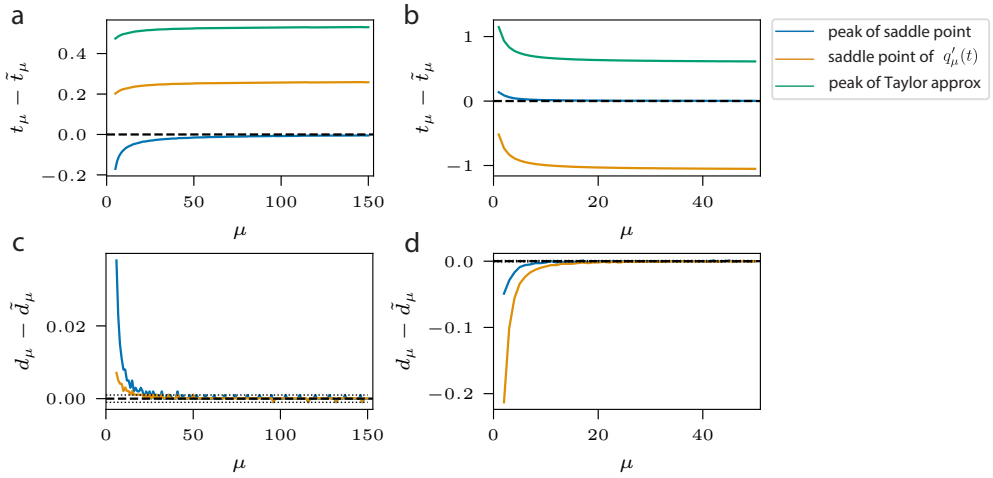

Fig. S2. Comparing approximations of t_μ to true peak times. In these plots, the differences between the true peak times t_μ and approximations $\tilde{t}_\mu$ are plotted across μ. Blue line, orange line, and green line correspond to the peaks of the saddle point approximation of $q_\mu(t)$, the roots of the saddle point approximation of $q'_\mu(t)$, and the peaks of the Taylor expansion approximation of $q_\mu(t)$, respectively. The python `scipy.signal` utility `find_peaks` is used to find the peak times numerically for the blue line. True peak times (dashed lines) are found by numerical quadrature of the integral expression in Eq. (A.2), followed by using `find_peaks`. In all plots, $\rho = 0$. **a**) Nonzero coefficient values $a_{-1} = -.3$, $a_0 = .1$, and $a_1 = 1.8$. **b**) Nonzero coefficient values $a_{-1} = .2$, $a_0 = .2$, and $a_1 = .8$. **c**)-**d**) Difference of peak differences $d_\mu = t_\mu - t_{\mu-1}$ with approximations $\tilde{d}_\mu$. Note that the orange and green lines coincide. Dotted lines denote the precision ceiling due to the step size $dt = 0.001$. **c**) Nonzero coefficient values $a_{-1} = -.3$, $a_0 = .1$, and $a_1 = 1.8$. **d**) Nonzero coefficient values $a_{-1} = .2$, $a_0 = .2$, and $a_1 = .8$.

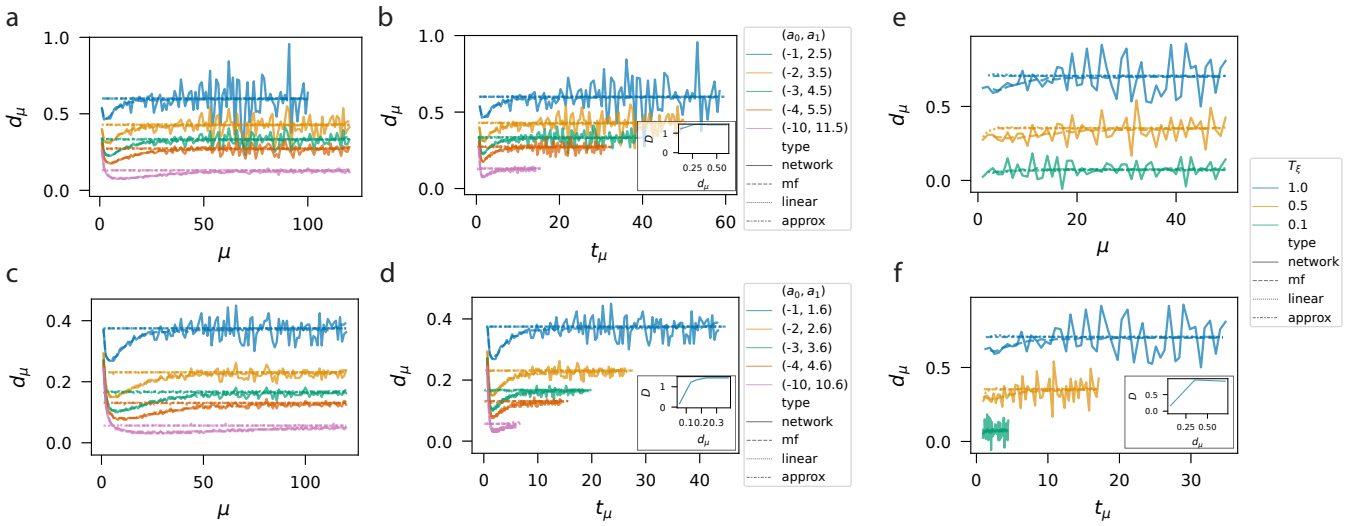

Fig. S3. Sequence dynamics over a large range of timescales. Peak differences d_μ as a function of μ and t_μ over a wide range of parameter combinations. Throughout, solid lines denote the full network simulations Eq. (1), dashed lines denote the mean-field simulations Eq. (9), dotted lines denote the linear approximate mean-field simulations Eq. (10), and the dotted-dashed lines denote the approximation given by Eq. (14). In all plots, $\rho = 0$. **a**)−**d**) Two nonzero terms a_0 and a_1. Color corresponds to coefficient combinations (a_0, a_1). Additional parameters: $N = 35,000$, $P = 140$. **a**), **c**) d_μ as a function of pattern index μ. **b**), **d**) d_μ as a function of peak time t_μ. **a**),**b**) Coefficient values sum to 1.5: $a_0 + a_1 = 1.5$. Corresponding legend is to the right of **b**. **c**),**d**) Coefficient values sum to $.6$: $a_0 + a_1 = .6$. Corresponding legend is to the right of **d**. **b**) inset shows the magnitude of the normalized integrated difference D between the mean field equation value of d_μ (dashed lines) and the baseline approximate value of d_μ (dot-dash lines) as a function of the approximate value of d_μ: $D = |(d^{\text{approx}})^{-1} \sum_\mu (d_\mu^{\text{mean field}} - d^{\text{approx}})|$. Note that each value of d^{approx} corresponds to a color in **b**. This measures the deviation of the mean-field tempo from the asymptotic tempo. **d**) description as in **b** but for different coefficient values. **e**)-**f**) Exponential kernel with parameters $\tau_1 = .25$, $m_1 = 2$, $\tau_2 = 1$, and $m_2 = 2$. Color denotes tutor signal interval T_ξ. Additional parameters: $N = 100,000$, $P = 80$. **e**) d_μ as a function of μ. **f**) d_μ as a function of t_μ. Inset as in **b**.

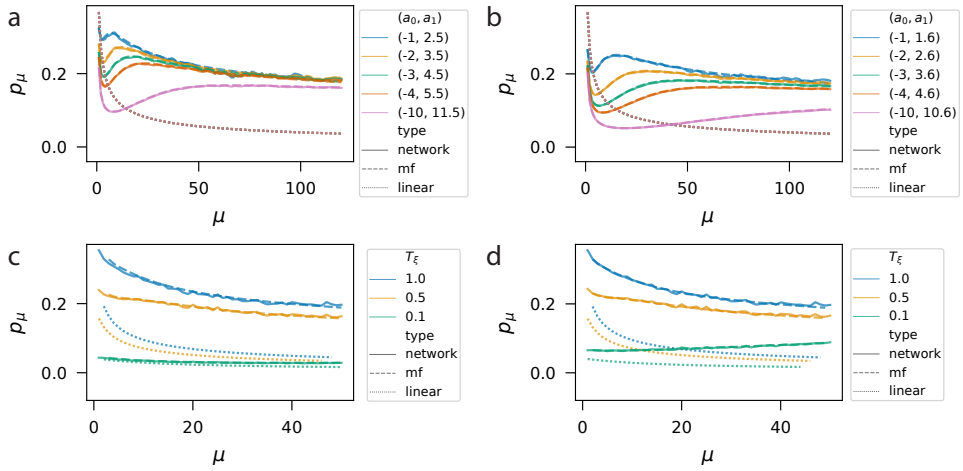

Fig. S4. Magnitude of sequences over a large range of timescales. Peak magnitudes p_μ as a function of μ over a wide range of parameter combinations. Throughout, solid lines denote the full network simulations Eq. (1), dashed lines denote the mean-field simulations Eq. (9), and dotted lines denote the linear approximate mean-field simulations Eq. (10). In all plots, $\rho = 0$. **a**)−**b**) Color corresponds to coefficient combinations (a_0, a_1). Additional parameters: $N = 35,000$, $P = 140$. **a**) Coefficient values sum to 1.5: $a_0 + a_1 = 1.5$. **b**) Coefficient values sum to $.6$: $a_0 + a_1 = .6$. **c**) Exponential kernel with parameters $\tau_1 = .25$, $m_1 = 2$, $\tau_2 = 1$, and $m_2 = 2$. Color denotes tutor signal interval T_ξ. **d**) As in **c**, but with the kernel magnitude doubled to $m_1 = m_2 = 4$. Additional parameters: $N = 100,000$, $P = 80$.

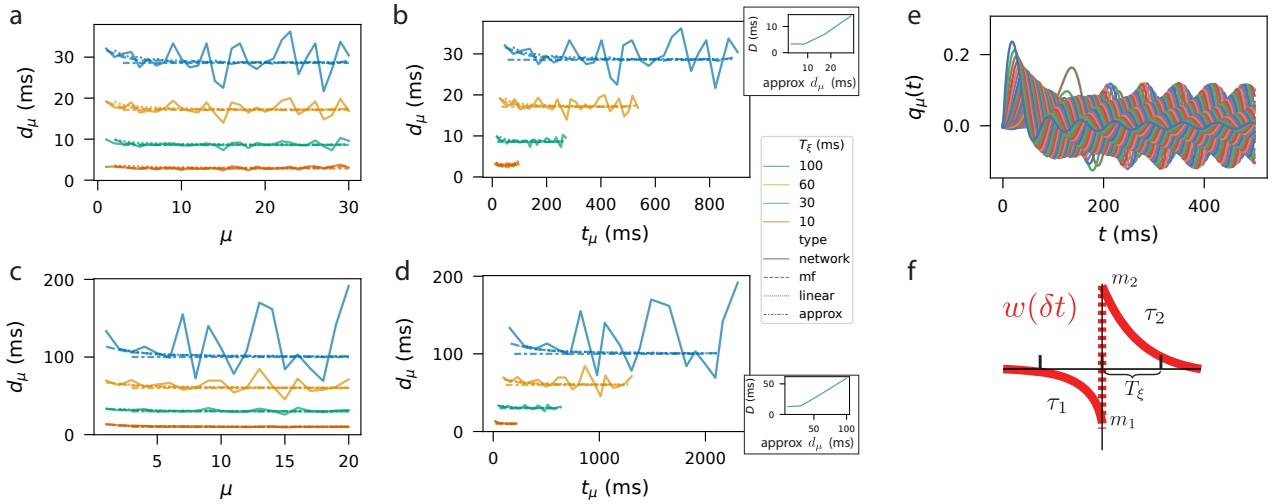

Fig. S5. Sequence dynamics over a large range of timescales for parameter values close to those measured in biology with double-sided exponential kernels. **a**)-**d**) Peak differences d_μ as a function of μ and t_μ for parameter values close to those measured in biology with double-sided exponential kernels. Throughout, solid lines denote the full network simulations Eq. (1), dashed lines denote the mean-field simulations Eq. (9), dotted lines denote the linear approximate mean-field simulations Eq. (10), and the dotted-dashed lines denote the approximation given by Eq. (14). Color denotes the tutor signal interval T_ξ (legend to right of panel c). Here we take a timescale of $\tau = 10$ milliseconds (ms) in Eq. (1), and throughout time is measured in milliseconds (ms). In all plots, $\rho = 0$. **a**)-**b**) Parameters for double-sided exponential kernel, adjusted from (3), are $\tau_1 = 16.8$ ms, $m_1 = 1092$, $\tau_2 = 16.8$ ms, and $m_2 = 3108$. Note that the absolute values of m_1 and m_2 are somewhat arbitrary since scaling them has a similar effect to multiple applications of the learning rule; their ratio is the more important factor, which we take to match that of (3). The large magnitude of m_1 and m_2 are needed to maintain stability for small T_ξ, but can be decreased for larger T_ξ. **a**) Peak difference d_μ as a function of pattern index μ. **b**) Peak difference d_μ as a function of peak time t_μ. Inset is as defined in Fig. S3b. **c**)-**d**) Parameters for double-sided exponential kernel are adjusted so that the tutor signal is stored faithfully, $\frac{m_2\tau_2 - m_1\tau_1}{m_1\tau_1^2 + m_2\tau_2^2}\tau = 1$. This results in $\tau_1 = 4.8$ ms, $m_1 = 1638$, $\tau_2 = 4.8$ ms, and $m_2 = 4662$. **c**) Peak difference d_μ as a function of pattern index μ. **d**) Peak difference d_μ as a function of peak time t_μ. Inset is as defined in Fig. S3b. **e**) Overlaps $q_\mu(t)$ for network with decay rates τ_1 and τ_2 taken directly from (3): $\tau_1 = 33.7$ ms, $m_1 = 1092$, $\tau_2 = 16.8$ ms, and $m_2 = 3108$. Color denotes μ from $\mu = 2$. **f**) Illustration of a double-sided exponential kernel, copied from Fig. 6a. Additional parameters: $N = 100,000$, $P = 40$.

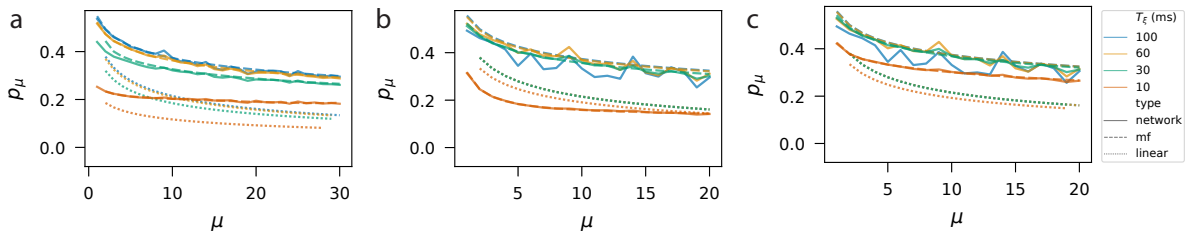

Fig. S6. Sequence magnitude over a large range of timescales for parameter values close to those measured in biology with double-sided exponential kernels. Peak magnitudes p_μ as a function of μ for parameter values close to those measured in biology with double-sided exponential kernels. Throughout, solid lines denote the full network simulations Eq. (1), dashed lines denote the mean-field simulations Eq. (9), and dotted lines denote the linear approximate mean-field simulations Eq. (10). Color denotes the tutor signal interval T_ξ. Here we take a timescale of $\tau = 10$ milliseconds (ms) in Eq. (1). In all plots, $\rho = 0$. **a**) Parameters for double-sided exponential kernel, adjusted from (3), are $\tau_1 = 16.8$ ms, $m_1 = 1092$, $\tau_2 = 16.8$ ms, and $m_2 = 3108$. Note that the absolute values of m_1 and m_2 are somewhat arbitrary since scaling them has a similar effect to multiple applications of the learning rule; their ratio is the more important factor, which we take to match that of (3). The large magnitude of m_1 and m_2 are needed to maintain stability for small T_ξ, but can be decreased for larger T_ξ. **b**) Parameters for double-sided exponential kernel are adjusted so that the tutor signal is stored faithfully, $\frac{m_2\tau_2 - m_1\tau_1}{m_1\tau_1^2 + m_2\tau_2^2}\tau = 1$. This results in $\tau_1 = 4.8$ ms, $m_1 = 1638$, $\tau_2 = 4.8$ ms, and $m_2 = 4662$. **c**) As in **b**, but with the kernel magnitude doubled to $m_1 = 3276$ and $m_2 = 9324$. Additional parameters: $N = 100,000$, $P = 40$.