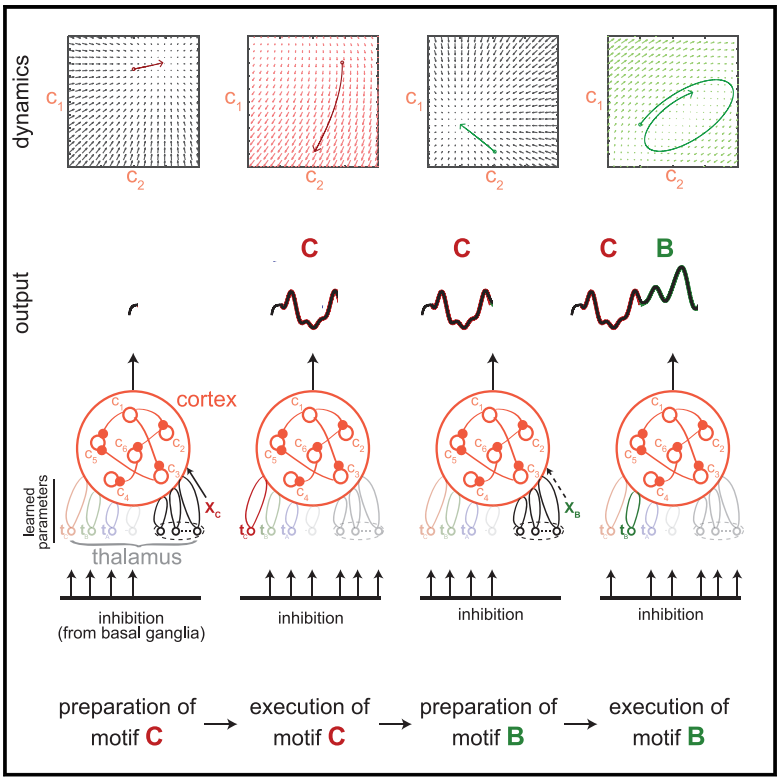
Cell Reports

Article

Thalamic control of cortical dynamics in a model of flexible motor sequencing

Graphical abstract



Authors

Laureline Logiaco, L.F. Abbott, Sean Escola

Correspondence

laureline.logiaco@gmail.com

In brief

Logiaco et al. develop a model of the thalamocortical architecture that supports learning of an extensible library of temporally varied motor motifs and chaining of these in arbitrary orders to control hierarchical behaviors. This relies on the surprising power of the relatively small thalamus to control the dynamics of the much larger cortex.

Highlights

- A single thalamocortical loop can vastly remap a large cortical circuit's dynamics
- Specific thalamic units can each drive distinct temporally complex motor outputs
- Another set of thalamic units can drive fast transitions between any two outputs
- This architecture yields extensible and flexible motor sequencing

Logiaco et al., 2021, Cell Reports *35*, 109090
June 1, 2021
https://doi.org/10.1016/j.celrep.2021.109090

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Thalamic control of cortical dynamics in a model of flexible motor sequencing

Laureline Logiaco,[1,2,3,*] L.F. Abbott,[1] and Sean Escola[2]
[1]Zuckerman Institute, Department of Neuroscience, Columbia University, New York, NY 10027, USA
[2]Zuckerman Institute, Department of Psychiatry, Columbia University, New York, NY 10027, USA
[3]Lead contact
*Correspondence: laureline.logiaco@gmail.com
https://doi.org/10.1016/j.celrep.2021.109090

SUMMARY

The neural mechanisms that generate an extensible library of motor motifs and flexibly string them into arbitrary sequences are unclear. We developed a model in which inhibitory basal ganglia output neurons project to thalamic units that are themselves bidirectionally connected to a recurrent cortical network. We model the basal ganglia inhibitory patterns as silencing some thalamic neurons while leaving others disinhibited and free to interact with cortex during specific motifs. We show that a small number of disinhibited thalamic neurons can control cortical dynamics to generate specific motor output in a noise-robust way. Additionally, a single "preparatory" thalamocortical network can produce fast cortical dynamics that support rapid transitions between any pair of learned motifs. If the thalamic units associated with each sequence component are segregated, many motor outputs can be learned without interference and then combined in arbitrary orders for the flexible production of long and complex motor sequences.

INTRODUCTION

Animals have the remarkable ability of flexibly performing long and complex sequences of movements (Geddes et al., 2018; Krakauer et al., 2019). In humans, dance illustrates this ability. A long dance can be decomposed into a sequence of short, stereotyped moves or motifs. These motifs form a library that can be flexibly combined by experienced dancers to create novel sequences—new dances—with minimal additional training. Furthermore, new motifs can be learned to extend the library without interfering with previously acquired dance moves. These phenomena raise important questions regarding sequence generation in the mammalian motor system that have not yet been addressed with computational models (Figure 1A). First, how is flexibility achieved? If the motor system could not generate sequences with previously unexperienced transitions between pairs of known motifs, this would severely limit flexibility. Second, how can a motif library be extended? New learning brings the risk of overwriting of prior knowledge, which the motor system must be robust to. Third, how can a high-level sequencing command—instructing which motifs to execute and in which order—be efficiently communicated to a neural network dedicated to the dynamic elaboration of the corresponding motor program? Of note, these are also open issues in machine learning (Merel et al., 2019b; Belkin et al., 2018; Geirhos et al., 2018; He and Jaeger, 2018; Riemer et al., 2019). We turn to the anatomy and physiology of the motor system for clues to and constraints on the answers to these questions.

We focus on the recurrent system comprising motor cortex, the basal ganglia, and thalamus, for which much evidence supports a role in the learning and execution of skilled motor sequences (Beloozerova and Marlinski, 2020; Geddes et al., 2018; Sauerbrei et al., 2020; Jin et al., 2014; Shenoy et al., 2013; Penhune and Steele, 2012; Jin and Costa, 2010; Mushiake and Strick, 1995). Although multiple other structures are important in motor control in general, these are particularly implicated in internally generated behaviors (Mushiake and Strick, 1995; van Donkelaar et al., 1999, 2000), the target of our investigation. The motor cortex appears to function as a dynamic pattern generator that produces neural activities needed to execute the muscle contractions associated with individual motifs (Churchland and Cunningham, 2014; Shenoy et al., 2013). The basal ganglia, on the other hand, have been linked to computations needed for arranging motifs into sequences. The striatum may control sequence structure (Geddes et al., 2018), and the output nuclei—the internal capsule of the globus pallidus (GPi) and substantia nigra pars reticulata (SNr)—generate sustained firing patterns that are specific to particular motifs (Jin et al., 2014). During sequence generation, the motor thalamus is typically considered to function as a relay, receiving strong inhibitory input from the basal ganglia (Edgerton and Jaeger, 2014; Deniau and Chevalier, 1985) and projecting to cortex (Harris et al., 2019). However, the motor thalamus also receives feedback from motor cortex (Harris et al., 2019). Thus, in addition to the conventional loop from cortex through the basal ganglia and thalamus and back to cortex, there is a second loop directly between cortex and thalamus. Although the former has long been studied (Mannella and Baldassarre, 2015; Alexander and Crutcher, 1990), only recently has the importance of the second loop been characterized (Rikhye et al., 2018; Schmitt et al., 2017;

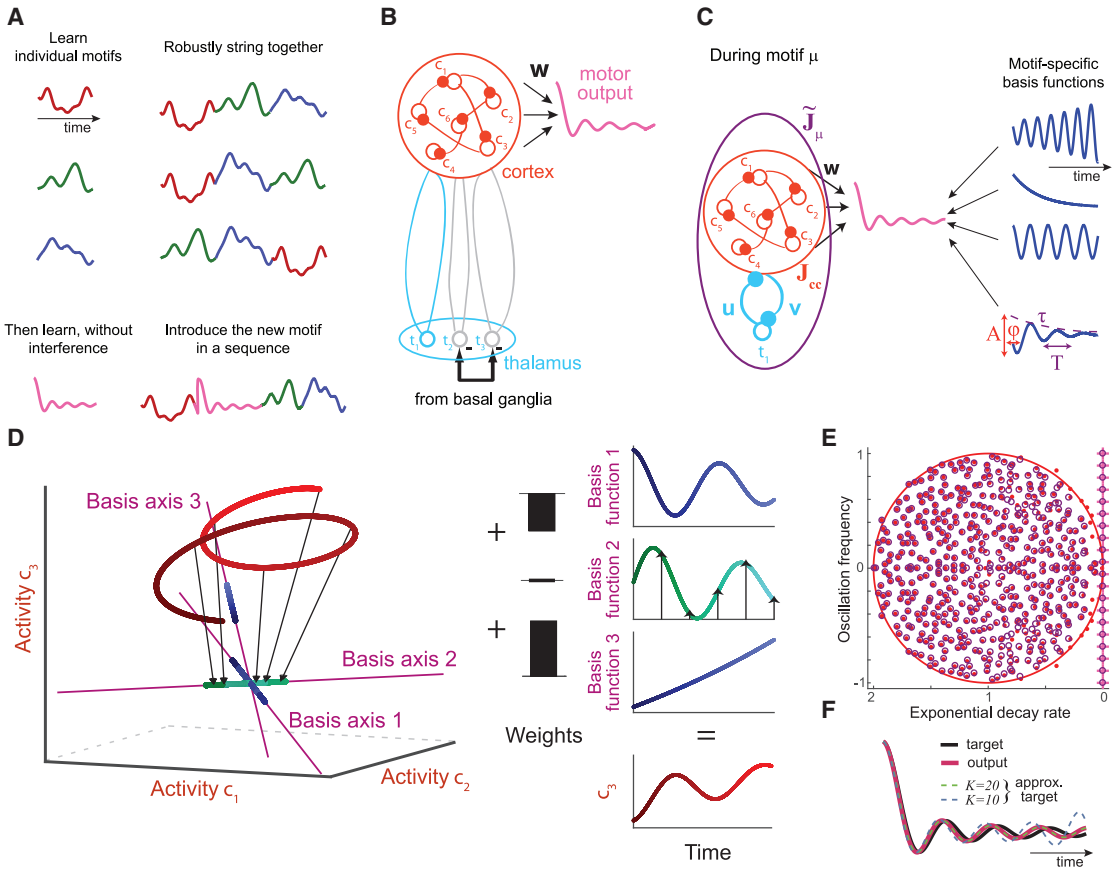


Figure 1. Motif sequencing and the cortex-thalamus circuit

(A) Flexible and extendable motor sequencing.

(B) Cortex-thalamus model schematic.

(C) Motif-specific dynamics. The effective connectivity matrix $\tilde{\mathbf{J}}_\mu$ defines the dynamics of motif μ (Equation 1), which can be summarized by an ensemble of basis functions (right). These basis functions are characterized by (1) exponential and oscillation timescales τ and T, which depend on the connectivity $\tilde{\mathbf{J}}_\mu$, and (2) initial amplitudes A and phases φ, which depend on the cortical activities at the start of motif μ. A sum of these basis functions composes the motor output (pink trace).

(D) (Left) Neural activity space for a 3-neuron network. The neural trajectory is shown in red. The basis functions are the projections of the neural activity onto specific directions in activity space (i.e., the eigenvectors in purple; STAR Methods section 4.2, with details in section 4.2.2.2). The arrows show the projection of the neural activity for the 2nd basis function. Dark to bright shades are early to late times. (Right) The activity of each neural unit (e.g., here, c_3) is a weighted sum of the basis functions.

(E) Control of the dynamics' timescale parameters (i.e., the eigenvalues) λ_i—related to τ and T in (C) through $\tau = 1/(Re\ \lambda_i - 1)$ and $T = 2\pi/Im\ \lambda_i$—by a thalamocortical loop during a motif. The target eigenvalues λ_k^{des} (for $k \leq 20$; pink "+" symbols) are included among the eigenvalues of the full cortex-thalamus circuit (purple circles) as a result of imposing the relationship between the corticothalamic and thalamocortical synaptic weights given by Equation 2. The corresponding eigenvalues of the unperturbed cortical network are shown for comparison (red dots), and the unit circle (indicating the bounds for an infinite random recurrent network; Bai, 1997) is shown for reference. Cortical network size: N = 500.

(F) Approximating a target motor motif (here, the sinc function; black curve) by a weighted sum of K basis functions (green dashed curve: K = 20; blue dashed curve: K = 10). The pink curve is the cortical output when the corresponding timescales (pink crosses in E) are included in the dynamics through tuning a thalamocortical loop (with K = 20).

See Figure S1 for other motifs.

Guo et al., 2017). Lacking, however, is a model of how motor function is supported by the interaction between these two loops.

We have developed a model incorporating thalamocortical loops controlled by basal ganglia outputs that can flexibly and extensibly generate sequences. Our model allows for a complete analysis that reveals the relationship between system output and the synaptic weights of the cortex-thalamus circuit. We find that the dynamics needed to execute a specific motif in a noise-robust manner can be generated solely by adjusting the synapses between cortex and motif-specific thalamic units, while leaving all synapses within cortex and to the motor output unchanged by learning. Furthermore, our analysis reveals the role for special purpose "preparatory" thalamocortical loops that robustly mediate fast transitions between motifs without requiring motif- or transition-specific synaptic weights.

To generate an arbitrary sequence, the basal ganglia switches between inhibiting different subsets of thalamic units to prepare

and execute different motifs. This framework ascribes to the basal ganglia the roles of sequence selection and transition timing and to the thalamocortical loop, in conjunction with motor cortex, the roles of motor preparation and motif execution. Our framework suggests that the cortex-thalamus architecture is well suited to flexibly control prolonged and complex sequential motor behavior.

RESULTS

Our aim is to develop an understanding of how cortical and subcortical motor areas cooperate to implement flexible motor sequencing. In particular, we wish to gain insight into how the structural anatomy of the motor system supports this complex function. In the development of our model, we are guided by three sources of evidence: anatomy; physiology; and the computational requirements of sequencing. These lead us to construct a model that is faithful to the known biology but simple enough to be analytically tractable and thus provide insights about the functioning of the motor system.

Our model comprises motor cortex, motor thalamus, and GPi/SNr (Figure 1B, left). Motor cortex is a highly recurrent structure (Kaneko, 2013), a feature of our model that is necessary to generate the dynamics needed to produce each motif. Furthermore, projection neurons in cortex synapse in the spinal cord for the control of muscles (Harrison et al., 2012), and thus, we model motor output as arising from cortical activity. Motor thalamus, conversely, has no excitatory recurrence (Arcelli et al., 1997), and this, as we shall see, is fundamental for the functioning of our model. It does, however, both project to and receive projections from motor cortex (Harris et al., 2019). Finally, we assume that GPi/SNr selects behavior via its inhibitory inputs to thalamus.

We rely on two major features of the physiology. First, previous experimental and theoretical work (Shenoy et al., 2013; Churchland and Cunningham, 2014; Sussillo et al., 2015) has provided strong evidence that motor cortex dynamics generates patterns of activity that form a basis for driving the muscle activity associated with movement. Specifically, these studies support the view, during individual movements, that activity in cortex can be captured by linear dynamics (Lara et al., 2018a; Sussillo et al., 2015) and that muscle activity can be reconstructed from a weighted sum of cortical activity (Russo et al., 2018). These results support using a linear network model during each motif. Second, activity patterns in GPi/SNr have been shown to be sustained during motif execution, with switches predominantly at the transitions between motifs (Jin et al., 2014). In our model, we assume that the synapses from basal ganglia to thalamus are strong enough such that, when active, they completely inhibit their thalamic targets (Edgerton and Jaeger, 2014)—which we assume to be a specific subset of thalamic neurons involved in learning new motifs. This assumption is further supported by the presence of recurrent inhibition within the thalamus (e.g., via the thalamic reticular nucleus; Arcelli et al., 1997), which—though ignored by our model—would only facilitate the ability of the basal ganglia to select among different thalamocortical loops via winner-take-all dynamics (Murray and Escola, 2017). This is compatible with the classic view that motor thalamus is in an inhibited state by default and is disinhibited by targeted removal of inhibitory input from the basal ganglia, which gates movement (Deniau and Chevalier, 1985; Edgerton and Jaeger, 2014; Kim et al., 2017; Aoki et al., 2019; but see Schwab et al., 2020 and STAR Methods, section 4.2.5). Importantly, this introduces a critical nonlinearity in our model by allowing thalamic units to be switched "on" and "off" across motifs. Specifically, inhibited thalamic units are silent while disinhibited ones are free to respond to their cortical inputs—in which case they participate to the motif-specific linear dynamics.

We further constrain our model by considering the computational requirement that robust motor sequencing must support the learning of new motifs without concern for interference with previously acquired ones. The structure of the cortex-thalamus circuit suggests a two-part solution. First, we restrict learning to the synapses within the thalamocortical loops and assume that the intracortical and output synapses are fixed. Second, we restrict the subsets of thalamic neurons that are active during different motifs to be non-overlapping. These constraints guarantee that motifs do not interfere and offer a procedure for adding new motifs: identify a new subset of thalamic units and set their synapses to and from cortex such that, when they are released from basal ganglia inhibition, the cortex-thalamus system drives the output to generate some new target movement (Figure 1B).

We can develop a mathematical description of our model with a final assumption that the dynamics in thalamus are more rapid than in cortex—which is expected given the absence of recurrent excitation in thalamus (Lim and Goldman, 2013; Seung et al., 2000)—and thus that we can approximate the thalamic response to cortical activity as instantaneous. (See the table in STAR Methods, section 4.2.5, for a complete list of the assumptions in our model and their supporting references and STAR Methods, section 4.2.6, for a demonstration that our results remain valid under more biologically plausible conditions.)

Together, the lines of evidence and assumptions above can be formalized within the standard "firing-rate" model framework (Gerstner et al., 2009; STAR Methods, section 4.2.5). The cortical activity, described by a vector $\mathbf{c}$, interacts with thalamic activity $\hat{\mathbf{t}}$ according to the rate equation $\dot{\mathbf{c}} = -\mathbf{c} + \mathbf{J}_{cc}\mathbf{c} + \mathbf{J}_{ct}\hat{\mathbf{t}}$, where $\mathbf{J}_{cc}$ and $\mathbf{J}_{ct}$ are the fixed intracortical and tunable thalamocortical synaptic weights, respectively. The cortical input to thalamus is $\mathbf{J}_{tc}\mathbf{c}$, where $\mathbf{J}_{tc}$ are the tunable corticothalamic weights. Consequently, the thalamic activity $\hat{\mathbf{t}}$—which accounts for inhibition from basal ganglia—is given as $\hat{\mathbf{t}} = \mathbf{S}_{\mu}\mathbf{J}_{tc}\mathbf{c}$, where $\mathbf{S}_{\mu}$ is a diagonal "selection matrix" whose only nonzero elements are "1"s at locations along the diagonal corresponding to thalamic units that are active during some motif μ. This selection matrix encodes the inhibition and disinhibition of thalamus caused by its basal ganglia inputs: multiplication by $\mathbf{S}_{\mu}$ sets the inhibited thalamic units to zero but leaves the disinhibited units free to respond to their cortical inputs. Note that the vector $\hat{\mathbf{t}}$ only models the thalamic units that are modulated by basal ganglia during sequence generation; other (non-plastic) thalamocortical loops may be absorbed into the non-plastic part of the network (STAR Methods, 4.2.5). Finally, the motor output y is generated as a weighted sum of the cortical activity with fixed output weights $\mathbf{w}$: $y = \mathbf{w}^T\mathbf{c}$.

The simplicity of this model allows us to combine our cortical and thalamic descriptions into a single closed equation for the cortical activity during motif μ:

$$\dot{\mathbf{c}} = -\mathbf{c} + \tilde{\mathbf{J}}_\mu \mathbf{c}, \quad \text{where} \quad \tilde{\mathbf{J}}_\mu \equiv \mathbf{J}_{cc} + \mathbf{J}_{ct}\mathbf{S}_\mu\mathbf{J}_{tc}. \qquad \text{(Equation 1)}$$

Thus, during a particular motif, the cortical dynamics are governed by the effective synaptic weights $\tilde{\mathbf{J}}_\mu$ given by the sum of the fixed intracortical connectivity $\mathbf{J}_{cc}$ and a corticothalamocortical perturbation $\mathbf{J}_{ct}\mathbf{S}_\mu\mathbf{J}_{tc}$. This perturbation, in turn, is determined by the pattern of inhibitory input into the thalamus from the basal ganglia as encoded in the diagonal elements of $\mathbf{S}_\mu$. Although we have presented these equations within the context of motif generation, the same framework can apply more generally. For instance, below, we will consider the dynamics of motor preparation as also arising from basal ganglia disinhibition of thalamocortical loops. In this case, $\mathbf{S}_{prep}$ will indicate the pattern of basal ganglia activity during preparation and the preparatory dynamics will be governed by $\mathbf{J}_{cc} + \mathbf{J}_{ct}\mathbf{S}_{prep}\mathbf{J}_{tc}$.

We briefly note major distinctions between our model and prior approaches. Several studies have modeled the neural and muscle activity of motor motifs with fully linear dynamical systems (Hennequin et al., 2014; Churchland and Cunningham, 2014; Lara et al., 2018a). However, these models generate a diversity of outputs only by setting motif-specific initial neural activities. Initial conditions also play an important role in defining our motifs (as discussed below), but by proposing that basal ganglia input changes the effective synaptic weights in cortex, our model benefits from increased expressivity, which we analytically characterize in the next section. Our proposal for the motor system—a switching linear model (Linderman et al., 2017)—falls into a larger category of approaches with nonlinear modulations of dynamics. However, by employing motif-specific thalamocortical loop weights, our model distinguishes itself by benefiting from rich shared cortical dynamics while still providing a mechanism for fine-tuning those dynamics per motif and avoiding interference between motifs.

We next address several key questions regarding the implementation of motor sequence generation by our model. First, what is the relationship between the corticothalamic and thalamocortical weights and the dynamics of the model, and what are the limitations to the dynamics that can be instantiated by tuning these weights? Second, how should these weights be set such that a particular motor output is generated? Third, can motif generation remain robust to noise in the system? Finally, a major question remains regarding the implementation of transitions between motifs. If the transitions between all possible pairs of motifs need to be learned, this would scale quadratically with the number of motifs. Can the motor system as represented in our model support efficient transitioning that avoids this poor scaling property? (We will show that, in fact, it can.)

Thalamus can act as a powerful modulator of cortical dynamics

A hypothesis of our model is that control of motif production arises from basal ganglia disinhibition of motif-specific thalamic units. We will show that this disinhibition can switch the dynamics of the motor system between a number of different configurations, one per motif, such that the thalamocortical network can implement a variety of motif-specific dynamics that enables robust and accurate motor outputs that go beyond the expressivity of the cortical network alone. Our cortical model is a stable recurrent linear network with fixed connectivity, mirroring previous models of motor cortical dynamics during reaches (e.g., Hennequin et al., 2014; Churchland and Cunningham, 2014; and Lara et al., 2018a). The dynamics of such a network is composed of an ensemble of oscillatory "basis functions" (Figures 1C and 1D; see table at the beginning of the STAR Methods section 4.2.2), whose temporal characteristics—oscillation frequency and decay rate—are typically limited to a specific range that depends on the statistics of the recurrent weight matrix (e.g., Figure 1E; and see Ahmadian et al., 2015; Bai, 1997; and Girko, 1985). For instance, if the cortical matrix is designed to favor stable dynamics, the basis functions will all decay much more rapidly (Hennequin et al., 2014) relative to the slower dynamics of many typical movements (Russo et al., 2018). In principle, these basis functions can be combined in different ways to generate a variety of motifs (Figures 1D and 1F), but—if there is a mismatch between the temporal characteristics of a desired motif and the temporal characteristics of the basis functions present in the cortical network—accurate motif production will require extremely large neural activities (such that the output relies on a "fine-tuned" cancellation of the activities; Figures S1F–S1L). In addition to being biologically implausible, such a solution would be highly sensitive to noise perturbations (Figures S1K and S1L). To generate such a motif robustly, the basis function ensemble must be modified such that basis functions with appropriate oscillatory and decay timescales are present within the dynamics (Figures S1A–S1E), and we propose to do this through thalamocortical loops.

Thalamic control of cortex requires that the activity of the large cortical network be altered by a much smaller number of thalamic units. To understand whether this is possible, we use the fact that our model is linear within each motif, which permits complete characterization of its dynamics. Specifically, the effective synaptic connectivity matrix $\tilde{\mathbf{J}}_\mu$ determines the basis functions of the combined thalamus-cortex system: their decay rates and oscillation frequencies are given by the real and imaginary parts of the "eigenvalues" of $\tilde{\mathbf{J}}_\mu$ (Figure 1C, right; Figure 1E; STAR Methods section 4.2.2.2). Thus, whether or not a motif can be realistically generated depends on the eigenvalues of $\tilde{\mathbf{J}}_\mu$ being set such that the motif can be constructed from a weighted sum of the basis functions (Figures 1C–1F) without requiring large neural activities. Because the neural activities directly relate to the weighting (STAR Methods, section 4.2.2.2), a plausible solution is one that constructs a motif by combining basis functions with relatively small weights. Although, for any motif μ, there is no unique appropriate set of basis functions, we wish to identify a small number of "desired" basis functions that, if present together in the basis function ensemble, would be sufficient for the network to accurately and robustly produce the motif with reasonable activities. Specifically, we wish to determine K decay rates and oscillation frequencies (i.e., eigenvalues) such that, when the corresponding basis functions are

summed, the result both faithfully reproduces the desired motif and does so under the constraint that the basis function weights are small (Figures S1A–S1E; STAR Methods). These two conditions enable accuracy, robustness, and plausible activities, while the condition of requiring a small K minimizes the number of basis functions that need to be introduced into the dynamics through thalamic control (Figures S1 and S2). Using these criteria, we can determine such a set of desired basis functions either numerically or, in some cases, analytically (STAR Methods, sections 4.2.2.2 and 4.2.7; Figure S1).

Given these considerations, we ask: can we efficiently employ thalamocortical loops in our model such that this set of desired basis functions that would enable the network to produce a given motif is certain to be present in the ensemble? To explore this, we consider the minimal case in which a single thalamic unit is left free to interact with cortex during motif μ while all others are inhibited by the basal ganglia. In this case, the effective connectivity matrix in Equation 1 simplifies to $\tilde{\mathbf{J}}_\mu = \mathbf{J}_{cc} + \mathbf{u}_\mu \mathbf{v}_\mu^T$, where $\mathbf{u}_\mu$ and $\mathbf{v}_\mu$ are the thalamocortical and corticothalamic synaptic weight vectors that are active during motif μ. We show, under light assumptions about the structure of $\mathbf{J}_{cc}$, that a relationship exists between $\mathbf{u}_\mu$ and $\mathbf{v}_\mu$ such that the ensemble of basis functions in our model is guaranteed to include the desired set (STAR Methods, section 4.2.2). Specifically, if the decay rates and oscillation frequencies of K desired basis functions are given by the real and imaginary parts of λ_i^{des} (for $i = 1, \ldots, K$) and the basis functions of the unperturbed cortical network $\mathbf{J}_{cc}$ are given by the N eigenvalues λ_i, the relationship is

$$\mathbf{v}_\mu = \mathbf{L}^T \mathrm{diag}(\mathbf{L}\,\mathbf{u}_\mu)^{-1} \mathbf{P}^+ \mathbb{1}, \qquad \text{(Equation 2)}$$

where the rows of $\mathbf{L}$ describe the directions in the high dimensional space of the neural activity along which the basis functions of the unperturbed cortical network lie (i.e., $\mathbf{L}$ is a matrix of the left eigenvectors of $\mathbf{J}_{cc}$) and $\mathbf{P}^+$ is the pseudoinverse of a matrix with elements $P_{ij} = 1/(\lambda_i^{des} - \lambda_j)$. For a given motif, our goal of introducing K desired basis functions into the dynamics in principle imposes only K constraints on the $2N$ parameters defining $\mathbf{u}_\mu$ and $\mathbf{v}_\mu$. A sufficient way to satisfy these constraints is to impose N relationships between $\mathbf{u}_\mu$ and $\mathbf{v}_\mu$ (through the pseudoinverse of $\mathbf{P}$; see Equation 12 in STAR Methods). The choice of how to make use of these N relationships is arbitrary; one can, for example, choose to completely define $\mathbf{v}_\mu$ in terms of a random $\mathbf{u}_\mu$ (as in Equation 2 above and Figures 2A–2C) or vice versa. In the following section, we will take advantage of the remaining N degrees of freedom to increase the robustness of the solution.

Figure 1E shows the λ_i^{des} for a set of desired basis functions and the λ_i for a random unperturbed cortical network (pink crosses and red dots, respectively). After setting the corticothalamic weights $\mathbf{v}_\mu$ according to Equation 2 (with any random thalamocortical weights $\mathbf{u}_\mu$), the cortex-thalamus system contains the desired basis functions (i.e., the eigenvalues of the effective connectivity matrix $\tilde{\mathbf{J}}_\mu$ include the λ_i^{des}; purple circles).

Equation 2 shows that, in principle, setting only N thalamocortical synaptic weights is sufficient to control all the timescales of the basis functions of the model cortex, a system with a much larger N^2 synaptic weights. However, there are limits to this control. If the number of desired basis functions is large or if their timescales differ too much from those of the unperturbed cortical network, it may be numerically difficult to satisfy Equation 2 (STAR Methods, section 4.2.2; Figure S2). This could, in principle, impose an effective constraint on the motifs that can be generated. Fortunately, a wide variety of motifs—including oscillatory signals resembling, for example, the muscle activities seen during primate behavior (Churchland et al., 2012; Russo et al., 2018)—can be well described by a relatively small number of basis functions (i.e., $K \ll N$; see Figure S1). In this case, obtaining the desired basis functions is stable (Figures 1E, S2, and S5) with synaptic weights within the thalamocortical loop that are similar in magnitude to the recurrent cortical weights (Figures 2C, 2F, and S4). Furthermore, the changes in the cortical timescales induced by individual motif-specific thalamocortical loops are sufficient to dramatically and qualitatively expand the diversity of the motifs that the network can produce (Figures S1 and 1E). Thus, despite the obvious challenge of remapping the dynamics of the large cortical network with a single thalamocortical loop, the use of thalamocortical control loops greatly extends the dynamics that could be achieved by using an unmodulated cortex model.

In summary, our analysis demonstrates that a single thalamocortical loop can powerfully modulate cortex by introducing into the circuit dynamics a set of useful basis functions for a specific motif.

Taming sensitivity to initial conditions

Generating the right set of basis functions within the cortical dynamics is necessary to produce a desired motif but omits a key remaining step: setting the initial cortical activities (Churchland et al., 2006; Shenoy et al., 2013; Churchland and Cunningham, 2014; Elsayed et al., 2016; Lara et al., 2018b; Zimnik and Churchland, 2021). In our model, these initial activities determine the amplitude—and, for oscillating functions, phase—of each basis function in the subsequent dynamics (Figures 1C and 1D). Because each basis function has its own relative contribution to the motor output—governed by the geometrical orientation of the basis functions and the output weights in the space of the dynamics—setting the amplitudes and phases appropriately is necessary to ensure that the summed output gives a desired motif (Equation 15 in STAR Methods, section 4.2.2.2). Here, we study the robustness of motor output to the presence of noise in the initial conditions. This is fundamental because any biologically plausible mechanism for setting the initial conditions must be able to tolerate the large neural variability observed in behaving animals (Churchland et al., 2010).

By expressing a desired motif as a weighted sum of basis functions, we can calculate the initial cortical activities $\mathbf{c}_\mu^{init}$ (STAR Methods, Equation 16) that will produce a given motif if set exactly. However, if the thalamocortical $(\mathbf{u}_\mu)$ and corticothalamic $(\mathbf{v}_\mu)$ weights satisfy the constraints given in the previous section (Equation 2) but are not further constrained (i.e., $\mathbf{u}_\mu$ is random), the system output can be catastrophically affected if the initial activity pattern is perturbed even slightly (Figure 2A). This occurs even in the stable regime where all basis functions in the network undergo decay (Figure 1E).

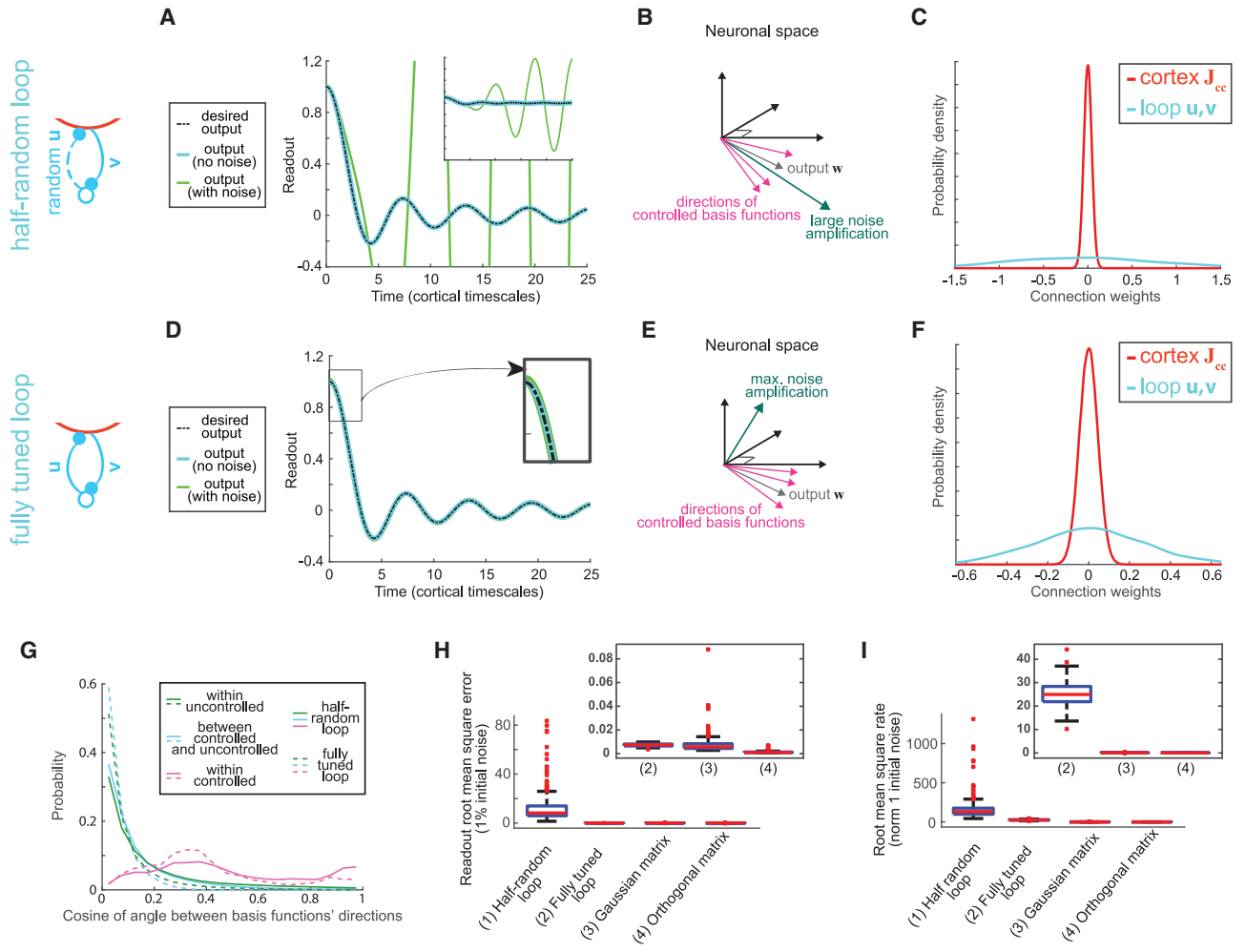


Figure 2. Noise-robust motif execution

(A–C) Corticothalamic weights $\mathbf{v}_\mu$ adjusted to control the dynamics' timescales as in Figure 1E (with Equation 2), while the thalamocortical weights $\mathbf{u}_\mu$ are random (as indicated with the dashed line in the schematic on the left).

(A) Desired and actual system outputs in the presence and absence of noise in the initial cortical activities ($\mathbf{c}_\mu^{\text{init}}$ from Equation 16). Noise is Gaussian i.i.d. with standard deviation equal to 1% of the root-mean-squared norm of the activities in the noiseless setting.

(B) Amplification of noise due to alignment of the directions of the basis functions. If the output weights $\mathbf{w}$ and some directions of large noise amplification overlap, the output is highly noise sensitive.

(C) Distributions of intracortical and loop weights with $\mathbf{u}_\mu$ sampled from a centered Gaussian distribution scaled such that $\mathbf{u}_\mu$ and $\mathbf{v}_\mu$ have equal norm (STAR Methods, section 4.2.5).

(D–F) Same as (A)–(C) but with the thalamocortical weights $\mathbf{u}_\mu$ optimized to minimize the effect of noise in the initial activity. Noise has a negligible effect on the output (D) because all directions of large noise amplification and the output vector are no longer aligned (E). Additionally, the distribution of the loop weights narrows (F).

(G) Distributions of the cosine of the angles between pairs of basis functions' directions (i.e., eigenvectors). Larger values indicate that a pair of directions are nearly parallel (STAR Methods, section 4.2.3).

(H) Root-mean-squared error of the output in the presence of 1% noise. Compared to the case of random $\mathbf{u}_\mu$ (1), the noise-induced error is substantially diminished after optimization (2) and is on par with errors observed using matrices with the same timescales as $\bar{\mathbf{J}}_\mu$ but with basis function directions that match those of random Gaussian matrices (3) or are orthogonal (4). See STAR Methods, section 4.2.3 for details.

(I) Average root-mean-squared norm of the activity vector $\mathbf{c}$ when the initial activity is sampled from a random noise vector of norm one (from Equation 20 in STAR Methods). Optimizing $\mathbf{u}_\mu$ reduces the length of the activity vector (1 versus 2), but not to lengths observed with random or orthogonal directions of basis functions (3 and 4), indicating that optimization does not fully eliminate non-normal amplification.

The data in (G)–(I) were generated from 50 random networks (i.e., random samples of $\mathbf{J}_{cc}$ and $\mathbf{w}$) with five randomly sampled $\mathbf{u}_\mu$ per network; (H) and (I) used five matrices with random and orthogonal directions of basis functions per network. N = 500 throughout. Boxplots indicate the median (red line), the 25th and 75th percentiles (edges of the box), the range of the data beyond these percentiles while staying no more than 1.5 times the interquartile interval away (whiskers), and outliers (crosses).

We can understand this through our analysis which reveals that, in order for a motif to be produced in a noise-robust way with reasonable neural activities, the requirement that the motif be decomposable into a sum of basis functions with small weights is incomplete. This requirement would be sufficient to ensure reasonable magnitude neural activities if the geometry of the network dynamics was "regular" (i.e., if the directions of the basis functions were close to orthogonal). If the basis functions are aligned, however, this sets up the potential for the neural activities—which themselves can be written as sums of basis functions—to require large weights and thus have large magnitudes (Figure 1D; Equations 16, 20, and 21 in STAR Methods). Then, as the dynamics evolve, large noise-amplifying activity transients can be observed, a phenomenon known as non-normal amplification (Ganguli et al., 2008; Murphy and Miller, 2009; Hennequin et al., 2012; Bondanelli and Ostojic, 2020). Indeed, in our case, we show that, by forcing the dynamics to include our desired basis functions via Equation 2, the directions of the basis functions will tend to be aligned with each other. Thus, it is imperative that the geometry of the cortical dynamics can be constrained through thalamic influence.

Fortunately, the conditions on the thalamocortical and corticothalamic weights required to generate a set of basis functions with desired decay rates and oscillation frequencies (Equation 2) do not completely specify both sets of these weights. This raises the possibility that they can be selected to make the motor output more robust to deviations in the initial conditions by controlling the geometry of the dynamics. In fact, noise robustness is only needed in the directions of the dynamics that are aligned with the output weights $\mathbf{w}$. Our model permits an analytic calculation of the output error in the presence of initial noise as a function of the thalamocortical weights, and therefore, we optimized these weights to minimize this error (STAR Methods, section 4.2.3). Importantly, this optimization does not affect the decay rates and oscillation frequencies of the basis functions but instead exclusively acts on the directions along which the basis functions lie and thus the properties of non-normal amplification in the system (STAR Methods, sections 4.2.2 and 4.2.3). After optimization, we find that production of the motor output is robust to initial noise (Figure 2D), with errors on par with those seen using "control" connectivity matrices that have the same basis functions as $\tilde{\mathbf{J}}_\mu$ but are constructed to have little or no non-normal amplification (Figure 2H). This is achieved even though the controlled basis functions stay relatively aligned (Figure 2G), meaning that the remaining non-normal amplification is restricted to non-output dimensions (Figures 2E, 2I, S3D, and S3E; STAR Methods, section 4.2.3). Therefore, after full optimization of the thalamocortical loop, the activities in the network are of moderate magnitudes (Figures 3E and 3F), with norms (Figure S3E) that indicate that the directions of larger activity patterns tend to not fully align with the readout weights (which themselves have unit norm). Likewise, moderately large neural activity patterns in movement-irrelevant dimensions have been observed in motor cortical dynamics (Russo et al., 2018; Saxena et al., 2021). Finally, we find that noise-robust solutions lead to smaller magnitudes of the thalamocortical loop weights compared to unoptimized networks (Figures 2C and 2F).

These results show that adjusting both the corticothalamic and thalamocortical synaptic weights is a mechanism by which a thalamocortical loop can achieve the dual goals of sculpting the cortical dynamics to provide the basis functions for motif production and stabilizing the motor output with respect to deviations in the neural activity.

Thalamocortical loops can prepare cortex to execute each motif

We now turn to the question of how the cortex-thalamus system can prepare the initial activities needed to generate a motif (Churchland et al., 2006; Shenoy et al., 2013; Churchland and Cunningham, 2014; Elsayed et al., 2016; Lara et al., 2018b). This preparation should not depend on the cortical activity at the start of the preparatory period or else the motor system would need to explicitly learn the transitions between all pairs of motifs, producing a quadratic dependence on the number of motifs and preventing improvisation of new sequences. Additionally, it should be possible to achieve preparation quickly (Lara et al., 2018b). Here, we assume that an input that specifies the upcoming motif can be delivered to the motor system during motor preparation. This input could arise, for example, from frontal cortex and reflect goals (Russo et al., 2020; Kornysheva and Diedrichsen, 2014) or from the sensory system to cue a behavioral response to a stimulus (Dacre et al., 2019). We also assume that the basal ganglia (Jin and Costa, 2010; Jin et al., 2014) can select preparation-specific thalamocortical loops that are active only during the preparatory period. Importantly, these preparatory loops can be reused for the transitions between all pairs of motifs. Thus, we propose that preparation is split computationally into two components: the specificity for the upcoming motif is determined by a motif-specific input, while the preparation dynamics are shaped generically through thalamic modulation of cortex (Figure 3A).

Turning to our model, the effective connectivity of the cortex-thalamus system during preparation is given by $\mathbf{J}_{\text{prep}} = \mathbf{J}_{\text{cc}} + \mathbf{J}_{\text{ct}}\mathbf{S}_{\text{prep}}\mathbf{J}_{\text{tc}}$, where $\mathbf{S}_{\text{prep}}$ determines the thalamic units that are active during preparation (Figure 3A). Then, if $\mathbf{x}_\mu$ is the input specific to upcoming motif μ, the cortical dynamics are given by $\dot{\mathbf{c}} = -\mathbf{c} + \mathbf{J}_{\text{prep}}\mathbf{c} + \mathbf{x}_\mu$. With these dynamics (which we can ensure are stable) and if $\mathbf{x}_\mu$ is set appropriately, the cortical activity will converge to the desired initial state for the upcoming motif $\mathbf{c}_\mu^{\text{init}}$ (Equation 24 in STAR Methods, section 4.2.4).

This "decay to steady state" mechanism is independent of the cortical activity both at the start of preparation (which is the end of the previous motif) and throughout the new motif. Thus, a single preparatory network can serve all possible motif transitions despite having no synaptic weights that are trained on specific transitions.

The speed of preparation depends on how quickly the desired initial activity state can be reached. Because the dynamics of convergence to steady state are independent of the input (Equation 26 in STAR Methods, section 4.2.4), if thalamocortical loops generate fast decay dynamics, they can be used for quick preparation of any upcoming motif. Tuning of thalamocortical loops is required because, when the unperturbed cortical network is chosen to include slow dynamics to aid in motif generation, its

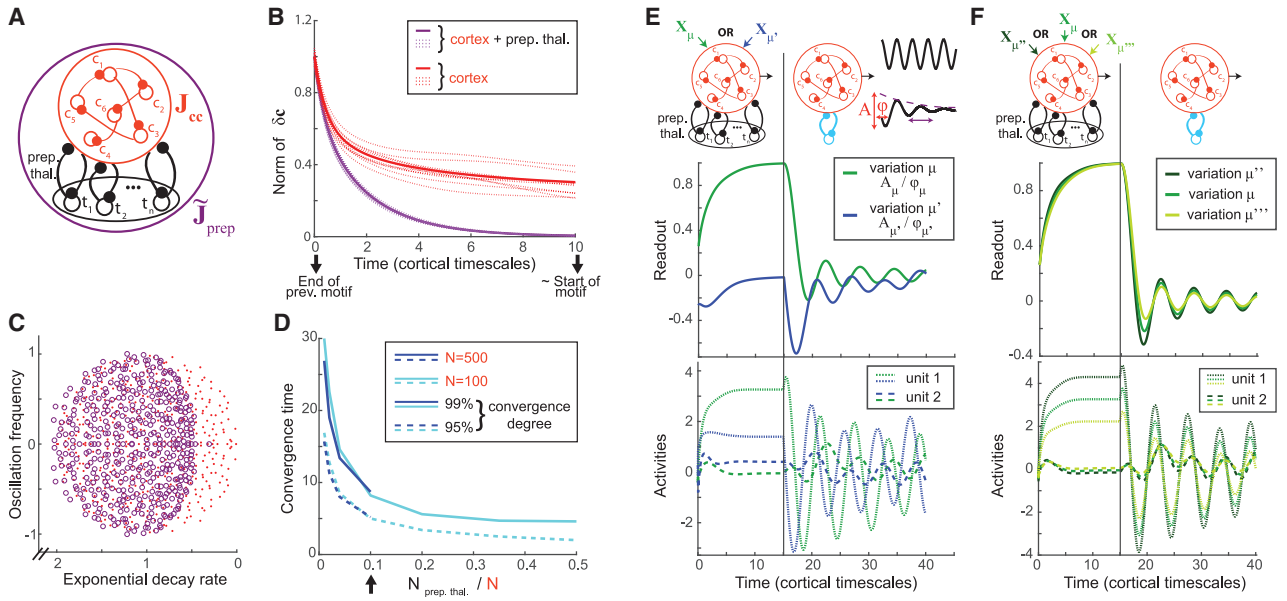


Figure 3. Motif preparation

(A) Effective thalamocortical network during the preparatory period.

(B) Decay of the magnitude of the difference between the cortical rates and their steady state for an optimized thalamocortical preparatory network (solid purple: average; dashed purple: individual trials) and for the unperturbed cortical network (red).

(C) Timescale parameters of the optimized preparatory cortex-thalamus network (purple circles), compared to those of the unperturbed cortical network (red dots). Conventions as in Figure 1E. See also Figures S6D and S6J.

(D) Time for the cortical activity vector to decay by either 95% or 99% of its initial value as a function of the number of thalamic units used in the preparatory network relative to the number of cortical units. For a fixed proportion of thalamic units, the decay time is similar between different numbers of cortical units. The arrow indicates the proportion of thalamic neurons used in other panels.

(E) Preparing and executing two motif variations μ (which is identical to the output in Figure 2) and μ' when using the same thalamocortical loop during motif execution. Top: schematic of the thalamocortical circuits involved. Example basis functions are also shown—for each motif variation, the basis functions' exponential decay rates and oscillation frequencies are fixed, but their initial amplitude A and phase φ are set via $\mathbf{x}$. Middle: network output. Bottom: two example cortical units. Network activities were initialized with random standard normal values.

(F) Three sample motif preparations and executions when systematically varying the neural activities at motif start along a line in neural space. Conventions are as in (E).

N = 500 for (A)–(C), (E), and (F). In all panels except (D), the thalamic population has 10% of the number of cortical units.

intrinsic convergence to steady state is too slow to support efficient preparation (Figure 3B, red).

To find corticothalamic and thalamocortical weights that encourage rapid preparation, we optimized the relevant weights of $\mathbf{J}_{ct}$ and $\mathbf{J}_{tc}$ to minimize the convergence time while ensuring smoothness in the network output (STAR Methods, section 4.2.4; Figure S6). Using a thalamic population that has 10% of the number of cortical units, convergence to steady state occurs in just a few time constants of the cortical rate model (Figure 3B, purple). This occurs because all basis functions of the preparatory network have fast exponential decay rates (Figures 3C, S6D, and S6J). Increasing the size of the thalamic preparatory population leads to faster convergence times (Figure 3D). Furthermore, after optimization, the thalamocortical loop weights can have magnitudes that are comparable to the intracortical weights (STAR Methods, section 4.2.4; Figures S6E and S6K). The size of the thalamic population needed for preparation is considerably larger than what is needed for the production of any single motif. This is because, during production, the readout is the only constrained direction, whereas during preparation, all units are constrained.

This preparatory mechanism suggests that movements can be divided into two categories. In the previous sections, we have considered the case where different movements require different dynamics, corresponding to different basis functions, that are implemented by different thalamocortical loops. There is also the case where the same dynamics can be reused to generate novel movements by exclusively modifying motor preparation. Different inputs during preparation change the activities at the start of motif production—and thus the amplitudes and phases of the basis functions—resulting in different outputs. We can call these "motif variants" to disambiguate them from different motifs that require all new dynamics. In Figure 3E, motif variants μ and μ' can share the same circuit dynamics during execution $(\tilde{\mathbf{J}}_{\mu})$ as long as their initial activities can be prepared with appropriate inputs. However, we stress that a distinctly shaped new movement cannot always be robustly and accurately constructed as a variant of some previously learned motif μ (or equivalently constructed from a fixed linear network; Figure S1) because the fixed dynamics constrain the basis functions' timescales and directions. The concept of motif variants is supported by results showing that, when primates prepare and reach to

different targets (Lara et al., 2018a; Churchland and Cunningham, 2014), the cortical activities across these movements can be well captured by shared linear dynamics. Furthermore, evidence suggests that structured variability at the end of the preparatory period correlates with output variability for similar movements (Churchland et al., 2006; Vyas et al., 2018), as is also the case in our model, specifically among motif variants (Figure 3F). This effect does not hold across different motifs, consistent with experiments that explored more varied movements (Al Borno et al., 2020; Sun et al., 2020).

In summary, we have shown that thalamocortical loops can modify the motor system dynamics to quickly drive the cortical activity toward a particular initial condition needed for an upcoming behavior. As we shall see in the following section, this enables near-seamless transitioning between motifs during sequence generation. Importantly, the same preparatory thalamocortical loop can be used to prepare any upcoming motif, with motif specificity coming from inputs to the system.

Switching between thalamocortical loops robustly and flexibly generates complex motor sequences

The components of the motor circuit discussed in the preceding sections can be combined to implement arbitrary motif sequences. The basal ganglia are the "selectors" of the system, dictating which sequence elements to perform and in which order. Via their inputs to thalamus, the basal ganglia alternate between disinhibiting different thalamocortical loops for the preparation of motif-specific initial conditions in cortex and the subsequent execution of those motifs (Figure 4A). The cortex-thalamus system is the "executor," implementing the necessary dynamics needed to prepare and execute each motif. During the preparatory period, the cortical activities converge toward a static pattern associated with the upcoming motif, while during execution, motif-specific activity is generated (Figure 4B).

In our model, basal ganglia act as a critical nonlinearity, turning on and off different thalamocortical loops and thus changing the cortical dynamics per motif. In contrast, a naive baseline linear network that could support similarly rich sequence generation would effectively require a different subnetwork for each motif (compare Figure 4C, left and right). The total number of units N_{units} required by our model scales as the sum of the number of motifs N_{motifs} and the number of basis functions K required for each motif (which is a lower bound on the size of the cortical network). A fully linear solution, on the other hand, would require $\sim KN_{motifs}$ units. These considerations emphasize how, during motif production, the control of many basis functions by a smaller number of thalamocortical loops—possibly as few as one, as we demonstrate here—is a key factor for the efficiency of our model. This makes it clear that the presence of strong nonlinear inhibition at a specific location in the thalamocortical architecture balances the goals of keeping N_{units} low while maintaining segregation of motif-specific circuit elements and providing a simple and plausible solution for adding motifs to the system.

Once a set of motifs is learned, any arbitrary motif sequence order can be generated by the network (Figure 4D). Just prior to the onset time for the next motif (here, we use 5 times the cortical rate model time constant), the active thalamocortical loop is switched to the preparatory loop and the preparatory input for the next motif is supplied to the circuit. During the transition period, the cortical activity approaches the appropriate initial condition for the upcoming motif. After this, removal of $\mathbf{x}_\mu$, inhibition of the preparatory loop, and disinhibition of the appropriate thalamic unit results in the execution of the next motif. Due to the smoothness constraint mentioned in the previous section, the motor output smoothly interpolates from the end of the prior motif to the start of the next one over the course of the transition period (even when those two output values differ substantially; STAR Methods, section 4.2.4; Logiaco and Escola, 2020). Notably, the need to invoke motor preparation dynamics for each motif transition—which arises in our network for purely computational reasons—argues against the "chunking" theory of motor sequences (Sakai et al., 2003; Abrahamse et al., 2013) and mirrors recent experimental results in primates, which show preparatory activity prior to each component of fast reach sequences (Zimnik and Churchland, 2021).

DISCUSSION

We have developed a model of the cortex-thalamus system that can produce sequences composed from an extensible library of motifs generated in arbitrary order. Mechanistically, this relies on switching linear dynamics where the switches—which are aligned to the transitions during sequence production—are triggered by basal ganglia disinhibition of specific thalamocortical loops with tuned weights.

The linearity of our model during motif execution makes it analytically tractable, while the nonlinear inactivation of thalamic units gives it flexibility. Indeed, any nonlinear dynamical system can be approximated by switching linear dynamics, with increasing precision (but decreasing tractability) when the switching frequency increases. In the context of motor sequencing, switching is useful to quickly stabilize and adjust the dynamics at each motif transition and to support motif-specific dynamical regimes. In contrast, although linear dynamical systems can have fixed points and interesting transient dynamics, they cannot modify their dynamics to suit different motifs. The thalamic modulation of cortical dynamics during different motifs therefore expands the expressivity of the linear networks previously used to model motor cortical dynamics (Hennequin et al., 2014; Churchland and Cunningham, 2014; Lara et al., 2018a).

Comparison to prior modeling work

We considered the case of a single thalamic unit active during motif execution that results in a "rank-one" perturbation to the circuit dynamics. This minimal change, coupled with the within-motif linearity of our model, allowed us to fully characterize the dynamics while accounting for the detailed structure of all the weights. This revealed that a single unit can robustly shape the activity of a recurrent network and expands a body of research using recurrent networks with dynamics defined by the sum of full-rank and low-rank matrices (Sussillo and Abbott, 2009; Rivkind and Barak, 2017; Mastrogiuseppe and Ostojic, 2018, 2019; Landau and Sompolinsky, 2018; Schuessler et al., 2020; Schüßler et al., 2020; Susman et al., 2021). Notably, the mathematical analysis of these networks used to focus on cases where the low-rank

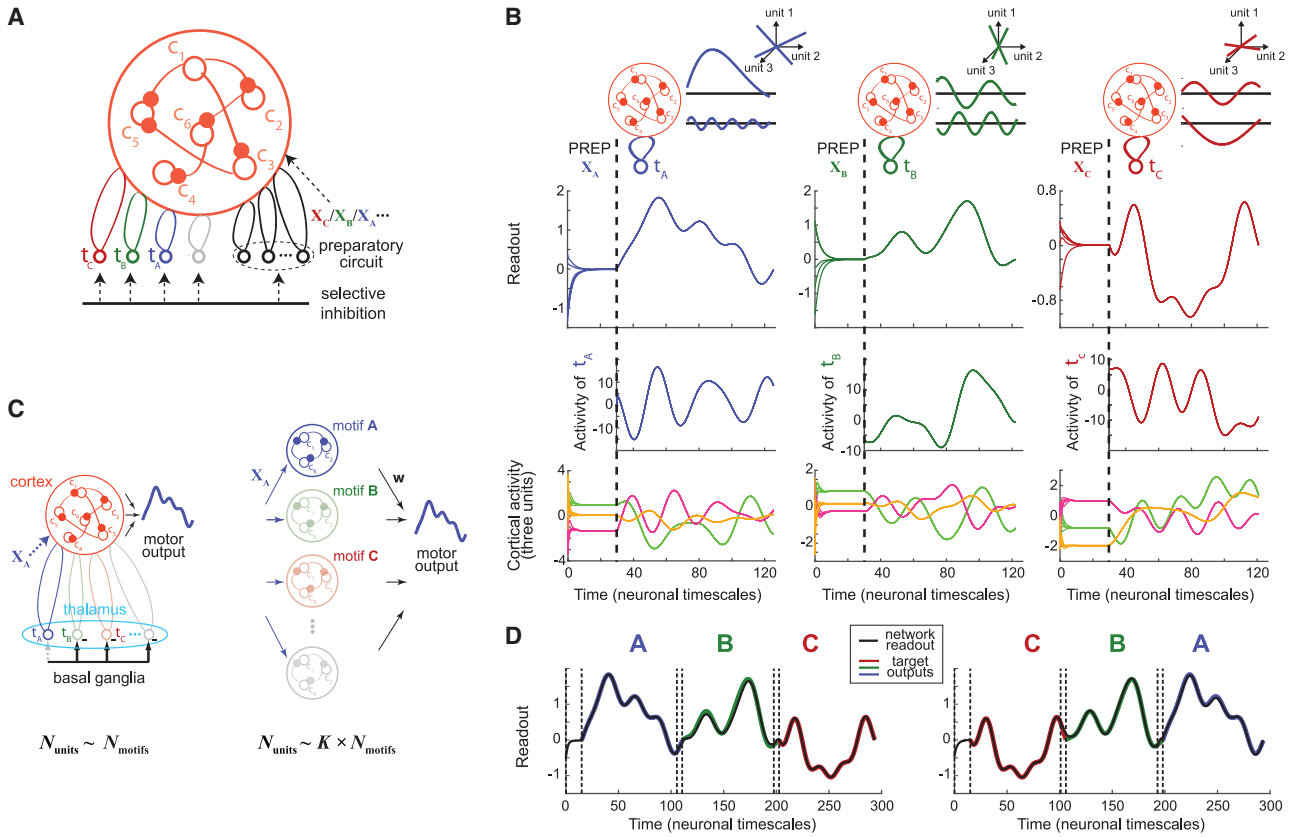


Figure 4. Flexible motor sequencing

(A) Schematic of the cortex-thalamus system. The input from the basal ganglia selects thalamic units needed for either motif execution or preparation. During preparation, the cortical population also receives an input $\mathbf{x}_\mu$ specific to the upcoming motif μ.

(B) Preparation and execution of three example motifs. When the preparatory thalamic units interact with cortex, the system output (upper) and cortical units (lower) converge to the values needed for the upcoming motif. (Here, for visualization purposes, the preparatory period was made much longer than needed.) Upon disinhibition of the motif-specific execution loops, the target motifs (upper) are composed from thalamic and cortical activities (middle and lower). These motifs are composed from a linear combination of motif-specific basis functions (top), with different (1) directions in neural space, (2) timescales, and (3) initial amplitudes and phases.

(C) Comparison of the thalamocortical network with a "baseline" linear network. These two architectures are shown during the production of motif "A," and active elements of the circuits are drawn in bright colors while inactive elements are grayed out. In the thalamocortical network (left), the total number of units in the network N_{units} is given as the number of motifs N_{motifs} added to the fixed sizes of the cortical network and of the preparatory network. As new motifs are added to the library, this implies a scaling of network size with N_{motifs}. In contrast, in a linear network (right), a subnetwork of size K (i.e., the number of basis functions used for motif construction) is required for each motif giving a total scaling of KN_{motifs}.

(D) Generation of sequences of arbitrary orders, using preparatory periods (between vertical dashed lines) before executing each motif.

perturbation dominates the resulting dynamics (Mastrogiuseppe and Ostojic, 2018; Landau and Sompolinsky, 2018) or on manipulating the timescales (the eigenvalues) of the system (Rivkind and Barak, 2017; Mastrogiuseppe and Ostojic, 2019; Schuessler et al., 2020). Instead, our analysis reveals that we can fully tune the low-rank matrix to match features of the detailed structure of both the full-rank matrix and the output weights, which enables control of both the timescales and the directions of the basis functions (their eigenvectors). By doing so, we can robustly shape high dimensional activity, which can then be used for the flexible production of complex temporal outputs.

Though flexible motor sequencing is routinely performed by animals, it remains a challenge in machine learning (Merel et al., 2019b; Merel et al., 2019a; Logiaco and Escola, 2020). Sequencing requires both that new motifs can be learned without destroying previously learned ones (He and Jaeger, 2018; Riemer et al., 2019) and that motif sequences can be generated in orders never experienced during training (Belkin et al., 2018; Geirhos et al., 2018; Merel et al., 2019a, 2019b)—two flexibility requirements that can lead to catastrophic failures with state-of-the-art methods (Logiaco and Escola, 2020). Guided by biological and formal principles, our thalamocortical network overcomes both of these challenges. First, different motifs can be learned with completely separate sets of parameters, preventing interference while still benefiting from the rich dynamics of the shared cortical network. Specifically, the capacity of the network for producing large numbers of motifs is only limited by the number of thalamic units. In addition, by

segregating the units that control cortical dynamics into their own non-recurrent brain region where activity cannot spread laterally—the thalamus—our model avoids the problem of interference between motifs. Finally, segregating the control units into thalamus facilitates two biological processes: inputting the switch commands into the circuit and learning the motif-specific synapses. Second, our preparatory network, which is motif independent, can implement any transition, including novel ones. We further show that transitions can be fast even under the constraint that the size of the preparatory thalamic population is a fraction of the size of cortex. Interestingly, recent results (Logiaco and Escola, 2020) demonstrate that the insights from the switching-linear thalamocortical model presented in this article are relevant to improving both robustness and accuracy in a continuously nonlinear network solving a hierarchical control task (Merel et al., 2019a).

Our implementation of motor preparation shares similarities with an alternative thalamic preparatory circuit that was recently developed (Kao et al., 2021). The key difference between the two approaches is that the thalamic preparatory network we propose is designed to generically handle any upcoming movement dynamics, whereas Kao et al. propose a preparatory network with synaptic weights that are matched to specific upcoming dynamics, yielding very efficient motor preparation. From a functional viewpoint, the two approaches have complementary strengths. The latter approach is clearly advantageous in a context in which a given dynamics is reused to create different motif variations (e.g., by changing the initial activities in cortex as in Figures 3E and 3F). However, in a context with multiple complex motifs requiring different dynamics (Figure 4), our preparatory network can be used to prepare all motifs while still implementing fast transitions. More generally, during motor preparation or motor production, our results emphasize that even a limited number of thalamic units can potently remap cortical dynamics, so that the thalamocortical circuit can extensibly learn and flexibly perform the motor sequencing task while using neural resources relatively sparingly.

Our model captures key experimental findings

Our model captures several important features of motor cortical activity, despite the fact that we did not impose biological constraints other than the type of dynamical regime and basic anatomy. Notably, we observe prominent oscillatory patterns of neural activity during motif execution (Russo et al., 2018; Churchland and Cunningham, 2014). We require an obligatory preparatory period before movement execution (Lara et al., 2018b; Shenoy et al., 2013). Variability in the cortical activity at the end of the preparatory period can correlate with variability in the motor output (Churchland et al., 2006; Vyas et al., 2018). Perhaps most interestingly, our model stipulates that preparatory periods are required even within ongoing motor sequences, specifically at the transition times between motifs. This mechanism is in contrast to the view that highly repeated motor sequences become chunks that are executed en bloc following their initiation (Sakai et al., 2003; Abrahamse et al., 2013). Recently, motor cortex recordings have shown that motor preparation occurs before starting each sequence element during rapidly executed sequences (Zimnik and Churchland, 2021), supporting our hypothesis that—at least in certain settings—chunking does not occur at the level of motor cortex.

Model limitations and extensions

Certain aspects of our model are considerably simplified compared to the full complexity of the biological circuit. Notably, we assumed (1) that, during specific motifs, all cortical units and all disinhibited thalamic units are in a linear regime and (2) that thalamus is instantaneous with respect to cortex. In STAR Methods, sections 4.2.5 and 4.2.6, we explain how it is possible to relax these two simplifications by using standard rectified units in place of linear units in both cortex and thalamus and by setting the thalamic time constant to be ten times faster than cortex rather than infinitely fast. We have verified that there exists a dynamical regime under which the model can still generate motif sequences with dynamics that closely follow our idealized theoretical framework (Figure S7).

To keep the exposition of our results focused on general mechanisms by which a motor sequential task can be solved by a recurrent network, we did not try to optimize the statistics of the recurrent connectivity matrix or of the motor motifs to fit a particular dataset. However, the thalamic control mechanisms proposed in this work could be combined with different putative types of cortical connectivity matrices, such as those that have been shown to capture specific statistics of cortical activity recorded while primates perform hand reaches (Hennequin et al., 2014; Kao et al., 2021).

Furthermore, sensory feedback is absent in our model. In many realistic settings, sensory feedback is likely to be key in organizing the sequence, as well as in providing error-correcting information (Penhune and Steele, 2012; Guo et al., 2020; Dacre et al., 2019). Our thalamocortical model could be extended by introducing a cerebellar module, which could fine-tune cortical dynamics in response to sensory feedback and/or contribute to motif selection and tuning. More generally, in our model, transitions between motifs are ''externally'' triggered by altering the pattern of basal ganglia input to the thalamus. In the brain, it is likely that the striatum is involved in this process by inhibiting neurons in the GPi/SNr (the output nuclei of the basal ganglia) at times determined by its monitoring of cortical activity. In particular, frontal brain regions, which project back to the basal ganglia (McFarland and Haber, 2000) and whose firing rates can reflect the abstract sequential structure of a task (Tanji and Shima, 1994; Clower and Alexander, 1998; Procyk et al., 2000; Russo et al., 2020), are good candidates for planning and controlling sequence generation in this way. Alternatively, other brain structures could participate in the mechanisms for thalamic disinhibition—the key functional feature of our model. For example, cerebellar outputs could modulate cortex-thalamus interactions (Nashef et al., 2021) through indirect disinhibitory signals, e.g., passing through the thalamic reticular nucleus (Arcelli et al., 1997).

Finally, our work does not address the mechanisms by which a biologically plausible learning rule could allow the brain to learn the synaptic weights of thalamocortical loops. The most straightforward interpretation of our model would suggest that plasticity occurs at the level of the synapses of the direct thalamocortical and corticothalamic projections—for which there is some

evidence (Pigeat et al., 2015; Oberlaender et al., 2012; Yu et al., 2012; Hsu et al., 2010; Castro-Alamancos and Calcagnotto, 1999). However, plasticity could also occur anywhere within an effective feedforward subnetwork between the thalamic and cortical populations involved in the dynamic production of motor commands. Our model assumes that the synaptic weights between units can be tuned to an accuracy of up to about 0.1% of the magnitude of the cortical weights (Figure S5), a similar constraint as for recurrent networks trained with an online algorithm (Sussillo and Barak, 2013). This fine-tuning requirement would be somewhat mitigated by mapping each model unit to a population of biological neurons and matching the model's synaptic weights to effective weights between populations (Mongillo et al., 2017). Further work will be needed to investigate whether, in the context of our model, the level of synaptic fine-tuning required for generating a complex motor sequence with a realistic number of units can be within a biologically plausible range.

Experimental predictions

Our model makes experimental predictions that can be tested in animals engaged in flexible motor sequencing tasks. First, we predict that changes in the activity patterns either in thalamus or in the basal ganglia should immediately precede, and be causally related to, changes in cortical dynamics. Recently developed data analysis tools (Linderman et al., 2017) can infer switch times between different dynamical regimes in neural population recordings. Thus, simultaneous recordings in motor cortex and either thalamus or GPi/SNr could be used to test this prediction. Similarly, we predict that switch times in thalamic activity patterns, as well as switch times in cortical dynamics, would reflect points of change in muscle activity delineating reusable behavioral primitives (Zimnik and Churchland, 2021). Further, motor preparation is predicted to involve the activation of a substantial thalamic population, considerably larger than for the execution of any single motif, which appears consistent with recent results (Nashef et al., 2021). Next, we predict that perturbative experiments in GPi/SNr and thalamus would have differential effects. A controlled alteration of activity patterns in the basal ganglia could modify the order of motifs while leaving individual motifs unchanged. Perturbing thalamocortical interactions during a motif, on the other hand, would affect cortical dynamics and would thus disrupt motor execution. This is in line with recent experimental results demonstrating the need for time-dependent thalamic input to motor cortex during movements (Sauerbrei et al., 2020) and would additionally suggest a critical role for the feedback from cortex to thalamus. In addition, our model posits that thalamocortical interactions can remap cortical dynamics to produce the basis functions needed for the execution of each motif. This is compatible with experimental reports showing substantial changes in motor cortical activity between different movements and/or contexts (Miri et al., 2017; Al Borno et al., 2020; Sun et al., 2020), as well as during prolonged brain machine interface training (Oby et al., 2019). Further analyses will be needed to investigate whether these large changes of cortical activity relate to changes in the effective cortical connectivity (Feulner and Clopath, 2021). Finally, our model also postulates that the thalamic neurons involved in shaping cortical dynamics during motif execution are segregated into motif-specific subpopulations. This prediction could be tested by correlating recordings in thalamus with behavior during a flexible sequencing task, ideally during the acquisition of new motifs. Interestingly, a study showed that, when training an animal on two distinct motor tasks, learning for each task is associated with its own synaptic subpopulation in a thalamic-recipient layer of motor cortex (Hayashi-Takagi et al., 2015; Kaneko, 2013). Our interpretation would be that these subpopulations are receiving inputs from motif-specific neurons in thalamus.

Conclusions

In conclusion, our corticothalamocortical model suggests a mechanism for flexible and robust sequence generation and leads to experimental predictions that can further our understanding of motor system function. In addition, our model reveals how complex cognitive processes may rely on neural systems operating on very different timescales. First, slow learning through the adjustment of synaptic weights can construct a library of cognitive building blocks, such as motifs. Then, assuming the network architecture is appropriately constrained—as in the case of motif-specific units in the thalamus—the flexible organization of these cognitive building blocks can be achieved online through a selection process like the one that we propose is implemented in the basal ganglia.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - 4.2.1 Model definition
  - 4.2.2 Eigenvalue control for motif sculpting
  - 4.2.3 Eigenvector control for motif robustness
  - 4.2.4 Motor preparation through thalamic loops
  - 4.2.5 Relation to more realistic neuronal dynamics
  - 4.2.6 Implementing the more realistic dynamics
  - 4.2.7 Numerics
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109090.

ACKNOWLEDGMENTS

We thank Christopher J. Cueva, Cora Ames, Andrew Zimnik, Zirong Gu, Francesca Mastrogiuseppe, and Omri Barak for useful discussions. This research was supported by NIH BRAIN award (U19 NS104649), NSF/NIH Collaborate Research in Computational Neuroscience award (R01 NS105349), NIH Director's Early Independence award (DP5 OD019897), the Gatsby Charitable Foundation award no. GAT3708, and NSF NeuroNex award (DBI-1707398), as well as the Leon Levy Foundation, and the Swartz Foundation.

AUTHOR CONTRIBUTIONS

L.L., L.F.A., and S.E. contributed to all aspects of the research, including project design, model development, mathematical analysis, programming, and writing of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 17, 2020
Revised: March 4, 2021
Accepted: April 16, 2021
Published: June 1, 2021

REFERENCES

Abrahamse, E.L., Ruitenberg, M.F.L., de Kleine, E., and Verwey, W.B. (2013). Control of automated behavior: insights from the discrete sequence production task. Front. Hum. Neurosci. *7*, 82.

Ahmadian, Y., Fumarola, F., and Miller, K.D. (2015). Properties of networks with partially structured and partially random connectivity. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *91*, 012820.

Al Borno, M., Vyas, S., Shenoy, K.V., and Delp, S.L. (2020). High-fidelity musculoskeletal modeling reveals that motor planning variability contributes to the speed-accuracy tradeoff. eLife *9*, e57021.

Alexander, G.E., and Crutcher, M.D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci. *13*, 266–271.

Aoki, S., Smith, J.B., Li, H., Yan, X., Igarashi, M., Coulon, P., Wickens, J.R., Ruigrok, T.J., and Jin, X. (2019). An open cortico-basal ganglia loop allows limbic control over motor output via the nigrothalamic pathway. eLife *8*, e49995.

Arcelli, P., Frassoni, C., Regondi, M.C., De Biasi, S., and Spreafico, R. (1997). GABAergic neurons in mammalian thalamus: a marker of thalamic complexity? Brain Res. Bull. *42*, 27–37.

Bai, Z.D. (1997). Circular law. Ann. Probab. *25*, 494–529.

Belkin, M., Hsu, D.J., and Mitra, P. (2018). Overfitting or perfect fitting? risk bounds for classification and regression rules that interpolate. In Advances in Neural Information Processing Systems, *Volume 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds. (Curran Associates), pp. 2300–2311.

Beloozerova, I.N., and Marlinski, V. (2020). Contribution of the ventrolateral thalamus to the locomotion-related activity of motor cortex. J. Neurophysiol. *124*, 1480–1504.

Boeddeker, C., Hanebrink, P., Drude, L., Heymann, J., and Haeb-Umbach, R. (2017). On the computation of complex-valued gradients with application to statistically optimum beamforming. arXiv, arXiv:1701.00392v2. https://arxiv.org/abs/1701.00392.

Bondanelli, G., and Ostojic, S. (2020). Coding with transient trajectories in recurrent neural networks. PLoS Comput. Biol. *16*, e1007655.

Castro-Alamancos, M.A., and Calcagnotto, M.E. (1999). Presynaptic long-term potentiation in corticothalamic synapses. J. Neurosci. *19*, 9090–9097.

Chalker, J.T., and Mehlig, B. (1998). Eigenvector statistics in non-hermitian random matrix ensembles. Phys. Rev. Lett. *81*, 3367–3370.

Churchland, M.M., and Cunningham, J.P. (2014). A dynamical basis set for generating reaches. Cold Spring Harb. Symp. Quant. Biol. *79*, 67–80.

Churchland, M.M., Afshar, A., and Shenoy, K.V. (2006). A central source of movement variability. Neuron *52*, 1085–1096.

Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat. Neurosci. *13*, 369–378.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56.

Clower, W.T., and Alexander, G.E. (1998). Movement sequence-related activity reflecting numerical order of components in supplementary and presupplementary motor areas. J. Neurophysiol. *80*, 1562–1566.

Dacre, J., Colligan, M., Ammer, J., Schiemann, J., Clarke, T., Chamosa-Pino, V., Claudi, F., Harston, J.A., Eleftheriou, C., Pakan, J.M.P., et al. (2019). Cerebellar-recipient motor thalamus drives behavioral context-specific movement initiation. bioRxiv. https://doi.org/10.1101/802124.

Deniau, J.M., and Chevalier, G. (1985). Disinhibition as a basic process in the expression of striatal functions. II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. Brain Res. *334*, 227–233.

Devergnas, A., Chen, E., Ma, Y., Hamada, I., Pittard, D., Kammermeier, S., Mullin, A.P., Faundez, V., Lindsley, C.W., Jones, C., et al. (2016). Anatomical localization of Cav3.1 calcium channels and electrophysiological effects of T-type calcium channel blockade in the motor thalamus of MPTP-treated monkeys. J. Neurophysiol. *115*, 470–485.

Edgerton, J.R., and Jaeger, D. (2014). Optogenetic activation of nigral inhibitory inputs to motor thalamus in the mouse reveals classic inhibition with little potential for rebound activation. Front. Cell. Neurosci. *8*, 36.

Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., and Cunningham, J.P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nat. Commun. *7*, 13239.

Feulner, B., and Clopath, C. (2021). Neural manifold under plasticity in a goal driven learning behaviour. PLoS Comput. Biol. *17*, e1008621.

Ganguli, S., Huh, D., and Sompolinsky, H. (2008). Memory traces in dynamical systems. Proc. Natl. Acad. Sci. USA *105*, 18970–18975.

Geddes, C.E., Li, H., and Jin, X. (2018). Optogenetic editing reveals the hierarchical organization of learned action sequences. Cell *174*, 32–43.e15.

Geirhos, R., Temme, C.R.M., Rauber, J., Schütt, H.H., Bethge, M., and Wichmann, F.A. (2018). Generalisation in humans and deep neural networks. In Advances in Neural Information Processing Systems, *Volume 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds. (Curran Associates), pp. 7538–7550.

Gerstner, W., Kistler, W.M., Naud, R., and Paninski, L. (2009). Neuronal Dynamics (Cambridge University).

Girko, V.L. (1985). Circular law. Theory Probab. Appl. *29*, 694–706.

Guo, Z.V., Inagaki, H.K., Daie, K., Druckmann, S., Gerfen, C.R., and Svoboda, K. (2017). Maintenance of persistent activity in a frontal thalamocortical loop. Nature *545*, 181–186.

Guo, J.-Z., Sauerbrei, B., Cohen, J.D., Mischiati, M., Graves, A., Pisanello, F., Branson, K., and Hantman, A.W. (2020). Dynamics of the cortico-cerebellar loop fine-tune dexterous movement. bioRxiv. https://doi.org/10.1101/637447.

Halley, A.C., and Krubitzer, L. (2019). Not all cortical expansions are the same: the coevolution of the neocortex and the dorsal thalamus in mammals. Curr. Opin. Neurobiol. *56*, 78–86.

Harris, J.A., Mihalas, S., Hirokawa, K.E., Whitesell, J.D., Choi, H., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., et al. (2019). Hierarchical organization of cortical and thalamic connectivity. Nature *575*, 195–202.

Harrison, T.C., Ayling, O.G., and Murphy, T.H. (2012). Distinct cortical circuit mechanisms for complex forelimb movement and motor map topography. Neuron *74*, 397–409.

Hayashi-Takagi, A., Yagishita, S., Nakamura, M., Shirai, F., Wu, Y.I., Loshbaugh, A.L., Kuhlman, B., Hahn, K.M., and Kasai, H. (2015). Labelling and optical erasure of synaptic memory traces in the motor cortex. Nature *525*, 333–338.

He, X., and Jaeger, H. (2018). Overcoming catastrophic interference using conceptor-aided backpropagation. https://openreview.net/forum?id=B1al7jg0b.

Hennequin, G., Vogels, T.P., and Gerstner, W. (2012). Non-normal amplification in random balanced neuronal networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *86*, 011909.

Hennequin, G., Vogels, T.P., and Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. Neuron *82*, 1394–1406.

Hirst, H.P., and Macey, W.T. (1997). Bounding the roots of polynomials. Coll. Math. J. *28*, 292–295.

Hsu, C.-L., Yang, H.-W., Yen, C.-T., and Min, M.-Y. (2010). Comparison of synaptic transmission and plasticity between sensory and cortical synapses on relay neurons in the ventrobasal nucleus of the rat thalamus. J. Physiol. *588*, 4347–4363.

Jin, X., and Costa, R.M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature *466*, 457–462.

Jin, X., Tecuapetla, F., and Costa, R.M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. Nat. Neurosci. *17*, 423–430.

Kaneko, T. (2013). Local connections of excitatory neurons in motor-associated cortical areas of the rat. Front. Neural Circuits *7*, 75.

Kao, T.-C., Sadabadi, M.S., and Hennequin, G. (2021). Optimal anticipatory control as a theory of motor preparation: a thalamo-cortical circuit model. Neuron, Published online March 30, 2021. https://doi.org/10.1016/j.neuron.2021.03.009.

Kim, J., Kim, Y., Nakajima, R., Shin, A., Jeong, M., Park, A.H., Jeong, Y., Jo, S., Yang, S., Park, H., et al. (2017). Inhibitory basal ganglia inputs induce excitatory motor signals in the thalamus. Neuron *95*, 1181–1196.e8.

Kornysheva, K., and Diedrichsen, J. (2014). Human premotor areas parse sequences into their spatial and temporal features. eLife *3*, e03043.

Krakauer, J.W., Hadjiosif, A.M., Xu, J., Wong, A.L., and Haith, A.M. (2019). Motor learning. In Comprehensive Physiology, D.M. Pollock, ed. (Wiley).

Landau, I.D., and Sompolinsky, H. (2018). Coherent chaos in a recurrent neural network with structured connectivity. PLoS Comput. Biol. *14*, e1006309.

Lara, A.H., Cunningham, J.P., and Churchland, M.M. (2018a). Different population dynamics in the supplementary motor area and motor cortex during reaching. Nat. Commun. *9*, 2754.

Lara, A.H., Elsayed, G.F., Zimnik, A.J., Cunningham, J.P., and Churchland, M.M. (2018b). Conservation of preparatory neural events in monkey motor cortex regardless of how movement is initiated. eLife *7*, e31826.

Lim, S., and Goldman, M.S. (2013). Balanced cortical microcircuitry for maintaining information in working memory. Nat. Neurosci. *16*, 1306–1314.

Linderman, S., Johnson, M., Miller, A., Adams, R., Blei, D., and Paninski, L. (2017). Bayesian learning and inference in recurrent switching linear dynamical systems. PMLR *54*, 914–922.

Logiaco, L., and Escola, G.S. (2020). Thalamocortical motor circuit insights for more robust hierarchical control of complex sequences. arXiv, arXiv:2006.13332v1. https://arxiv.org/abs/2006.13332.

Mannella, F., and Baldassarre, G. (2015). Selection of cortical dynamics for motor behaviour by the basal ganglia. Biol. Cybern. *109*, 575–595.

Mastrogiuseppe, F., and Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. Neuron *99*, 609–623.e29.

Mastrogiuseppe, F., and Ostojic, S. (2019). A geometrical analysis of global stability in trained feedback networks. Neural Comput. *31*, 1139–1182.

McFarland, N.R., and Haber, S.N. (2000). Convergent inputs from thalamic motor nuclei and frontal cortical areas to the dorsal striatum in the primate. J. Neurosci. *20*, 3798–3813.

Mehta, D., Chen, T., Tang, T., and Hauenstein, J.D. (2018). The loss surface of deep linear networks viewed through the algebraic geometry lens. arXiv, arXiv:1810.07716v1. https://arxiv.org/abs/1810.07716v1.

Mensi, S., Naud, R., Pozzorini, C., Avermann, M., Petersen, C.C.H., and Gerstner, W. (2012). Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. J. Neurophysiol. *107*, 1756–1775.

Merel, J., Ahuja, A., Pham, V., Tunyasuvunakool, S., Liu, S., Tirumala, D., Heess, N., and Wayne, G. (2019a). Hierarchical visuomotor control of humanoids. arXiv, arXiv:1811.09656v2. https://arxiv.org/abs/1811.09656.

Merel, J., Botvinick, M., and Wayne, G. (2019b). Hierarchical motor control in mammals and machines. Nat. Commun. *10*, 5489.

Mezzadri, F. (2007). How to generate random matrices from the classical compact groups. Not. Am. Math. Soc. *54*, 592–604.

Miri, A., Warriner, C.L., Seely, J.S., Elsayed, G.F., Cunningham, J.P., Churchland, M.M., and Jessell, T.M. (2017). Behaviorally selective engagement of short-latency effector pathways by motor cortex. Neuron *95*, 683–696.e11.

Mongillo, G., Rumpel, S., and Loewenstein, Y. (2017). Intrinsic volatility of synaptic connections - a challenge to the synaptic trace theory of memory. Curr. Opin. Neurobiol. *46*, 7–13.

Murphy, B.K., and Miller, K.D. (2009). Balanced amplification: a new mechanism of selective amplification of neural activity patterns. Neuron *61*, 635–648.

Murray, J.M., and Escola, G.S. (2017). Learning multiple variable-speed sequences in striatum via cortical tutoring. eLife *6*, e26084.

Mushiake, H., and Strick, P.L. (1995). Pallidal neuron activity during sequential arm movements. J. Neurophysiol. *74*, 2754–2758.

Nashef, A., Mitelman, R., Harel, R., Joshua, M., and Prut, Y. (2021). Area-specific thalamocortical synchronization underlies the transition from motor planning to execution. Proc. Natl. Acad. Sci. USA *118*, e2012658118.

Oberlaender, M., Ramirez, A., and Bruno, R.M. (2012). Sensory experience restructures thalamocortical axons during adulthood. Neuron *74*, 648–655.

Oby, E.R., Golub, M.D., Hennig, J.A., Degenhart, A.D., Tyler-Kabara, E.C., Yu, B.M., Chase, S.M., and Batista, A.P. (2019). New neural activity patterns emerge with long-term learning. Proc. Natl. Acad. Sci. USA *116*, 15210–15215.

Ostojic, S., and Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. PLoS Comput. Biol. *7*, e1001056.

Penhune, V.B., and Steele, C.J. (2012). Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. Behav. Brain Res. *226*, 579–591.

Pigeat, R., Chausson, P., Dreyfus, F.M., Leresche, N., and Lambert, R.C. (2015). Sleep slow wave-related homo and heterosynaptic LTD of intrathalamic GABAAergic synapses: involvement of T-type Ca2+ channels and metabotropic glutamate receptors. J. Neurosci. *35*, 64–73.

Procyk, E., Tanaka, Y.L., and Joseph, J.P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. Nat. Neurosci. *3*, 502–508.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K.D. (2010). The asynchronous state in cortical circuits. Science *327*, 587–590.

Riemer, M., Cases, I., Ajemian, R., Liu, M., Rish, I., Tu, Y., and Tesauro, G. (2019). Learning to learn without forgetting by maximizing transfer and minimizing interference. arXiv, arXiv:1810.11910v3. https://arxiv.org/abs/1810.11910.

Rikhye, R.V., Gilra, A., and Halassa, M.M. (2018). Thalamic regulation of switching between cortical representations enables cognitive flexibility. Nat. Neurosci. *21*, 1753–1763.

Rivkind, A., and Barak, O. (2017). Local dynamics in trained recurrent neural networks. Phys. Rev. Lett. *118*, 258101.

Russo, A.A., Bittner, S.R., Perkins, S.M., Seely, J.S., London, B.M., Lara, A.H., Miri, A., Marshall, N.J., Kohn, A., Jessell, T.M., et al. (2018). Motor cortex embeds muscle-like commands in an untangled population response. Neuron *97*, 953–966.e8.

Russo, A.A., Khajeh, R., Bittner, S.R., Perkins, S.M., Cunningham, J.P., Abbott, L.F., and Churchland, M.M. (2020). Neural trajectories in the supplementary motor area and motor cortex exhibit distinct geometries, compatible with different classes of computation. Neuron *107*, 745–758.e6.

Sakai, K., Kitaguchi, K., and Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. Exp. Brain Res. *152*, 229–242.

Sauerbrei, B.A., Guo, J.-Z., Cohen, J.D., Mischiati, M., Guo, W., Kabra, M., Verma, N., Mensh, B., Branson, K., and Hantman, A.W. (2020). Cortical pattern generation during dexterous movement is input-driven. Nature *577*, 386–391.

Saxena, S., Russo, A.A., Cunningham, J.P., and Churchland, M.M. (2021). Motor cortex activity across movement speeds is predicted by network-level strategies for generating muscle activity. bioRxiv. https://doi.org/10.1101/2021.02.01.429168.

Schmitt, L.I., Wimmer, R.D., Nakajima, M., Happ, M., Mofakham, S., and Halassa, M.M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. Nature *545*, 219–223.

Schuessler, F., Dubreuil, A., Mastrogiuseppe, F., Ostojic, S., and Barak, O. (2020). Dynamics of random recurrent networks with correlated low-rank structure. Phys. Rev. Research *2*, 013111.

Schüßler, F., Mastrogiuseppe, F., Dubreuil, A.M., Ostojic, S., and Barak, O. (2020). The interplay between randomness and structure during learning in rnns. In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds. (Curran Associates).

Schwab, B.C., Kase, D., Zimnik, A., Rosenbaum, R., Codianni, M.G., Rubin, J.E., and Turner, R.S. (2020). Neural activity during a simple reaching task in macaques is counter to gating and rebound in basal ganglia-thalamic communication. PLoS Biol. *18*, e3000829.

Seung, H.S., Lee, D.D., Reis, B.Y., and Tank, D.W. (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. Neuron *26*, 259–271.

Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical control of arm movements: a dynamical systems perspective. Annu. Rev. Neurosci. *36*, 337–359.

Sherman, S.M. (2001). Tonic and burst firing: dual modes of thalamocortical relay. Trends Neurosci. *24*, 122–126.

Sherman, S.M. (2016). Thalamus plays a central role in ongoing cortical functioning. Nat. Neurosci. *19*, 533–541.

Sun, X., O'Shea, D.J., Golub, M.D., Trautmann, E.M., Vyas, S., Ryu, S.I., and Shenoy, K.V. (2020). Skill-specific changes in cortical preparatory activity during motor learning. bioRxiv. https://doi.org/10.1101/2020.01.30.919894.

Susman, L., Mastrogiuseppe, F., Brenner, N., and Barak, O. (2021). Quality of internal representation shapes learning performance in feedback neural networks. Phys. Rev. Res. *3*, 013176.

Sussillo, D., and Abbott, L.F. (2009). Generating coherent patterns of activity from chaotic neural networks. Neuron *63*, 544–557.

Sussillo, D., and Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput. *25*, 626–649.

Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K.V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. Nat. Neurosci. *18*, 1025–1033.

Tanji, J., and Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. Nature *371*, 413–416.

Tao, T. (2013). Outliers in the spectrum of iid matrices with bounded rank perturbations. Probab. Theory Relat. Fields *155*, 231–263.

Trefethen, L.N. (2005). Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators (Princeton University).

van Donkelaar, P., Stein, J.F., Passingham, R.E., and Miall, R.C. (1999). Neuronal activity in the primate motor thalamus during visually triggered and internally generated limb movements. J. Neurophysiol. *82*, 934–945.

van Donkelaar, P., Stein, J.F., Passingham, R.E., and Miall, R.C. (2000). Temporary inactivation in the primate motor thalamus during visually triggered and internally generated limb movements. J. Neurophysiol. *83*, 2780–2790.

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science *274*, 1724–1726.

van Vreeswijk, C., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. Neural Comput. *10*, 1321–1371.

Vyas, S., Even-Chen, N., Stavisky, S.D., Ryu, S.I., Nuyujukian, P., and Shenoy, K.V. (2018). Neural population dynamics underlying motor learning transfer. Neuron *97*, 1177–1186.e3.

Yu, X., Chung, S., Chen, D.Y., Wang, S., Dodd, S.J., Walters, J.R., Isaac, J.T., and Koretsky, A.P. (2012). Thalamocortical inputs show post-critical-period plasticity. Neuron *74*, 731–742.

Zimnik, A.J., and Churchland, M.M. (2021). Independent generation of sequence elements by motor cortex. Nat. Neurosci. *24*, 412–424.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
thalContCrtx	This paper	https://github.com/LaurelineLogiaco/thalContCrtx_seq

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Laureline Logiaco (laureline.logiaco@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The code for simulating the model described in this study is available on github (https://github.com/LaurelineLogiaco/thalContCrtx_seq).

METHOD DETAILS

Here, we formally define our model and derive several properties that we have illustrated in the main text figures. We then explain how additional biological constraints – such as finite timescales and firing rate positivity – can be added to the model while staying within the validity domain of our simplified mathematical framework. In the main text, we have aimed to use intuitive words to describe the mechanisms by which our model produces output; however here we will use the corresponding technical terms to make the text as clear as possible for the computationally-oriented reader. We list the equivalencies between the names used in the main text and the mathematical terminology in the table below.

Correspondence between intuitive and technical terms	
Main text phrase	Technical term
Basis function of the dynamics	eigenmode
Timescale parameter for each basis function	eigenvalue λ
Direction/axis of basis function	eigenvector

4.2.1 Model definition

Here, we give a short derivation of the simplified rate equations we use in the main text (Equation 1). A longer derivation linking these rate equations to more biologically plausible neuronal dynamics is provided in STAR Methods section 4.2.5.

We consider a rate description of population neural dynamics (Gerstner et al., 2009), where the cortical activities $\mathbf{c}$ and thalamic activities $\mathbf{t}$ interact according to:

$$\begin{cases} \dot{\mathbf{c}} = -\mathbf{c} + \mathbf{J}_{\mathbf{cc}}\,\mathbf{c} + \mathbf{J}_{\mathrm{ct}}\,[\mathbf{t}]^{+} + \mathbf{x} \quad \text{(Equation 3)} \\ \tau_t\,\dot{\mathbf{t}} = -\mathbf{t} + \mathbf{J}_{\mathrm{tc}}\,\mathbf{c} + \mathbf{b}, \quad \text{(Equation 4)} \end{cases}$$

where $\mathbf{J}_{\mathbf{cc}}$ is the recurrent cortical connectivity matrix, $\mathbf{J}_{\mathrm{ct}}$ is the thalamocortical weight matrix, $\mathbf{J}_{\mathrm{tc}}$ is the corticothalamic weight matrix, and $\mathbf{x}$ and $\mathbf{b}$ are external inputs. Finally, $[\,]^{+}$ is the rectified nonlinearity. Note that these equations assume that cortical dynamics stays in a quasi-linear regime so that the cortical rates do not need to be passed through rectification. Also, note that the dynamics are expressed in units of the cortical timescale (that we normalized to one in our figures).

In addition, we analyze the model in the limit in which the thalamic dynamics are fast compared to cortical dynamics – which would be expected given that positive recurrent feedback, which is present in cortex but not in thalamus (Arcelli et al., 1997; Sherman,

2016), often results in slower timescales of neuronal population dynamics (Seung et al., 2000; Lim and Goldman, 2013). In this limit, $\mathbf{t} = \mathbf{J}_{tc}\,\mathbf{c} + \mathbf{b}$. In addition, we consider that during a particular stage 's' of the task (e.g., during motif execution or preparation), the thalamic units are separated into two groups indexed by a diagonal 'selection' matrix $\mathbf{S}_s$: those who are fully shut off by their basal ganglia inputs, which correspond to zeros in the diagonal of $\mathbf{S}_s$; and those who are left free to interact linearly with cortex, which correspond to ones in the diagonal of $\mathbf{S}_s$. Then, during the task stage s, we can find a mapping between $[\mathbf{t}]^+$ and the thalamic activity vector $\widehat{\mathbf{t}} = \mathbf{S}_s\mathbf{J}_{tc}\,\mathbf{c}$ (see STAR Methods 4.2.5 for details) such that we can write:

$$\dot{\mathbf{c}} = -\mathbf{c} + \mathbf{J}_{cc}\,\mathbf{c} + \mathbf{J}_{ct}\,\mathbf{S}_s\mathbf{J}_{tc}\,\mathbf{c} + \mathbf{x}_s \qquad \text{(Equation 5)}$$

Depending on the particular stage considered – motif production or motor preparation – the characteristics of the cortical input $\mathbf{x}_s$ and of the selection matrix $\mathbf{S}_s$ will be different, as we explain below.

Motif execution dynamics

During the execution of a given motif dynamics μ, we study the minimal case when a single motif-specific thalamic unit is left free to interact with cortex, such that the matrix $\mathbf{S}_s = \mathbf{S}_\mu$ has a single non-zero entry on row μ. Therefore, $\mathbf{J}_{ct}\,\widehat{\mathbf{t}} = \mathbf{u}_\mu\mathbf{v}_\mu^\top\mathbf{c}$ where $\mathbf{u}_\mu$ is the μ^{th} column of $\mathbf{J}_{ct}$ and $\mathbf{v}_\mu^\top$ is the μ^{th} line of $\mathbf{J}_{tc}$. In addition, there is no input needed during motif, so $\mathbf{x}_s = 0$, and we get:

$$\dot{\mathbf{c}} = -\mathbf{c} + \mathbf{J}_{cc}\mathbf{c} + \mathbf{u}_\mu\mathbf{v}_\mu^\top\mathbf{c} := -\mathbf{c} + \tilde{\mathbf{J}}_\mu\mathbf{c}. \qquad \text{(Equation 6)}$$

Hence, the thalamocortical loop involved in the execution of motif μ effectively implements a rank-one perturbation $\mathbf{u}_\mu\mathbf{v}_\mu^\top$ of the connectivity $\mathbf{J}_{cc}$ to create an effective motif-specific connectivity $\tilde{\mathbf{J}}_\mu$.

Motor preparation dynamics

We assume that for the preparation of any motif, a specific population of N_{prep} thalamic neurons interact with cortex such that $\mathbf{S}_s = \mathbf{S}_{prep}$ where $\mathbf{S}_{prep}$ has N_{prep} non-zero entries that are separate from the entries corresponding to motif execution units. In addition, for the preparation of a given motif variation which starts with the ideal cortical activity pattern $\mathbf{c}_\mu^{init}$, the input $\mathbf{x}_\mu$ is added to the circuit, so that the dynamics is:

$$\dot{\mathbf{c}} = -\mathbf{c} + \mathbf{J}_{ct}\mathbf{S}_{prep}\mathbf{J}_{tc} + \mathbf{x}_\mu := -\mathbf{c} + \tilde{\mathbf{J}}_{prep}\mathbf{c} + \mathbf{x}_\mu. \qquad \text{(Equation 7)}$$

If the dynamics are asymptotically stable (which is one of the aims of the design of the connectivity $\mathbf{J}_{ct}\mathbf{S}_{prep}\mathbf{J}_{tc}$, as we will explain in STAR Methods section 4.2.4), then by choosing $\mathbf{x}_\mu = -(\tilde{\mathbf{J}}_{prep} - \mathbf{I})\mathbf{c}_\mu^{init}$, the activities indeed converge toward $\mathbf{c}_\mu^{init}$.

Now that we have formally defined the dynamics of our model, we will describe how the thalamocortical perturbation weights can shape the cortical dynamics to support flexible motor sequencing.

4.2.2 Eigenvalue control for motif sculpting

In this section, we focus on the dynamics of the circuit during the production of the motif μ, associated with the effective connectivity matrix $\tilde{\mathbf{J}}_\mu$. For the sake of equation compactness, in this section we omit the index μ for the associated thalamocortical loop vectors $\mathbf{u}$ and $\mathbf{v}$ as well as for the eigenvalues $[\tilde{\lambda}_i]_{1\le i\le N}$ and the left and right eigenvector matrices $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{R}}$ of $\tilde{\mathbf{J}}_\mu$.

The eigenvalues $\tilde{\lambda}_1 \ldots \tilde{\lambda}_i \ldots \tilde{\lambda}_N$ of the effective motif-specific connectivity $\tilde{\mathbf{J}}_\mu$ obey the characteristic equation:

$$0 = \det\left(\mathbf{J}_{cc} - \tilde{\lambda}_i\,\mathbf{I} + \mathbf{u}\mathbf{v}^\top\right) = \left(1 + \mathbf{v}^\top\left(\mathbf{J}_{cc} - \tilde{\lambda}_i\,\mathbf{I}\right)^{-1}\mathbf{u}\right)\det\left(\mathbf{J}_{cc} - \tilde{\lambda}_i\,\mathbf{I}\right),$$

where we assumed that $\mathbf{J}_{cc}$ is an invertible matrix whose eigenvalues are all distinct from all $[\tilde{\lambda}_i]_{1\le i\le N}$ so that $\mathbf{J}_{cc} - \tilde{\lambda}_i\,\mathbf{I}$ is invertible, and we used the matrix determinant lemma to write the second equality. In addition, these assumptions we make on $\mathbf{J}_{cc}$ imply that $\det(\mathbf{J}_{cc} - \tilde{\lambda}_i\,\mathbf{I}) \neq 0$, so the above equation implies:

$$1 = \mathbf{v}^\top\left(\tilde{\lambda}_i\,\mathbf{I} - \mathbf{J}_{cc}\right)^{-1}\mathbf{u}\,. \qquad \text{(Equation 8)}$$

We then further assume that $\mathbf{J}_{cc}$ is diagonalizable and use the eigendecomposition procedure to write $\mathbf{J}_{cc} = \mathbf{R}\mathrm{diag}(\lambda)\mathbf{L}$, where we have concatenated the eigenvalues $\lambda_1, \ldots, \lambda_N$ in the vector λ, and $\mathbf{L}$ and $\mathbf{R}$ are matrices regrouping the left and right eigenvectors of the matrix $\mathbf{J}_{cc}$. This allows us to expand $(\tilde{\lambda}_i\,\mathbf{I} - \mathbf{J}_{cc})^{-1} = \mathbf{R}\,\mathrm{diag}(\tilde{\lambda}_i - \lambda)^{-1}\,\mathbf{L}$, where $\mathrm{diag}(\tilde{\lambda}_i - \lambda)^{-1}$ is a diagonal matrix whose element on the j^{th} row and column is $1/(\tilde{\lambda}_i - \lambda_j)$. We can then rewrite Equation 8 to get:

$$1 = \mathbf{v}^\top\mathbf{R}\,\mathrm{diag}\left(\tilde{\lambda}_i - \lambda\right)^{-1}\mathbf{L}\mathbf{u}\,. \qquad \text{(Equation 9)}$$

Note that Equation 9 is valid for all eigenvalues $\tilde{\lambda}_i$ of $\tilde{\mathbf{J}}_\mu$. As a consequence, in order for the eigenspectrum of $\tilde{\mathbf{J}}_\mu$ to include an ensemble of $K \le N$ desired eigenvalues λ_i^{des}, then – with the $K \times N$ matrix $\mathbf{P}$ defined to have elements $P_{ij} = 1/\left(\lambda_i^{des} - \lambda_j\right)$ – the following system of equations holds:

$$\mathbb{1} = \mathbf{P}\,\mathrm{diag}(\mathbf{R}^\top\mathbf{v})\,\mathbf{L}\mathbf{u} \qquad \text{(Equation 10)}$$

$$= \mathbf{P}\,\mathrm{diag}(\mathbf{L}\mathbf{u})\,\mathbf{R}^\top\mathbf{v}, \qquad \text{(Equation 11)}$$

where $\mathbb{1}$ is a vector of ones. Note that these two expressions are equivalent and that they restate Equation 9 as a linear equation for either **u** or **v** instead of an implicit equation for both. These equations then allow one to get an expression for either the vector **u** (from Equation 10) or the vector **v** (from Equation 11) that imposes the eigenvalues λ_i^{des} in the eigenspectrum of $\tilde{\mathbf{J}}_\mu$. In the following, without loss of generality, we will proceed with the latter.

It is interesting to examine the extreme case when all eigenvalues are desired to change (i.e., $K = N$). It is then easy to see that under the mild assumptions that *(i)* **P** is full-rank – notably requiring that all the eigenvalues λ_i and λ_i^{des} are all distinct from one another; and *(ii)* **u** is not orthogonal to any of the rows of **L**, then there actually exist a solution for the weights **v** such that $\forall 1 \leq i \leq N, \tilde{\lambda}_i = \lambda_i^{\text{des}}$. This solution is $\mathbf{v} = \mathbf{L}^\top \text{diag}(\mathbf{Lu})^{-1}\mathbf{P}^{-1}\mathbb{1}$. However, although this solution is well-defined analytically, it can suffer from numerical stability issues (Figures S2A–S2C). Indeed, if several desired eigenvalues λ_i^{des} are way outside of the ensemble of initial eigenvalues λ_j, the statistics of the difference $\lambda_i^{\text{des}} - \lambda_j$ will be on a larger scale compared to the differences among the initial eigenvalues, and therefore the matrix **P** will have correlated rows and will be hard to invert. Indeed, more formally, in this case, we can write $\lambda_i^{\text{des}} - \lambda_j = \lambda_i^{\text{des}} - (\bar{\lambda} + \varepsilon_j)$, with $\varepsilon_j \ll \lambda_i^{\text{des}} - \bar{\lambda}$ and $\bar{\lambda} = \sum_j \lambda_j / N$. Hence, to zeroth order in ε_j, $P_{ij} \approx 1/(\tilde{\lambda}_i - \bar{\lambda})$ which is independent of the column index j, indicating that different rows will be almost constant and therefore correlated to one another. This explains why, numerically, the control of all eigenvalues succeeds if the target eigenvalues have the same distribution as the initial eigenvalues (Figure S2A) but not if they are drawn from a very different distribution (Figure S2B). These correlations in the matrix **P** also relate to the necessity for the perturbation weights $\mathbf{uv}^\top$ to be large when many desired eigenvalues are different from the ensemble of initial eigenvalues (Figure S2B; Tao, 2013).

In order for eigenvalue control to be stable enough – both numerically and from the view-point of biological learning – while being computationally more powerful than linear reservoir computing, we therefore examine how to design a plausible solution of an underdetermined version of Equation 11 where there are $1 < K \ll N$ desired eigenvalues outside of the initial ensemble of eigenvalues.

4.2.2.1 Choosing a solution for loop weights implementing eigenvalue control that favors stable dynamics

When the number of desired eigenvalues of the perturbed matrix $\tilde{\mathbf{J}}_\mu$ is much less than the dimensionality of the system, then there are indeed several vectors **v** that solve Equation 11 and we are interested in choosing a solution that is favorable for the stability of the dynamical system. To this end, we solved Equation 11 using the pseudoinverse $\mathbf{P}^+$ of the matrix **P**. More specifically, our approach is to fix:

$$\mathbf{d} \equiv \mathbf{P}^+\mathbb{1} = \mathbf{R}^\top \mathbf{v} \odot \mathbf{Lu} \qquad \text{(Equation 12)}$$

where $\odot$ is the element-wise product. Though this restricts the ensemble of possible thalamocortical weight solutions verifying the equality in Equation 11, there are still infinitely many **u**, **v** that would lead to the same vector **d**. We will take advantage of this remaining freedom in the upcoming section 4.2.3, while here we just remark that for any choice of **u** (still with the only mild restriction that **u** is not orthogonal to any of the rows of **L**), then we can choose:

$$\mathbf{v} = \mathbf{L}^\top \text{diag}(\mathbf{Lu})^{-1}\mathbf{d} = \mathbf{L}^\top \text{diag}(\mathbf{Lu})^{-1}\mathbf{P}^+\mathbb{1},$$

which is Equation 2 from the main text.

As we will now show, the reason why this solution favors the stability of the motif dynamics is that besides imposing the K desired eigenvalues λ_i^{des} among the eigenvalues $\tilde{\lambda}_i$ of $\tilde{\mathbf{J}}_\mu$, it also indirectly constrains the $N - K$ remaining $\tilde{\lambda}_i$ such that their norm tends to be small. This consequently discourages the presence of eigenmodes with very slow exponential decay. Indeed, as indicated above, Equation 2 fixes **d**, such that the following holds for any pair **u**, **v** satisfying Equation 2:

$$1 = \sum_{j=1}^{N} \frac{d_j}{\tilde{\lambda}_i - \lambda_j} = \frac{\sum_{j=1}^{N} d_j \prod_{k \neq j} (\tilde{\lambda}_i - \lambda_k)}{\prod_{k=1}^{N} (\tilde{\lambda}_i - \lambda_k)},$$

which implies

$$0 = \sum_{j=1}^{N} \left(d_j \prod_{k \neq j} \left(\tilde{\lambda}_i - \lambda_k \right) \right) - \prod_{k=1}^{N} \left(\tilde{\lambda}_i - \lambda_k \right). \qquad \text{(Equation 13)}$$

This equation defines the roots of a polynomial where $\tilde{\lambda}_i$ is the variable, and all coefficients are fixed given an initial matrix $\mathbf{J}_{\text{cc}}$ – which constrains the initial eigenvalues $\lambda_1, \ldots, \lambda_N$ – and given $d_1, \ldots, d_N$ which are constrained by $\mathbf{d} \equiv \mathbf{P}^+\mathbb{1}$. Hence, Equation 13 imposes exactly the N values that $\tilde{\lambda}_i$ can take. In other words, when fixing **d** by using the pseudoinverse of **P**, we constrain the whole eigenspectrum of $\tilde{\mathbf{J}}_\mu$ (Figure S2D). Moreover, we choose a very particular solution for **d**: the Moore–Penrose pseudoinverse or the minimal norm solution, which – as can be seen from Equation 13 – in turn tends to minimize the maximum magnitude of the coefficients of the characteristic polynomial (except for the leading coefficient associated with $\tilde{\lambda}_i^N$ which is fixed to -1). This further tends to minimize the Lagrange and Cauchy upper bounds on all of the roots of the characteristic polynomial (Hirst and Macey, 1997). As a consequence, the modulus of the non-controlled eigenvalues tends to be minimized. This is visible in Figure 1E. The purple circles corresponding to the $\tilde{\lambda}_i$ that are not used to match the desired eigenvalues (i.e., that do not overlap with a pink cross) are indeed closer to the center of mass of the distribution than the red dots indicating the initial eigenvalues λ_i. Note-worthily, as a consequence, our method also leads to smaller eigenvalue norm compared to an alternative eigenvalue control procedure (Schuessler et al., 2020) that, given the initial

matrix $\mathbf{J}_{cc}$ and a loop vector $\mathbf{u}$, chooses instead the vector $\mathbf{v}$ that has the smallest possible norm while imposing a number of desired eigenvalues in the effective matrix $\tilde{\mathbf{J}}_\mu$. More specifically, as a consequence of encouraging smaller norm for the $N-K$ eigenvalues that are not set to the desired targets λ_i^{des}, Equation 2 also favors a smaller maximal real part for these eigenvalues (Figure S2E) which supports the stability of the dynamics. There are also other advantages of using Equation 2 rather than the minimal $\mathbf{v}$ norm solution. Indeed, by fixing the whole eigenspectrum of $\tilde{\mathbf{J}}_\mu$, Equation 2 facilitates the analysis of the effect of the connectivity perturbation on the circuit dynamics as well as the numerical optimization of the weights, as we will explain in the Methods Secs. 4.2.3 and 4.2.7 respectively.

In addition, given that eigenvalue control through Equation 2 requires us to take the pseudoinverse of the matrix $\mathbf{P}$, whether or not eigenvalue control will be numerically successful can be predicted by the condition number of the matrix $\mathbf{P}$. In line with our analysis in the above paragraph that clarifies the conditions under which the rows of $\mathbf{P}$ are correlated, the condition number of $\mathbf{P}$ is larger when K increases relative to N, or when the distance between the eigenspectrum of $\mathbf{J_{cc}}$ and the desired eigenvalues increases (Figure S2F). Therefore, eigenvalue control will be successful when the number and relative distance of the desired eigenvalues are not too large for a given cortical size N.

4.2.2.2 Using eigenvalue control to approximate any desired output

Here we explain how, during motif production, this control of the dynamics' timescales by a thalamocortical loop can help sculpt the network output into a desired shape. Under the conditions that the dynamics during a motif are linear, the output is shaped through characteristics of the eigenmodes which act as basis functions. We will clarify in this section why these characteristics come in two categories: first, the exponential and oscillation timescales, which are determined by the effective connectivity through its eigenvalues; and second, the initial amplitude and phase, which can be set by the activity pattern at the beginning of the motif dynamics $\mathbf{c}(0)$ (Figures 1C and 1D).

More specifically, under the very mild assumption that the (left) eigenvectors $\tilde{\mathbf{L}}$ of $\tilde{\mathbf{J}}_\mu$ form a complete basis, the readout weights $\mathbf{w}^\top$ can be expressed as a linear combination of these eigenvectors: $\mathbf{w}^\top = (\mathbf{w}^\top \tilde{\mathbf{L}}^{-1})\,\tilde{\mathbf{L}} = (\mathbf{w}^\top \tilde{\mathbf{R}})\,\tilde{\mathbf{L}}$. Therefore, the readout is also a weighted sum (with weights $\mathbf{w}^\top \tilde{\mathbf{R}}$) of the N eigenmodes whose dynamics are given as:

$$\tilde{\mathbf{L}}\mathbf{c}(t) = \mathrm{diag}\left(e^{(\tilde{\lambda}-1)t}\right)\tilde{\mathbf{L}}\,\mathbf{c}(0), \qquad \text{(Equation 14)}$$

where $\tilde{\lambda}$ is a vector of eigenvalues of $\tilde{\mathbf{J}}_\mu$, and for any vector $\mathbf{v}$ we use the notation $e^{\mathbf{v}}$ for a vector whose elements are the exponentials of the elements of $\mathbf{v}$. We emphasize that here the time variable is defined such that time 0 corresponds to the beginning of motif execution.

Notice that given that the activities are real, the eigenvalues $\tilde{\lambda}$ and eigenvectors $\tilde{\mathbf{L}}$ are either real or come in complex conjugate pairs, such that the imaginary parts of the eigenmodes cancel. Therefore, for each eigenvalue $\tilde{\lambda}_i$, the relevant time course of the eigenmode corresponding to $\tilde{\mathbf{L}}_i$: the i^{th} row of $\tilde{\mathbf{L}}$, $\forall i \in [\mathbb{1},\ldots,N]$, is:

$$\mathrm{Re}\left(\tilde{\mathbf{L}}_i\,\mathbf{c}(t)\right) = e^{\mathrm{Re}\left(\left(\tilde{\lambda}_i-1\right)t\right)}\left(A_i\cos\left(\mathrm{Im}\left(\tilde{\lambda}_i\right)t+\phi_i\right)\right), \text{ where} \qquad \text{(Equation 15)}$$

$$\phi_i = \arctan\frac{\mathrm{Im}\left(\tilde{\mathbf{L}}_i\,\mathbf{c}(0)\right)}{\mathrm{Re}\left(\tilde{\mathbf{L}}_i\,\mathbf{c}(0)\right)} \text{ and } A_i = \frac{\mathrm{Re}\left(\tilde{\mathbf{L}}_i\,\mathbf{c}(0)\right)}{\cos\phi_i}.$$

This explicitly shows how the dynamics of the network breaks down into the exponentially modulated sinusoids that we show in Figures 1C and 1D and that we call basis functions.

As we have mentioned at the beginning of the results section, in order for the network to get a chance to accurately and robustly approximate a given a target trajectory y_μ for motif μ, we can find a small number $K \approx 4-20$ of basis functions with appropriate timescales $[\lambda_i^{des}]_{1\le i\le K}$ that can be combined with complex weights $[\alpha_i^{des}]_{1\le i\le K}$ whose magnitudes are not very large (i.e., they scale with the maximum magnitude of the desired output y_μ) such that $\hat{y}_\mu = \sum_{i=1}^{K}\alpha_i^{des}e^{(\lambda_i^{des}-1)t}$ fits y_μ very well and robustly. For some well-defined functions like the sinc function we show in Figures 1 and 2, appropriate timescales and complex weights can be determined analytically (for the sinc function, we just regularly sampled the frequencies given by the Fourier transform of this function). However, even when it is not possible to determine appropriate $[\lambda_i^{des}]_{1\le i\le K}$ and $[\alpha_i^{des}]_{1\le i\le K}$ analytically, it is easy to do numerically because the number of basis functions K required to get a good fit is small. More specifically, we choose the minimum K value that results into an acceptable error in $\hat{y}_\mu$ compared to y_μ while making sure that the $[\alpha_i^{des}]_{1\le i\le K}$ are not too large (see STAR Methods Sec. 4.2.7 for technical details, and Figures S1A–S1E for examples). By choosing an approximation $\hat{y}_\mu$ which is composed of basis functions that combine with weights of reasonable magnitude, we will allow a thalamocortical network whose dynamics includes the corresponding basis functions to produce the corresponding motif in a noise-robust manner (and with reasonable activity norm), as we will explain below and in the next STAR Methods section 4.2.3.

Indeed, the previous paragraphs of the STAR Methods section showed how it is possible to adjust the weight vector $\mathbf{v}$ to ensure that the eigenvalues $\tilde{\lambda}_i$ of the effective connectivity $\tilde{\mathbf{J}}_\mu$ includes the desired eigenvalues $[\lambda_i^{\text{des}}]_{1\leq i\leq K}$. The only other requirement that needs to be met in order for the network to produce the right output is therefore to control the initial amplitudes A_i and phases ϕ_i of the eigenmodes. We will show below how this can be achieved by adjusting the activity $\mathbf{c}(0)$ at the beginning of the motif execution. Indeed, the N real values of $\mathbf{c}(0)$ are linked through a linear system of N equations with the initial amplitude (and phase, in case of a complex eigenvalue) of the real part of each eigenmode composing the readout at time 0.

More precisely, we can then collect all the eigenvalues of $\tilde{\mathbf{J}}_\mu$ into the vector $\tilde{\boldsymbol{\lambda}}_\mu$ with $[\boldsymbol{\lambda}_\mu]_{i\leq K} = \lambda_i^{\text{des}}$, and define a vector $\boldsymbol{\alpha}_\mu$ such that $[\boldsymbol{\alpha}_\mu]_{i\leq K} = \alpha_i^{\text{des}}$ and $[\boldsymbol{\alpha}_\mu]_{i>K} = 0$. Taking care of ordering the columns of the eigenvector matrix $\tilde{\mathbf{R}}$ in correspondence with the elements of $\tilde{\boldsymbol{\lambda}}_\mu$, we note that $\hat{y}_\mu(0) = \mathbb{1}^\top \boldsymbol{\alpha}_\mu = (\mathbf{w}^\top \tilde{\mathbf{R}})\, \tilde{\mathbf{L}}\, \mathbf{c}(0) = \mathbb{1}^\top \text{diag}(\mathbf{w}^\top \tilde{\mathbf{R}})\, \tilde{\mathbf{L}}\, \mathbf{c}(0)$. Therefore, the initial activities $\mathbf{c}(0)$ need to be set according to:

$$\mathbf{c}(0) = \mathbf{c}_\mu^{\text{init}} \equiv \tilde{\mathbf{R}}\, \text{diag}\left(\mathbf{w}^\top \tilde{\mathbf{R}}\right)^{-1} \boldsymbol{\alpha}_\mu. \qquad \text{(Equation 16)}$$

Note that as a consequence, even with a fixed effective connectivity $\tilde{\mathbf{J}}_\mu$, different motif variations μ, μ' – corresponding to different amplitudes and phases of the eigenmodes composing the readout – can be composed with different initial rates $\mathbf{c}_\mu^{\text{init}}$, $\mathbf{c}_{\mu'}^{\text{init}}$ (Figures 3E and 3F). Though this strategy works for generating a certain variety of motifs, it has some limitations as *(i)* the properties of the eigenvectors can be responsible for associating some desired phases and amplitudes with unreasonable initial activity norm, and *(ii)* the timescales of a linear network with a fixed effective connectivity are restricted to a particular set (Figure 1E; Ahmadian et al., 2015; Hennequin et al., 2014). This constrains how well a fixed linear reservoir can robustly approximate arbitrary target motifs, which also means that such a model cannot approximate arbitrary motifs under the constraints of reasonable activity norm and of using a biologically plausible mechanism to set the initial activity of the network (Figures S1F–S1L).

Thus, the ability to modify the effective connectivity of the circuit – in order to impose desired eigenvalues in the dynamics as we described in this section and to control the eigenvector of the dynamics as described in the next section – is critical to plausibly produce motifs through the circuit dynamics (Figure S1).

4.2.3 Eigenvector control for motif robustness

For the above-proposed low-rank connectivity perturbation mechanism of motif production to be biologically plausible, the network output should also *(i)* be robust to noise in the initial activity pattern $\mathbf{c}_\mu^{\text{init}}$ and *(ii)* be associated with a reasonable activity magnitude. However, a naive 'half-random' rank-one connectivity perturbation approach which is only focused on eigenvalue control (STAR Methods Sec. 4.2.2) can lead to strong noise sensitivity (Figure 2). In this section, we show why this happens and how to leverage the remaining freedom in the low-rank perturbation to fix this.

4.2.3.1 Eigenvalue control through a half-random loop induces eigenvector correlations that compromise readout robustness

As we showed above, our procedure for eigenvalue control – Equation 2 – actually fixes all the eigenvalues of the dynamics, such that control networks with the same eigenvalues but well-behaved eigenvectors actually show good noise robustness (Figures 2H and 2I). In addition, eigenvector correlations have the potential to create large amplification of initial network activity (Ganguli et al., 2008; Murphy and Miller, 2009; Hennequin et al., 2012; Bondanelli and Ostojic, 2020). It is therefore natural to ask how our eigenvalue control procedure affects eigenvectors and how this can in turn affect noise robustness.

To get an expression for the left ($\tilde{\mathbf{L}}$) and right ($\tilde{\mathbf{R}}$) eigenvector matrices of the perturbed connectivity $\tilde{\mathbf{J}}_\mu$, we can start from the basic property of $\tilde{\mathbf{l}}_i^\top$, the i^{th} row of $\tilde{\mathbf{L}}$, to write:

$$\tilde{\lambda}_i\, \tilde{\mathbf{l}}_i^\top = \tilde{\mathbf{l}}_i^\top (\mathbf{J_{cc}} + \mathbf{u}\mathbf{v}^\top) \quad \rightarrow \quad \tilde{\lambda}_i\, \tilde{\mathbf{l}}_i^\top \mathbf{r}_j = \lambda_j \tilde{\mathbf{l}}_i^\top \mathbf{r}_j + \tilde{\mathbf{l}}_i^\top \mathbf{u}\mathbf{v}^\top \mathbf{r}_j, \qquad \text{(Equation 17)}$$

where, in the second step, we right-multiplied by $\mathbf{r}_j$, the j^{th} column of $\mathbf{R}$.

We will now consider $\mathring{\mathbf{l}}_i$ which is a particularly scaled version of $\tilde{\mathbf{l}}_i$ satisfying $\mathring{\mathbf{l}}_i^\top \mathbf{r}_i = 1$. Later, we will take care of renormalizing these arbitrarily scaled eigenvectors $\mathring{\mathbf{l}}_i$ – and $\mathring{\mathbf{r}}_i$ – to get the properly normalized eigenvectors $\tilde{\mathbf{l}}_i^\top$ and $\tilde{\mathbf{r}}_i$ such that $\forall i$, $\tilde{\mathbf{l}}_i^\top \tilde{\mathbf{r}}_i = 1$ as needed. With the former scaling, we can rewrite Equation 17 for the case of $i = j$ to get

$$\tilde{\lambda}_i = \lambda_i + \mathring{\mathbf{l}}_i^\top \mathbf{u}\, \mathbf{v}^\top \mathbf{r}_i \quad \rightarrow \quad \mathring{\mathbf{l}}_i^\top \mathbf{u} = \frac{\tilde{\lambda}_i - \lambda_i}{\mathbf{v}^\top \mathbf{r}_i}.$$

Then, substituting this back into Equation 17 for $i \neq j$ gives

$$\tilde{\lambda}_i \mathring{\mathbf{l}}_i^\top \mathbf{r}_j = \lambda_j \mathring{\mathbf{l}}_i^\top \mathbf{r}_j + \frac{\tilde{\lambda}_i - \lambda_i}{\mathbf{v}^\top \mathbf{r}_i} \mathbf{v}^\top \mathbf{r}_j \quad \rightarrow \quad \mathring{\mathbf{l}}_i^\top \mathbf{r}_j = \frac{\mathbf{v}^\top \mathbf{r}_j \left(\tilde{\lambda}_i - \lambda_i\right)}{\mathbf{v}^\top \mathbf{r}_i \left(\tilde{\lambda}_i - \lambda_j\right)}.$$

This relationship is valid for all j, so we can use matrix notation to write

$$\mathring{\mathbf{l}}_i^{\top}\mathbf{R} = \frac{\tilde{\lambda}_i - \lambda_i}{\mathbf{v}^{\top}\mathbf{r}_i}\,\mathbf{v}^{\top}\mathbf{R}\mathrm{diag}\left(\tilde{\lambda}_i - \lambda\right)^{-1},$$

or

$$\mathring{\mathbf{l}}_i^{\top} = \frac{\tilde{\lambda}_i - \lambda_i}{\mathbf{v}^{\top}\mathbf{r}_i}\,\mathbf{v}^{\top}\mathbf{R}\mathrm{diag}\left(\tilde{\lambda}_i - \lambda\right)^{-1}\mathbf{L}.$$

Following similar steps, we can also write an equation for the right eigenvectors $\mathring{\mathbf{r}}_i$ that are normalized such that $\mathbf{l}_i^{\top}\mathbf{r}_i^{\circ} = \mathbb{1}$:

$$\mathring{\mathbf{r}}_i = \frac{\tilde{\lambda}_i - \lambda_i}{\mathbf{l}_i^{\top}\mathbf{u}}\,\mathbf{R}\mathrm{diag}\left(\tilde{\lambda}_i - \lambda\right)^{-1}\mathbf{L}\mathbf{u}.$$

Finally we need to renormalize the eigenvectors such that, $\forall i$, $\tilde{\mathbf{l}}_i^{\top}\tilde{\mathbf{r}}_i = 1$. We remark that

$$\mathring{\mathbf{l}}_i^{\top}\mathring{\mathbf{r}}_i = \frac{\left(\tilde{\lambda}_i - \lambda_i\right)^2}{\mathbf{v}^{\top}\mathbf{r}_i\,\mathbf{l}_i^{\top}\mathbf{u}}\,\mathbf{v}^{\top}\mathbf{R}\mathrm{diag}\left(\tilde{\lambda}_i - \lambda\right)^{-2}\mathbf{L}\mathbf{u}.$$

Thus, if we define the normalization factor $a_i \equiv (\mathbf{v}^{\top}\mathbf{R}\mathrm{diag}(\tilde{\lambda}_i - \lambda)^{-2}\mathbf{L}\mathbf{u})^{-\frac{1}{2}}$, then we can write the normalized eigenvectors as

$$\begin{cases} \tilde{\mathbf{l}}_i^{\top} = \dfrac{\mathbf{v}^{\top}\mathbf{r}_i}{\tilde{\lambda}_i - \lambda_i}\,a_i\,\mathring{\mathbf{l}}_i^{\top} = a_i\,\mathbf{v}^{\top}\mathbf{R}\mathrm{diag}\left(\tilde{\lambda}_i - \lambda\right)^{-1}\mathbf{L} \\ \tilde{\mathbf{r}}_i = \dfrac{\mathbf{l}_i^{\top}\mathbf{u}}{\tilde{\lambda}_i - \lambda_i}\,a_i\,\mathring{\mathbf{r}}_i = a_i\,\mathbf{R}\mathrm{diag}\left(\tilde{\lambda}_i - \lambda\right)^{-1}\mathbf{L}\mathbf{u}. \end{cases} \quad \text{(Equation 18)}$$

Note that these eigenvector equations can finally be written compactly in matrix form, to get:

$$\begin{cases} \tilde{\mathbf{R}} = \mathbf{R}\,\mathrm{diag}(\hat{\mathbf{u}})\mathbf{A}\,\mathrm{diag}(\mathbf{a})^{-\frac{1}{2}} \\ \tilde{\mathbf{L}} = \mathrm{diag}(\mathbf{a})^{-\frac{1}{2}}\mathbf{A}^{\top}\mathrm{diag}(\hat{\mathbf{v}})\mathbf{L}, \end{cases} \quad \text{(Equation 19)}$$

where $\hat{\mathbf{u}} \equiv \mathbf{L}\mathbf{u}$, $\hat{\mathbf{v}}^{\top} \equiv \mathbf{v}^{\top}\mathbf{R}$, we define $\mathbf{A}$ such that $A_{ij} = 1/(\tilde{\lambda}_j - \lambda_i)$, and $\mathbf{a} = [\mathbf{A}^{\top}\mathrm{diag}(\hat{\mathbf{u}}) \odot \mathbf{A}^{\top}\mathrm{diag}(\hat{\mathbf{v}})]\mathbb{1}$.

Using Equation 18, we can now reflect on the properties of these eigenvectors of $\tilde{\mathbf{J}}_{\mu}$. More specifically, we can make deductions about the alignment between eigenvectors – which we assessed in Figure 2G by the cosine of the angle between them θ_{ij}, defined as $\cos\theta_{ij} = \mathrm{Re}(\tilde{\mathbf{r}}_i^{H}\tilde{\mathbf{r}}_j)/(\left|\tilde{\mathbf{r}}_i\right|\left|\tilde{\mathbf{r}}_j\right|)$, where the superscript H denotes conjugate transposition. The value $\cos\theta_{ij}$ approaches 1 (or, equivalently, the eigenvectors are almost parallel) if the real and imaginary parts of the two eigenvectors have very similar directions, irrespective of the norm of these vectors. Hence, from Equation 18, we can predict that $\cos\theta_{ij}$ will be large for the vectors $\tilde{\mathbf{r}}_i$ and $\tilde{\mathbf{r}}_j$ if the vectors $\mathbf{R}\mathrm{diag}(\tilde{\lambda}_i - \lambda)^{-1}\mathbf{L}\mathbf{u}$ and $\mathbf{R}\mathrm{diag}(\tilde{\lambda}_j - \lambda)^{-1}\mathbf{L}\mathbf{u}$ have similar directions. This happens when $\tilde{\lambda}_i$ and $\tilde{\lambda}_j$ are much closer to one another than typical eigenvalues in the initial spectrum are, and/or if these eigenvalues are far from the initial spectrum as this drives both above-mentioned vectors toward $\mathbf{0}$ (Figures 1E, 2G, S3A, and S3B). These considerations most notably allow us to conclude that the eigenvector correlations are likely to be larger in cases when there are many controlled eigenvalues that are sizably different from the original eigenvalues (Figure 2G versus Figure S3G). In addition, we can similarly understand why relatively large eigenvector correlations are also seen for 'uncontrolled' eigenvalues (that do not match any of the K desired eigenvalues in the matrix $\mathbf{P}$) that happen to be very close to one another, even at the center of the eigenvalue distribution (for instance among the eigenvectors whose eigenvalues have very small norm, Figures S3A and S3B).

Why are these eigenvector correlations causing large and amplified activity transients in the network – including in the readout direction, therefore leading to readout noise sensitivity? This can be understood by examining the expression for the expected integral of the square activity norm with correlated versus 'control' non-correlated eigenvectors (i.e., these eigenvectors verify $\tilde{\mathbf{R}}_{\mathrm{norm}}^{H} = \tilde{\mathbf{R}}_{\mathrm{norm}}^{-1}$ and $\tilde{\mathbf{L}}_{\mathrm{norm}}^{H} = \tilde{\mathbf{L}}_{\mathrm{norm}}^{-1}$ where H denotes conjugate transposition and norm stands for 'normal': the name of a matrix with 'uncorrelated' unitary eigenvectors).

In the latter case, we get:

$$a_{\mathrm{norm}} = E_{\delta\mathbf{c}_0}\left[\int_0^{t_\mu} dt\;\mathbf{c}^{\top}\;\mathbf{c}\right] = E_{\delta\mathbf{c}_0}\left[\int_0^{t_\mu} dt\;\mathrm{Tr}\left(\mathbf{c}\,\mathbf{c}^{H}\right)\right] = \varepsilon^2\sum_i \frac{e^{2\mathrm{Re}\left(\tilde{\lambda}_i - 1\right)\,t_\mu} - 1}{2\mathrm{Re}\left(\tilde{\lambda}_i - 1\right)}, \quad \text{(Equation 20)}$$

where ε^2 is the variance of the initial i.i.d. rates $\delta\mathbf{c}_0$ and t_μ is the duration of the motif, and we proceeded from the expression of the dynamics of $\mathbf{c}(t)$ as a function of the eigenmode's dynamics (Equation 14). Notice that the unitary eigenvectors – which we abusively

refer to as 'orthogonal' in the main text to ease interpretability – cancel during the algebraic developments and do not contribute to the final expression. In addition, notice how the different timescales of eigenmodes (i.e., $1/\mathrm{Re}\,(\tilde{\lambda}_i - 1)$) are simply added together and do not otherwise interact.

In contrast, when the eigenvectors are correlated, the expected squared activity norm deviates from Equation 20 to give:

$$a_{\mathrm{corr}} = \varepsilon^2 \mathrm{Tr}\left[\tilde{\mathbf{R}}\left(\left(\tilde{\mathbf{L}}\,\tilde{\mathbf{L}}^H\right) \odot \mathbf{K}\right)\tilde{\mathbf{R}}^H\right], \qquad \text{(Equation 21)}$$

where we defined the matrix $\mathbf{K}$ with elements $K_{ij} = (e^{(\tilde{\lambda}_i + \tilde{\lambda}_j^H - 2)t_\mu} - 1)/(\tilde{\lambda}_i + \tilde{\lambda}_j^H - 2)$.

When comparing Equations 20 and 21, it is clear that the correlated eigenvectors in Equation 20 lead to magnified interactions between the dynamics of eigenmodes which combine into the neural activities: pairs of eigenvalues interact in the matrix $\mathbf{K}$, and these interaction terms can be multiplied by large numbers contained in the matrices $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{R}}$ arising because these ill-conditioned, correlated matrices are the inverses of one another. Under certain conditions (Bondanelli and Ostojic, 2020), these eigenvector correlations can lead to very large amplified dynamics where the activities will undergo transient growth even though all eigenvalues generate asymptotically stable dynamics. In addition, it is possible to observe a more 'moderate' type of amplification that leads to a slower decrease of activity compared to what would be expected from a matrix with the same eigenspectrum but orthogonal eigenvectors, as tends to occur in random Gaussian matrices (Chalker and Mehlig, 1998).

These considerations lead to the definition for our 'control' matrices in Figures 2H, 2I, S3E, and S3H: they have the same eigenvalues as our motif production matrix $\tilde{\mathbf{J}}_\mu$, but their eigenvectors are generated either from the eigenvectors of a random Gaussian matrix or a normal matrix. Therefore, these control matrices are of the form $\mathbf{R}_c \mathrm{diag}(\tilde{\lambda})\mathbf{R}_c^{-1}$ – where the eigenvector matrices $\mathbf{R}_c$ were generated according to one of the following procedures:

- ***Eigenvectors from Gaussian matrices***

we created random Gaussian matrices from which we selected those with the same number of real eigenvalues as their matched $\tilde{\mathbf{J}}_\mu$. We then extracted their eigenvectors through eigendecomposition, and finally aligned real eigenvectors to real eigenvalues and complex conjugate eigenvector pairs to complex conjugate eigenvalue pairs of $\tilde{\boldsymbol{\lambda}}$.

- ***Eigenvectors from Normal matrices:***

we created random normal eigenvectors with appropriate numbers of real and complex conjugate pairs with the help of the method described in Mezzadri (2007). We first created N random real orthogonal eigenvectors $\mathbf{R}_c^{\mathrm{Re}}$ using the QR-decomposition based methodology developed in Mezzadri (2007) to uniformly sample the orthogonal group. We then created the appropriate n_{Im} number of complementary pairs of complex conjugate eigenvectors $\mathbf{R}_c^{\mathrm{Im}}$ through multiplication of an n_{Im} subset of the columns of $\mathbf{R}_c^{\mathrm{Re}}$, by the eigenvectors of an n_{Im} dimensional real random orthogonal matrix created with the same methodology and selected to only have complex conjugate eigenvalues (which was very common). Then, by concatenating the $N - n_{\mathrm{Im}}$ columns of $\mathbf{R}_c^{\mathrm{Re}}$ that we had left aside, with the complementary n_{Im} columns of $\mathbf{R}_c^{\mathrm{Im}}$, we could create a complete set of random unitary eigenvectors with the appropriate number of real and complex conjugate columns.

4.2.3.2 Readout robustness through full control of the loop weights

Our procedure for eigenvalue control fixes all eigenvalues by constraining one vector of the rank-one perturbation – here, $\mathbf{v}$ – while leaving the vector $\mathbf{u}$ unconstrained (STAR Methods section 4.2.2; Figure S2D). Here, we will show how we can take advantage of these remaining degrees of freedom to modify eigenvectors so as to improve the robustness of the readout. More specifically, we will choose the vector $\mathbf{u}$ to optimize a cost function $C(\mathbf{u})$ which we define as the integrated squared deviation in the network output due to a Gaussian fluctuation $\boldsymbol{\eta}$ in the initial conditions with i.i.d. elements $\eta_i \sim \mathcal{N}(0, \sigma^2(\mathbf{u}))$. Importantly, to get robustness to noise whose magnitude scales with the 'signal', we choose to scale the variance of $\boldsymbol{\eta}$ with the time-averaged squared norm of the activities $\mathbf{c}(t)$ when the initial conditions are exactly $\mathbf{c}_\mu^{\mathrm{init}}$. Therefore, defining $\hat{y}_{\mu,\boldsymbol{\eta}}$ as the network output when adding noise $\boldsymbol{\eta}$ to the initial conditions $\mathbf{c}_\mu^{\mathrm{init}}$, we get:

$$C(\mathbf{u}) = \frac{1}{t_\mu} E_{\boldsymbol{\eta}} \int_0^{t_\mu} dt \left(\hat{y}_{\mu,\boldsymbol{\eta}} - \hat{y}_\mu\right)^2 = \frac{1}{t_\mu} E_{\boldsymbol{\eta}} \left[\int_0^{t_\mu} dt \left(\mathbf{w}^\top \tilde{\mathbf{R}}\, \mathrm{diag}\left(e^{(\tilde{\lambda}-1)t}\right)\tilde{\mathbf{L}}\boldsymbol{\eta}\right)^2\right] = \frac{\sigma^2(\mathbf{u})}{t_\mu} \mathbf{w}^\top \tilde{\mathbf{R}}\left(\left(\tilde{\mathbf{L}}\tilde{\mathbf{L}}^\top\right) \odot \boldsymbol{\Lambda}\right)\tilde{\mathbf{R}}^\top \mathbf{w} \qquad \text{(Equation 22)}$$

and:

$$\sigma^2(\mathbf{u}) = \frac{1}{N t_\mu} \int_0^{t_\mu} dt \|\mathbf{c}\|^2 = \frac{1}{N t_\mu} \int_0^{t_\mu} dt \|\tilde{\mathbf{R}}\, \mathrm{diag}\left(e^{(\tilde{\lambda}-1)t}\right)\tilde{\mathbf{L}}\, \mathbf{c}_\mu^{\mathrm{init}}\|^2 = \frac{1}{N t_\mu} \mathbf{c}_\mu^{\mathrm{init}\,\top} \tilde{\mathbf{L}}^\top \left(\left(\tilde{\mathbf{R}}^\top \tilde{\mathbf{R}}\right) \odot \boldsymbol{\Lambda}\right)\tilde{\mathbf{L}}\, \mathbf{c}_\mu^{\mathrm{init}}, \qquad \text{(Equation 23)}$$

where we defined $E_{\boldsymbol{\eta}}$ as an expectation value, t_μ as the duration of the motif, $\odot$ as the component-wise Hadamard product, and $\boldsymbol{\Lambda}$ as a matrix with components:

$$\Lambda_{ij} = \frac{e^{(\tilde{\lambda}_i + \tilde{\lambda}_j - 2)t_\mu} - 1}{\tilde{\lambda}_i + \tilde{\lambda}_j - 2}. \qquad \text{(Equation 24)}$$

Note that $C(\mathbf{u})$ and $\sigma^2(\mathbf{u})$ depend on $\mathbf{u}$ through the eigenvector matrices $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{L}}$ (both directly and indirectly through $\mathbf{c}_\mu^{\mathrm{init}}$ in Equation 16).

Of note our expressions for the eigenvectors of $\tilde{\mathbf{J}}_\mu$, Equation 18, make it clear that the cost function C (Equation 22) can be in general expressed as a ratio of polynomials where the variables are the entries of the weight vector $\mathbf{u}$ that we want to optimize. This implies that there are several local minima in weight space for C (see Mehta et al., 2018 for a similar result in linear feedforward networks), which justifies the use of a non-local optimization method to minimize C (see STAR Methods section 4.2.7).

After optimizing the vector $\mathbf{u}$ to minimize $C(\mathbf{u})$, we indeed observe subtle changes in the eigenvector directions (Figures 2G, S3A, S3B, and S3G) that succeed at decreasing the network output's sensitivity to noise in the initial conditions to similar levels as control matrices with the same eigenvalues but well-behaved eigenvectors (Figures 2H and S3H). This can be achieved even though the amount of non-normal amplification – as quantified by the average activity norm of the network's response when initialized with random uncorrelated values – is somewhat reduced but still large compared to the control matrices (Figure 2I). This implies that amplification of noise added to the initial conditions can still be present, but is constrained to directions that are quasi-orthogonal to the readout, which we indeed verified numerically (Figure S3D).

Note that, when optimizing $C(\mathbf{u})$, the scaling of the noise with the norm of the activity during noiseless motif production – as opposed to a constant scaling – fulfills two different functions:

- First, it prevents the development of a dynamics that would give the illusion of noise robustness by cranking up the activity norm during noiseless production. Indeed, as mentioned above, in cases when there are many controlled eigenvalues leading to levels of non-normal amplification that cannot be fully reduced when optimizing the vector $\mathbf{u}$, the optimization can force the directions of maximal noise amplification to be orthogonal to the output weights $\mathbf{w}$. To achieve this, a naive solution is to make all the directions along which the activity can become large – which tend to correlate with the eigenvectors – nearly orthogonal to $\mathbf{w}$. However, this would mean that the eigenvectors corresponding to the controlled eigenvalues would also be poorly aligned with $\mathbf{w}$ and thus that the cortical activities would need to be very large so that the projections of the needed basis functions onto $\mathbf{w}$ would be of sufficient magnitude to construct the output (as is visible from Equation 16). By scaling the noise term in our cost function with the norm of the activity during noiseless motif production, we prevent this implausible solution and instead create a trade-off between how much noise can be redirected to output null directions and the signal to noise ratio of the activity during motif production.
- Second, we effectively implement a form of soft-bounding of the activity norm, such that – conversely to the cases when the eigenvectors are not shaped to favor motif production – the activity norm cannot become extremely large during noiseless motif production (for instance, compare the maximum norms for fully-tuned loops to some large outlier norms for the half-random loop and the control matrices in Figure S3E).

However, Equations 22 and 23 do leave some freedom on the activity norm during motif production. Therefore, the activity norm is still free to be moderately large – especially when adding noise to the dynamics. This indicates that, similarly to what is observed in motor cortex (Russo et al., 2018), the maximal activity patterns can occur in directions that are not fully aligned with the readout. This is for instance visible in Figure S3E, as the activity norm for the fully-tuned loop is still larger than the magnitude of the network output. Indeed, given that the output weights are a Gaussian random vector of expected norm one, this indicates that the angle between the output vector and the activity vector is relatively large. This phenomenon can be understood as a consequence of leveraging the relative absence of constraints placed on non-output directions. Indeed, given that the optimization procedure uses a limited number of parameters to control a much higher-dimensional dynamical system, it takes advantage of any degree of freedom – for instance by not forcing the alignment of $\mathbf{w}$ with the directions along which larger activity patterns lie during noiseless motif production, which may be difficult to realize while simultaneously orthogonalizing the directions of noise amplification with $\mathbf{w}$.

Remarkably, even though we defined $C(\mathbf{u})$ to only ensure that the activities would be of reasonable magnitude when starting from the pattern $\mathbf{c}_\mu^{\text{init}}$ corresponding to a particular set of amplitudes and phases of the controlled eigenmodes, this property of the activities also appeared to often generalize to other amplitudes and phases of these modes (Figure 3). If needed, to further expand the ensemble of initial amplitudes and phases of the eigenmodes that correspond to reasonable activity norms, it would be conceivable to modify the scaling of the noise in $C(\mathbf{u})$ to an average over the mean activity when initializing the network from several initial conditions.

An interesting additional consequence of our procedure that adjusts the full thalamocortical loop to improve noise robustness is that it also decreases both the average norm, and the variance of the norm of the perturbation $\mathbf{u}\mathbf{v}^\top$ (given a particular set of desired eigenvalues but over different instantiations of $\mathbf{w}$ and $\mathbf{J}_{cc}$, Figures S3C and S4).

For motifs that are associated with relatively large changes of the eigenspectrum – and for which the motif's dynamics is therefore more non-normal – a final noteworthy consequence of the minimization of $C(\mathbf{u})$ is an improvement of the robustness of the dynamics with respect to noise in the synaptic weights (Figure S5F). This effect was somewhat expected as a consequence of the relative decrease of non-normality of the effective connectivity matrix when optimizing $\mathbf{u}$ (Figures 2G and 2I), as non-normality is associated with larger sensitivity of the eigenspectrum to this type of perturbation (Trefethen, 2005). In consequence, though the framework that we propose in this paper does assume some relatively precise tuning of the weights, the dynamics can still be robust to small levels of noise in the weights (noise matrix with Frobenius norm 0.1% of $\mathbf{J}_{cc}$'s Frobenius norm, Figure S5).

In conclusion, we have shown that it is possible to fully tune a single thalamocortical loop to both *(i)* modulate the eigenspectrum of the dynamics to include a small number K of eigenvalues, and *(ii)* robustly readout a desired linear combination of these desired

eigenmodes. In this article, we demonstrate this by either pushing the control of the dynamics to $K = 20$ eigenvalues that are pretty different from those of $\mathbf{J}_{cc}$ ('hard' scenario, Figures 2 and 3); or considering in milder cases with $K = 4$ eigenvalues ('easier' scenario, Figures 4, S3F, and S3G).

4.2.4 Motor preparation through thalamic loops

Previously, we have explained how the successful production of a motif μ relies on the activities being initialized by a motif-specific pattern $\mathbf{c}_\mu^{\text{init}}$. This brings into question the biological implementation needed to drive the activities toward the right pattern before motif initiation. We propose to combine the use of a motif-specific cortical input and motif-independent thalamocortical loops for implementing this motif-preparation mechanism. Notice how this implementation will enable motif transitions during sequences without any network weights being specifically tuned to peculiar transitions between two given motifs, so that motifs can be learned independently before stringing them into arbitrary sequence orders.

More specifically, we assume that the corticothalamic circuit during the preparatory period is given as $\tilde{\mathbf{J}}_{\text{prep}} = \mathbf{J}_{cc} + \mathbf{J}_{ct}\mathbf{S}_{\text{prep}}\mathbf{J}_{tc} \equiv \mathbf{J}_{cc} + \mathbf{U}\mathbf{V}^\top$, where $\mathbf{S}_{\text{prep}}$ selects an ensemble of thalamic units that are specific to the preparatory period (see above section 4.2.1). When adding a cortical input $\mathbf{x}_\mu$, the effective dynamics reads:

$$\dot{\mathbf{c}} = \left(\tilde{\mathbf{J}}_{\text{prep}} - \mathbf{I}\right)\mathbf{c} + \mathbf{x}_\mu.$$

If all the eigenvalues of $\tilde{\mathbf{J}}_{\text{prep}}$ have real part less than one (which we will be able to enforce as described below), regardless of the initial activities, the network will then converge toward $\mathbf{c}_\mu^{\text{init}}$ – the steady-state that can be found by setting $\dot{\mathbf{c}} = 0$ – if:

$$\mathbf{x}_\mu = -\left(\tilde{\mathbf{J}}_{\text{prep}} - \mathbf{I}\right)\mathbf{c}_\mu^{\text{init}}. \qquad \text{(Equation 25)}$$

Importantly, the dynamics of decay to steady-state are independent of the input and therefore of the identity of the upcoming motif, as can be seen by considering the variable $\delta\mathbf{c} = \mathbf{c} - \mathbf{c}_\mu^{\text{init}}$:

$$\dot{\delta\mathbf{c}} = \left(\tilde{\mathbf{J}}_{\text{prep}} - \mathbf{I}\right)\delta\mathbf{c} \quad \rightarrow \quad \delta\mathbf{c}(t) = \mathbf{R}_{\text{prep}}\,\text{diag}\left(e^{\left(\lambda_{\text{prep}}-1\right)t}\right)\mathbf{L}_{\text{prep}}\,\delta\mathbf{c}_0, \qquad \text{(Equation 26)}$$

where $\delta\mathbf{c}_0$ is the vector containing the values of $\delta\mathbf{c}$ at the beginning of the preparatory period, $\mathbf{R}_{\text{prep}}$ and $\mathbf{L}_{\text{prep}}$ are the right and left eigenvector matrices of $\tilde{\mathbf{J}}_{\text{prep}}$ respectively, and λ_{prep} is a vector containing the eigenvalues of $\tilde{\mathbf{J}}_{\text{prep}}$.

The starting state of the preparatory period $\delta\mathbf{c}_0$ depends on the ending state of the previous motif which we assume is unknown, and to implement a good motor preparation it is desirable that – regardless of this state – $\delta\mathbf{c}$ goes fast to 0. Therefore, to reach this goal, we will minimize the average – over random iid values for the entries of $\delta\mathbf{c}_0$ – of the integrated squared norm of $\delta\mathbf{c}$.

In addition, ideally, while a motif is being prepared, the readout should smoothly interpolate between the previous and upcoming motifs, even though the single units may undergo fast and large transients. Therefore, we also include a smoothness contribution to our cost function, that we design to be the integrated square of the first derivative of the readout.

Our total cost function is therefore:

$$C(\mathbf{U},\mathbf{V}) = \frac{1}{N}\,E_{\delta\mathbf{c}_0}\left[\int_0^\infty dt\,||\delta\mathbf{c}||^2\right] + \beta\,E_{\delta\mathbf{c}_0}\left[\int_0^\infty dt\left(\frac{d}{dt}\mathbf{w}^\top\delta\mathbf{c}\right)^2\right]$$

$$\propto \frac{1}{N}\text{Tr}\left(\mathbf{R}_{\text{prep}}\left(\left(\mathbf{L}_{\text{prep}}\,\mathbf{L}_{\text{prep}}^\top\right)\odot\mathbf{Z}\right)\mathbf{R}_{\text{prep}}^\top\right) + \beta\,\mathbf{w}^\top\mathbf{R}_{\text{prep}}\left(\left(\mathbf{L}_{\text{prep}}\,\mathbf{L}_{\text{prep}}^\top\right)\odot\boldsymbol{\Gamma}\right)\mathbf{R}_{\text{prep}}^\top\mathbf{w}\,, \qquad \text{(Equation 27)}$$

where $Z_{ij} = -1/\left(\lambda_i^{\text{prep}} + \lambda_j^{\text{prep}} - 2\right)$ and $\Gamma_{ij} = \left(\lambda_i^{\text{prep}} - 1\right)\left(\lambda_j^{\text{prep}} - 1\right)Z_{ij}$. Finally, N is the number of cortical units and β is a hyperparameter which trades off the relative importance of transition speed (left part of the cost function) and readout smoothness (right part of the cost function). Notice that this cost is not impacted by the shape of the initial rates' distribution. Notice also that while our cost function relates to the linear-quadratic problem from control theory ((Kao et al., 2021)), using semi-numerical gradient-descent optimization - with an analytically computed gradient through the eigenvalues and eigenvectors, see Methods section 4.2.7 - is fast and enables direct control of the number of thalamic units involved (Figures 3 and S6C) and direct regularization of the loop weights (instead of an indirect regularization through the norm of the control input, Figures S6J and S6K). Notably, the ability to control the number of thalamic units was very important in the context of our article which emphasizes how, even with the biological constraint that the size of thalamus is much smaller than the size of cortex (Halley and Krubitzer, 2019), thalamus can act as a powerful controller of cortex. Finally, we also want to stress that the simple optimization procedure suggested above can be easily extended to continuously nonlinear dynamics (Logiaco and Escola, 2020).

After optimization, the resulting weights $\mathbf{U}\mathbf{V}^\top$ indeed lead to a fast decay of the rate dynamics toward steady-state, which is supported by the fact that all eigenvalues of $\tilde{\mathbf{J}}_{\text{prep}}$ have real part significantly smaller than one (Figures 3C, S6A, S6D, and S6J). The convergence is faster when increasing the rank of $\mathbf{U}\mathbf{V}^\top$ (Figures 3D and S6C). The longer the preparation phase is, the closer the

network activities get to the appropriate pattern $\mathbf{c}_{\mu}^{\text{init}}$ to start the next motif, and therefore the more accurate the upcoming motif will be (Figures S6G–S6I). In addition, consistently with recent experimental results (Al Borno et al., 2020), if the upcoming motif is 'hard' – in our case, meaning that it requires a larger remapping of the cortical dynamics – a small variation of the activities relative to $\mathbf{c}_{\mu}^{\text{init}}$ at the end of the preparatory period will lead to a larger variation of the output during motif execution. This is because it is more difficult to optimize the cost $C(\mathbf{u})$ for these movements (Equation 22) which means that their accuracy will be relatively more dependent on the network activities being closer to $\mathbf{c}_{\mu}^{\text{init}}$ at the start of motif execution.

In addition, during motif preparation, the smoothness of the readout arises from a 'slower' eigenvalue (with a real part that is a little larger, Figure S6A) that corresponds to an eigenvector that aligns with the readout (Figure S6B). Larger values of β cause this eigenvalue to have larger real part (Figure S6A), and an appropriate value of β creates a readout timescale that monotonically fills the gap between motifs during motor preparation (Figure 4, $\beta = 0.05$). Therefore, while the initial state can be prepared quickly through fast non-monotonic activity transients in single units, the readout can undergo a slightly slower and smoother evolution - more specifically, an approximate exponential relaxation - that interpolates between two motifs (Figure S6F). Note that this interpolation mechanism will work regardless of the values at which the previous motif ends or the next motif starts. Also, note that this mechanism does not enable a precise fit of the desired output during motif transitions -though nonlinear dynamics can mitigate this issue (Logiaco and Escola, 2020)- which is the reason why we use short transition periods during which our network is prepared to produce the next motif. Data shows that motor preparation can indeed be short (Lara et al., 2018b). This short preparation of the dynamic generator circuit is however compatible with the planning of the movement occurring earlier in a separate effective 'premotor' circuit (Zimnik and Churchland, 2021).

Our cost function Equation 22 is only designed for functional purposes, in line with our general approach. Without additional constraints, the synaptic weights of the loops $\mathbf{U}\mathbf{V}^{\top}$ tend to get large (Figure S6E). However, both terms of the cost function are impacted surprisingly little by the scale of the resulting matrix $\mathbf{U}\mathbf{V}^{\top}$ such that almost identical performances are observed over a wide range of norms of this perturbation (Figures S6F–S6I). This scaling of the weights basically leaves unchanged the eigenvalues with larger real part – which are the limiting factor for the speed of the dynamics – while scaling the norm of an ensemble of very negative eigenvalues whose number matches the rank of $\mathbf{U}\mathbf{V}^{\top}$ (Figure S6D). As expected given the relative robustness of the loss relative to the norm of the perturbation, it is also possible to constrain the norm of the perturbation during optimization, which leads to tighter bounds of the eigenspectrum without sizably affecting the performance (Figures S6J–S6L).

4.2.5 Relation to more realistic neuronal dynamics

We are now going to relate the simplified rate equations that we used in this article (Equations 3 and. 4), to more constrained and realistic neuronal dynamics including a reasonable timescale in thalamus (ten times faster than cortex, compatible with the absence of recurrent excitation in thalamus and/or differences in synaptic channels' timescales) and a rectifying non-linearity ensuring rate positivity. More specifically, we will show that we can relate the simpler rate equations that we presented in Methods section 4.2.1 to the following model:

$$\begin{cases} \dot{\mathbf{c}}_{\mathbf{abs}} = -\left(\mathbf{c}_{\mathbf{abs}} - \overline{\mathbf{c}}\right) + \mathbf{J}_{\mathbf{cc}}\left(\left[\mathbf{c}_{\mathbf{abs}}\right]^{+} - \overline{\mathbf{c}}\right) + \mathbf{J}_{\mathbf{ct}}^{\mathbf{act}}\left[\mathbf{t}_{\mathbf{act}}\right]^{+} \\ \dot{\mathbf{t}}_{\mathbf{act}} = \frac{1}{\tau_t}\left(-\left(\mathbf{t}_{\mathbf{act}} - \overline{\mathbf{t}_{\mathbf{act}}}\right) + \mathbf{J}_{\mathbf{tc}}^{\mathbf{act}}\left[\mathbf{c}_{\mathbf{abs}}\right]^{+}\right) \end{cases} \quad \text{(Equation 28)}$$

where the subscript 'abs' indicates that these variables are (mostly) positive and the subscript 'act' indicates the subset of thalamic units that are interacting with cortex (and not silenced by the basal ganglia). In this section, we will explain how – under certain assumptions – these rate equations can be mathematically related to both the population activity of spiking neurons and the simplified switching linear dynamics we use in the main text. In the next section (STAR Methods Sec. 4.2.6), we will then describe a methodology for extending the main text's results to the richer dynamics of Equation 28. The results of numerically implementing this method to demonstrate that our framework can be extended to the more complex dynamics of Equation 28 are shown in Figure S7.

The ensemble of assumptions that we are making along the way, as well as their experimental and theoretical justifications, are summarized in the table below.

Summary of model assumptions and their justifications

	Assumption	Source of evidence	References
1	linear readout from recurrent motor cortical population dynamics as an approximate motor output	anatomy/ electrophysiology	Harrison et al., 2012; Kaneko, 2013/Shenoy et al., 2013; Churchland et al., 2012; Churchland and Cunningham, 2014; Russo et al., 2018
2	quasi-linear input-output relation in balanced neural populations; fit of motor cortical population activity with linear dynamics during individual behaviors	theory/ electrophysiology	van Vreeswijk and Sompolinsky, 1996; 1998; Renart et al., 2010/Churchland and Cunningham, 2014; Lara et al., 2018a
3	non-recurrent thalamus bidirectionally connected to cortex	anatomy	Arcelli et al., 1997; Sherman, 2016; Harris et al., 2019

(*Continued on next page*)

Continued

	Assumption	Source of evidence	References
4	slower cortical dynamics (relative to thalamus) because of cortical excitatory recurrence (absent in the thalamus)	theory/ electrophysiology	Seung et al., 2000; Lim and Goldman, 2013/Shenoy et al., 2013; Churchland and Cunningham, 2014
5	quasi-rectified linear tonic firing in thalamus	electrophysiology	Sherman, 2001; Devergnas et al., 2016; but see Kim et al., 2017
6	strong inhibitory connections from basal ganglia to thalamus	electrophysiology/ optogenetics	Deniau and Chevalier, 1985; but see Schwab et al., 2020/Edgerton and Jaeger, 2014; Kim et al., 2017; Aoki et al., 2019
7	sequence-element-related activity in basal ganglia: (1) sustained during motifs and (2) phasic at switch times	electrophysiology/ optogenetics	Jin and Costa, 2010; Jin et al., 2014/Geddes et al., 2018

To link our tractable switching linear mathematical framework (Sec. 4.2.1) with more biologically plausible neuronal dynamics, we first note that – under some plausible dynamical regimes – the firing rate in a population of spiking neurons can be well-approximated using a two-stage process that leads to nonlinear rate equations (Ostojic and Brunel, 2011) that are similar to Equation 28. First, the synaptic inputs are linearly filtered – for instance, for Leaky-Integrate-and-Fire neurons, with an exponential filter through feeding the input into a first order linear differential equation – which gives rise to a 'voltage-like' variable. Second, this 'voltage-like' variable is passed through a static non-linearity, giving rise to an effective rate variable. The static non-linearity can be qualitatively described as rectified-linear above a certain threshold value θ of the voltage-like variable $\mathbf{v}$ (symbolized as $[\mathbf{v}-\theta]^+$). Directly following this framework, we can write dynamics for cortical populations with 'voltage-like' variables $\mathbf{v}$ and resting voltage $\mathbf{v_{rest}}$; recurrently interacting through the effective connectivity matrix $\mathbf{J_{cc}}$; responding to a thalamic input resulting from the projection of the thalamic rates $[\mathbf{t_{abs}}]^+$ through the thalamocortical weights $\mathbf{J}_{ct}$, and to an input $\mathbf{p}_\mu$ which stays constant during the motif μ:

$$\dot{\mathbf{v}} = -(\mathbf{v}-\mathbf{v_{rest}}) + \mathbf{J_{cc}}[\mathbf{v}-\theta]^+ + \mathbf{J}_{ct}[\mathbf{t_{abs}}]^+ + \mathbf{p}_\mu. \quad \text{(Equation 29)}$$

This equation directly maps to the nonlinear rate dynamics of Equation 28. Note that the connectivity matrices are 'effective' in the sense that rate units can impact one another both positively and negatively. This effective connectivity may be mapped to a biological network with separated inhibition and excitation (Gerstner et al., 2009), by assuming that each rate unit actually maps to two subpopulations: one with excitatory neurons who have slower population dynamics described by Equation 29, and another one with just inhibitory neurons that have faster dynamics (Mensi et al., 2012). Then, an effective excitatory connection between units naturally map to a projection from the excitatory neurons of the sending unit to the excitatory neurons of the receiving unit. In addition, an effective inhibitory connection between populations can be mapped to a disynaptic pathway involving the excitatory neurons of the sending population projecting toward the fast inhibitory neurons of the receiving population that inhibit their respective excitatory 'neighbors' (Gerstner et al., 2009).

We now turn to showing how these nonlinear rate equations relate to the simplified switching linear dynamics we use in the main text. We will take the example of the dynamics during a particular motif μ corresponding to a thalamic selection matrix $\mathbf{S}_\mu$ as in Equation 6, but the derivation directly generalizes to any thalamic selection matrix $\mathbf{S}_s$ with an arbitrary number and location of ones on the diagonal.

We can define $\mathbf{c_{abs}} = \mathbf{v} - \theta$ and assume that the cortical dynamics stay in the linear regime to get:

$$\dot{\mathbf{c}}_{\mathbf{abs}} = -(\mathbf{c_{abs}} - \bar{\mathbf{c}}) + \mathbf{J_{cc}}(\mathbf{c_{abs}} - \bar{\mathbf{c}}) + \mathbf{J}_{ct}[\mathbf{t_{abs}}]^+, \quad \text{(Equation 30)}$$

where $\bar{\mathbf{c}} = (\mathbf{I} - \mathbf{J_{cc}})^{-1}(\mathbf{v_{rest}} - \theta + \mathbf{p}_\mu)$. Following a similar model for the population firing rate of spiking neurons in the thalamus as for cortex, we write the thalamic rates $[\mathbf{t_{abs}}]^+ = [\mathbf{v_{thal}} - \theta_{\mathbf{thal}}]^+$ as the rectification of a voltage-like variable $\mathbf{t_{abs}}$ itself undergoing linear dynamics which, in absence of input, revolve around a baseline $\bar{\mathbf{t}}$:

$$\tau_t \dot{\mathbf{t}}_{\mathbf{abs}} = -\left(\mathbf{t_{abs}} - \bar{\mathbf{t}}\right) + \mathbf{J}_{tc}\,\mathbf{c_{abs}} + \mathbf{q}_\mu \quad \text{(Equation 31)}$$

The inputs to thalamus include cortical projections through the effective weights $\mathbf{J}_{tc}$. Positive cortico-thalamic connections can occur through direct projections from cortex, while inhibitory effective connections can correspond to an indirect cortical projection relayed by the thalamic reticular nucleus (Arcelli et al., 1997). In addition, thalamus receives an input from the basal ganglia $\mathbf{q}_\mu$.

We consider the case for which $\mathbf{q}_\mu$ consists of either sustained strong inhibition during a motif, or a complete release from inhibition (zero drive) in a small subset of thalamic neurons which we will index as rows with the value 1 in the diagonal of an otherwise zero matrix $\mathbf{S}_\mu$. In addition, we consider the limit in which thalamus is very fast relative to cortex ($\tau_t \to 0$). Indeed, thalamus lacks recurrent excitation, making it react at the relatively fast timescales that are intrinsic to a single-neuron, while the recurrently connected cortical populations can be much slower (Seung et al., 2000; Lim and Goldman, 2013). Hence, we assume that thalamus almost instantly follows its input, so Equation 31 becomes:

$$\mathbf{S}_\mu\,\mathbf{t}_{\mathbf{abs}} = \mathbf{S}_\mu\,\bar{\mathbf{t}} + \mathbf{S}_\mu\,\mathbf{J}_{\mathrm{tc}}\,\mathbf{c}_{\mathbf{abs}} \quad \text{(Equation 32)}$$

Finally, we assume that these released thalamic neurons indexed by 1's in the diagonal of $\mathbf{S}_\mu$ then interact with cortex within the linear regime, which allows us to write:

$$[\mathbf{t}_{\mathbf{abs}}]^+ = \mathbf{S}_\mu\,\mathbf{t}_{\mathbf{abs}} \quad \text{(Equation 33)}$$

We can then finally write, from Equations 30, 32, and 33:

$$\dot{\mathbf{c}} = -\mathbf{c} + (\mathbf{J}_{\mathbf{cc}} + \mathbf{J}_{\mathrm{ct}}\mathbf{S}_\mu\mathbf{J}_{\mathrm{tc}})\,\mathbf{c} \quad \text{(Equation 1)}$$

where we defined $\mathbf{c} = \mathbf{c}_{\mathbf{abs}} - \mathbf{a}_\mu$ with $\mathbf{a}_\mu \equiv (\mathbf{I} - \mathbf{J}_{\mathbf{cc}} - \mathbf{J}_{\mathrm{ct}}\,\mathbf{S}_\mu\mathbf{J}_{\mathrm{tc}})^{-1}\,((\mathbf{I} - \mathbf{J}_{\mathbf{cc}})\,\bar{\mathbf{c}} + \mathbf{J}_{\mathrm{ct}}\,\mathbf{S}_\mu\,\bar{\mathbf{t}})$. Note that $\mathbf{c}$ is 'centered' around zero and thus not (mostly) positive, unlike $\mathbf{c}_{\mathbf{abs}}$. Note also that $\mathbf{a}_\mu$ is a motif-specific input that stays constant during the duration of the motif (or alternatively during a phase of motif preparation, see the next section 4.2.6 for more details).

This is the equation for the cortical rates $\mathbf{c}$ that we used throughout the main text. We stress that this means that the target readout patterns that we are displaying in the main text are deviations from a motif-specific baseline readout value $\mathbf{w}^\top\,\mathbf{a}_\mu$. We also emphasize that, for each cortical loop associated with the corticothalamic weight vector $\mathbf{v}$ and the thalamocortical weight vector $\mathbf{u}$, our main text theory only constrains the matrix $\mathbf{u}\mathbf{v}^\top$, leaving some freedom for the norm of the individual weight vectors. When displaying or using individual loop weights, we therefore make the choice to scale these vectors such that the thalamocortical and corticothalamic weight vectors have the same norm (i.e., we multiply $\mathbf{u}$ by $\sqrt{\|\mathbf{u}\|\;\|\mathbf{v}\|/\|\mathbf{u}\|}$ and $\mathbf{v}$ by $\sqrt{\|\mathbf{u}\|\;\|\mathbf{v}\|/\|\mathbf{v}\|}$).

Note that our framework can be generalized in several ways. First, there can be additional, non-plastic thalamic loops that interact with cortex – either for the whole duration of some motif(s) or constantly – and that would simply take part, along with cortex, in an effective 'fixed' recurrent network that the plastic thalamic loop would have to modulate during a motif. Second, even though we only considered here that all the 'plastic' thalamic loops that are not involved in producing the current motif are shut off by basal ganglia, this assumption might be relaxed to some degree. For instance, the same framework could be applied to any scenario where the thalamic neurons who are not involved in shaping the currently produced motifs have dynamics that are dominated by strong approximately constant external inputs that dwarf the feedback from cortex. Indeed, these driven thalamic neurons would then act as an external constant input added to the effective cortical dynamics, whose effect could therefore easily be canceled by simple substraction of an 'effective copy' of this signal on the readout.

4.2.6 Implementing the more realistic dynamics

In this section, we will show that it is possible to find a dynamical regime of the more biologically constrained equations Equation 28 under which our simplified theoretically tractable framework is still valid to understand how thalamocortical loops support motor sequencing. Importantly, the result of this approach gives a lower-bound on the performance of a more biologically plausible network, under the strong constraints that our equations derived for centered linear dynamics still hold such that there is a clear mapping between the dynamics shown in the main text and the more complex dynamics studied here. In all likelihood, better performance would occur if using an online learning rule that would account for the additional complexity of the dynamics (Logiaco and Escola, 2020). We however focus on preserving our theoretical framework to show that it can also give the right intuitions and explanations when the dynamics includes more biological details.

More specifically, we will constrain our dynamics such that for a particular motif, our theory – which assumes an instantaneous thalamus – is still used to set both the effective eigenspectrum (through the tuning of the corticothalamic weights $\mathbf{v}$ in Equation 2) and the initial centered cortical rates $\mathbf{c}_\mu^{\mathrm{init}}$ (as a function of the desired amplitudes α_μ for the eigenmodes; Equation 16):

$$\begin{cases} \mathbf{v} = \mathbf{L}^\top \mathrm{diag}(\mathbf{L}\mathbf{u})^{-1}\mathbf{P}^+\mathbb{1} & \text{(Equation 2)} \\ \mathbf{c}(0) = \mathbf{c}_\mu^{\mathrm{init}} \equiv \tilde{\mathbf{R}}\,\mathrm{diag}\left(\tilde{\mathbf{R}}^\top\mathbf{w}\right)^{-1}\alpha_\mu\,, & \text{(Equation 16)} \end{cases}$$

where $\mathbf{u}$ is the thalamocortical weight vector from the single thalamic unit interacting with cortex during a motif, and $\mathbf{v}$ is the corresponding corticothalamic weight vector. However, we will now describe how other variables of the network are going to be adjusted to account for the reasonable thalamic timescale in Equation 28 (10 times faster than cortex, see the table at the beginning of the STAR Methods section 4.2.5). This means that the timescale in thalamus induces deviations that we are able to treat as a correlated and biased noise introduced in our ideal model (Equation 1). In addition, we will adjust the biases in Equation 28 to ensure that the added rectification of the rates only minimally impact the dynamics.

To achieve this, we will first work with the linearized version of Equation 28, writing:

$$\begin{cases} \dot{\mathbf{c}}_{\mathbf{abs}} = -(\mathbf{c}_{\mathbf{abs}} - \bar{\mathbf{c}}) + \mathbf{J}_{\mathbf{cc}}(\mathbf{c}_{\mathbf{abs}} - \bar{\mathbf{c}}) + \mathbf{J}_{\mathrm{ct}}^{\mathrm{act}}\,\mathbf{t}_{\mathbf{act}} \\ \dot{\mathbf{t}}_{\mathbf{act}} = \dfrac{1}{\tau_t}\left[-(\mathbf{t}_{\mathbf{act}} - \overline{\mathbf{t}_{\mathbf{act}}}) + \mathbf{J}_{\mathrm{tc}}^{\mathrm{act}}\,\mathbf{c}_{\mathbf{abs}}\right] \end{cases} \quad \text{(Equation 34)}$$

where we remind the reader that the index 'act' indicates the subset of the thalamic units that are interacting with cortex at some time during a sequence (instead of being shut off by basal ganglia). Note that, in line with Sec. 4.2.5, the vectors $\bar{\mathbf{c}}$ and $\overline{\mathbf{t}_{\mathbf{act}}}$ actually include a motif-specific external input.

We can then express the dynamics using a single vector $\mathbf{z}_{\mathbf{abs}}$ to concatenate both the cortical and non-silent thalamic rates: for $i \leq N$, $[\mathbf{z}_{\mathbf{abs}}]_i = [\mathbf{c}_{\mathbf{abs}}]_i$, and for $i > N$, $[\mathbf{z}_{\mathbf{abs}}]_i = [\mathbf{t}_{\mathbf{act}}]_i$. Similarly, we define a concatenated vector of biases $\bar{\mathbf{z}}$ regrouping $\bar{\mathbf{c}}$ and $\overline{\mathbf{t}_{\mathbf{act}}}$. Hence, we can write:

$$\dot{\mathbf{z}_{\mathbf{abs}}} = \mathbf{M}_{\mathbf{eff}}\, \mathbf{z}_{\mathbf{abs}} + \mathbf{V}_{\mathbf{eff}}\, \bar{\mathbf{z}}\,, \qquad \text{(Equation 35)}$$

where we defined the matrices $\mathbf{M}_{\mathbf{eff}}$ and $\mathbf{V}_{\mathbf{eff}}$ such that:

- for $i \leq N$ and $j \leq N$, $[\mathbf{M}_{\mathbf{eff}}]_{ij} = [\mathbf{J}_{\mathrm{cc}}]_{ij} - \delta_{ij}$ and $[\mathbf{V}_{\mathbf{eff}}]_{ij} = \delta_{ij} - [\mathbf{J}_{\mathrm{cc}}]_{ij}$
- for $i \leq N$ and $j > N$, $[\mathbf{M}_{\mathbf{eff}}]_{ij} = [\mathbf{J}^{\mathrm{act}}_{\mathrm{ct}}]_{i,\,(j-N)}$ and $[\mathbf{V}_{\mathbf{eff}}]_{ij} = 0$
- for $i > N$ and $j \leq N$, $[\mathbf{M}_{\mathbf{eff}}]_{ij} = \frac{1}{\tau_t}\,[\mathbf{J}^{\mathrm{act}}_{\mathrm{tc}}]_{(i-N),\,j}$ and $[\mathbf{V}_{\mathbf{eff}}]_{ij} = 0$
- for $i > N$ and $j > N$, $[\mathbf{M}_{\mathbf{eff}}]_{ij} = -\frac{\delta_{ij}}{\tau_t}$ and $[\mathbf{V}_{\mathbf{eff}}]_{ij} = \frac{\delta_{ij}}{\tau_t}$.

Finally, we get:

$$\dot{\mathbf{z}} = \mathbf{M}_{\mathbf{eff}}\, \mathbf{z} \qquad \text{(Equation 36)}$$

where $\mathbf{z} = \mathbf{z}_{\mathbf{abs}} + \mathbf{b}$ and $\mathbf{b} \equiv \mathbf{M}_{\mathbf{eff}}^{-1}\, \mathbf{V}_{\mathbf{eff}}\, \bar{\mathbf{z}}$ is a bias that is fixed during a particular motif or during the preparatory period. We will refer to $\mathbf{z}$ as the 'centered rates' (representing deviations of the rates above or below some average value). Finally, we define effective readout weights $\mathbf{w}_{\mathrm{eff}}$ for this effective $\mathbf{z}$ dynamics, such that $\forall i \leq N$, $\mathbf{w}_{\mathrm{eff},i} = \mathbf{w}_i$ and else $\mathbf{w}_{\mathrm{eff},i} = 0$. We will now examine how to derive appropriate thalamocortical weight vectors and constant biases, first tackling the question of the motif dynamics and then addressing the case of the preparatory period.

4.2.6.1 Motif dynamics

We will start by defining, for a given motif, a full vector of ideal initial conditions for the centered rates $\mathbf{z}_\mu$ of the network as a natural extension of the ideal initial centered rates derived in the instantaneous thalamus framework: for $i \leq N$, $[\mathbf{z}_\mu]_i = [\mathbf{c}_\mu^{\mathrm{init}}]_i$, and $[\mathbf{z}_\mu]_{N+1} = \mathbf{v}^\top\, \mathbf{c}_\mu^{\mathrm{init}}$. (Note that the exact value of the initial thalamic rates has little impact on the circuit's dynamics as thalamic units have a faster timescale ensuring that they quickly follow their synaptic inputs.) Then, we can express the dynamics of $\mathbf{z}$ as a function of the eigenvectors and the eigenvalues of $\mathbf{M}_{\mathbf{eff}}$ such that $\mathbf{z}(t) = \mathbf{R}_{\mathbf{eff}}\, \mathrm{diag}(e^{\lambda_{\mathbf{eff}}\, t})\, \mathbf{L}_{\mathbf{eff}}\, \mathbf{z}_\mu$.

Tuning $\mathbf{u}$ *to minimize the thalamic timescale effects while enforcing instantaneous-loop eigenvalue control with* $\mathbf{v}$

Now that we expressed centered $\mathbf{z}$ dynamics, we can write an equation to ensure that, during motif production, the readout is robust to both noise in the initial conditions and network response deviations that accumulate over the course of motif production due to the fact that the thalamus is not instantaneous. More precisely, while the idealized eigenspectrum is constrained by setting the thalamocortical weight vectors $\mathbf{v}$ through Equation 2, the corticothalamic weight vector $\mathbf{u}$ is adjusted to minimize the following cost function C_{m}:

$$C_{\mathrm{m}} = E_{\delta \mathbf{z}_0}\left[\int_{t=0}^{t=t_\mu} \left|\left|\left(\mathbf{w}_{\mathrm{eff}}^\top\, \mathbf{R}_{\mathbf{eff}}\, \mathrm{diag}\left(e^{\lambda_{\mathbf{eff}}\, t}\right)\, \mathbf{L}_{\mathbf{eff}}\, (\mathbf{z}_\mu + \delta \mathbf{z}_0)\right) - \alpha_\mu^\top\, e^{(\tilde{\lambda}-1)t}\right|\right|^2 dt\right], \qquad \text{(Equation 37)}$$

where t_μ is the motif duration, $\alpha_\mu^\top\, e^{(\tilde{\lambda}-1)t} = \hat{y}(t)$ is the weighted eigenmode sum which would form the readout in the idealized instantaneous thalamic network governed by the matrix $\mathbf{J}_{\mathrm{cc}} + \mathbf{u}\mathbf{v}^\top$, and $\delta \mathbf{z}_0$ is a zero-centered and uncorrelated noise in the initial conditions with a standard deviation $\sigma_{\delta \mathbf{z}_0}$ scaling as 5% of the square root of the mean squared activity norm during the motif (similarly to Equation 23).

After a few algebraic developments, we find:

$$\begin{cases} C_{\mathrm{m}} = \dfrac{\mathbf{w}_{\mathrm{eff}}^\top}{t_\mu}\mathbf{R}_{\mathbf{eff}}\left[\left(\left(\mathbf{L}_{\mathbf{eff}}\left(\mathrm{diag}\left(\sigma^2_{\delta \mathbf{z}_0}\mathbb{1}\right) + \mathbf{z}_\mu\, \mathbf{z}_\mu^\top\right)\mathbf{L}_{\mathbf{eff}}^\top\right) \odot \Lambda_{\mathrm{eff}}\right)\mathbf{R}_{\mathbf{eff}}^\top\right. \\ \left. -2\left(\left(\mathbf{L}_{\mathbf{eff}}\, \mathbf{z}_\mu\, \left(\mathbf{c}_\mu^{\mathrm{E}}\right)^\top \left(\tilde{\mathbf{L}}^{\mathrm{E}}\right)^\top\right) \odot \mathbf{B}\right)\left(\tilde{\mathbf{R}}^{\mathrm{E}}\right)^\top\right]\mathbf{w}_{\mathrm{eff}} + \dfrac{\mathbf{w}^\top}{t_\mu}\tilde{\mathbf{R}}\left[\left(\tilde{\mathbf{L}}\, \mathbf{c}_\mu\, \mathbf{c}_\mu^\top\, \tilde{\mathbf{L}}^\top\right) \odot \Lambda\right]\tilde{\mathbf{R}}^\top\mathbf{w} \\ \sigma^2_{\delta \mathbf{z}_0} = 0.05^2\, \dfrac{1}{t_\mu\, N_z}\mathbf{z}_\mu^\top\, \mathbf{L}_{\mathbf{eff}}^\top\left[\left(\mathbf{R}_{\mathbf{eff}}^\top\, \mathbf{R}_{\mathbf{eff}} \odot \Lambda_{\mathrm{eff}}\right)\right]\mathbf{L}_{\mathbf{eff}}\, \mathbf{z}_\mu \end{cases}$$

where the twiddled letters relate to the eigendecomposition of $\mathbf{J}_{\mathrm{cc}} + \mathbf{u}\mathbf{v}^\top$, and the upper E index indicates an extension of arrays with a last additional zero row (initial rates vector) or last additional zero row and column (left and right eigenvector matrix). Also, we introduced $[\Lambda_{\mathrm{eff}}]_{ij} = \frac{e^{([\lambda_{\mathrm{eff}}]_i + [\lambda_{\mathrm{eff}}]_j)\, t_\mu} - 1}{[\lambda_{\mathrm{eff}}]_i + [\lambda_{\mathrm{eff}}]_j}$; $[\Lambda]_{ij} = \frac{e^{([\tilde{\lambda}]_i + [\tilde{\lambda}]_j - 2)\, t_\mu} - 1}{[\tilde{\lambda}]_i + [\tilde{\lambda}]_j - 2}$ (Equation 23) where $\tilde{\lambda}$ is the vector containing the eigenvalues of $\mathbf{J}_{\mathrm{cc}} + \mathbf{u}\mathbf{v}^\top$

(note that $\tilde{\lambda}$ is independent of **u**); and $[\mathbf{B}]_{ij} = \frac{e^{([\lambda_{\text{eff}}]_i + [\tilde{\lambda}^{\text{E}}]_j)\, t_\mu} - 1}{[\lambda_{\text{eff}}]_i + [\tilde{\lambda}^{\text{E}}]_j}$ where for $j \leq N$, $\left[\tilde{\lambda}^{\text{E}}\right]_j = [\tilde{\lambda}]_j - 1$ and $[\tilde{\lambda}^{\text{E}}]_{N+1} = 0$. Finally, we also defined N_z as the number of elements in **z**, i.e., here $N+1$.

Hence, by optimizing the vector **u** to minimize the cost C_{m} while fixing the vectors **v** and $\mathbf{c}_\mu^{\text{init}}$ according to Equations 2 and 16, we ensure that the effective network governing the motif dynamics is designed to ensure eigenvalue control in the limit of an infinitely fast thalamus while correcting for the deviations due to the finite thalamic timescale.

Designing biases added to the rate equations to impose positive rates during motif production

Now that we have found thalamocortical weight vectors **u** and **v** which still perform a similar effective cortical eigenvalue control as in the main text while accounting for the thalamus' internal dynamics, we can consider the question of the biases in the rate dynamics that impose positive rates $\mathbf{z}_{\mathbf{abs}} = \mathbf{z} - \mathbf{M}_{\mathbf{eff}}^{-1}\, \mathbf{V}_{\mathbf{eff}}\, \bar{\mathbf{z}}$ in the circuit. Then, for a vector ε containing the desired positive lower bound of the rates, we can write:

$$\forall t, \mathbf{z}_{\mathbf{abs}}(t) \geq \varepsilon \Leftrightarrow -\mathbf{M}_{\mathbf{eff}}^{-1}\, \mathbf{V}_{\mathbf{eff}}\, \bar{\mathbf{z}} - \varepsilon \geq \max_t\left(-\mathbf{R}_{\mathbf{eff}}\, \text{diag}\left(e^{\lambda_{\text{eff}}\, t}\right)\, \mathbf{L}_{\mathbf{eff}}\, \mathbf{z}_\mu\right) = \mathbf{b}_\mu\,,$$

where the maximum is taken element-wise. We therefore set

$$-\mathbf{M}_{\mathbf{eff}}^{-1}\, \mathbf{V}_{\mathbf{eff}}\, \bar{\mathbf{z}} = (\mathbf{b}_\mu + \varepsilon) \Leftrightarrow$$

$$\bar{\mathbf{z}} = -\mathbf{V}_{\mathbf{eff}}^{-1}\, \mathbf{M}_{\mathbf{eff}}\, (\mathbf{b}_\mu + \varepsilon)\,, \qquad \text{(Equation 38)}$$

where we chose $\varepsilon = 0.1$ in our simulations.

This ensures that the non-centered rate variables $\mathbf{z}_{\mathbf{abs}}$ stay positive while undergoing the dynamics in Equation 35 if, at the beginning of motif μ, these rates are exactly set to $\mathbf{z}_{\mathbf{abs}\,\mu} = \mathbf{z}_\mu - \mathbf{M}_{\mathbf{eff}}^{-1}\, \mathbf{V}_{\mathbf{eff}}\, \bar{\mathbf{z}}$. However, a deviation from the right initial conditions can cause the linear rate trajectories to go beneath zero. Having a good preparatory period will therefore prove critical to ensuring that the linear dynamics solution during a motif stays close to the solution of a rectified linear dynamical system. Fortunately, as we will now show, we can design an efficient thalamic preparatory network obeying the constraints of positive rates and non-instantaneous thalamus.

4.2.6.2 Preparatory period dynamics

We again first consider the centered rates **z** with linear dynamics (Equation 36) to find thalamocortical connections accommodating a non-instantaneous thalamus and then find the biases needed to maintain positivity of the non-centered rates.

Optimizing thalamocortical connectivity for motif preparation with a non-instantaneous thalamus

We optimize the loop weights $\mathbf{J}_{\text{tc}}^{\text{act}}$ and $\mathbf{J}_{\text{ct}}^{\text{act}}$ – where we remind the reader that the index 'act' indicates that these matrices are restricted to the subset of thalamic neurons engaged in motor preparation – such that $\mathbf{z}(t)$ from Equation 36 converges fast to its steady-state (which is zero in the absence of an input). In addition, in keeping with the cost function in the case of an instantaneous thalamus (Equation 27), we add a term to favor smoothness of the readout, to get the cost function:

$$C_{\text{p}} = \frac{1}{N + N_{\text{thal}}} E_{\mathbf{z}_0}\left[\int_0^\infty \mathbf{z}(t)^2\, dt\right] + \beta\, E_{\mathbf{z}_0}\left[\int_0^\infty dt \left(\frac{d}{dt}\mathbf{w}_{\text{eff}}^\top \delta\mathbf{z}\right)^2\right]$$

$$\propto \frac{1}{N + N_{\text{thal}}} \text{Tr}\left(\mathbf{R}_{\mathbf{eff}}\left(\left(\mathbf{L}_{\mathbf{eff}}\, \mathbf{L}_{\mathbf{eff}}^\top\right) \odot \mathbf{A}\right)\mathbf{R}_{\mathbf{eff}}^\top\right) + \beta\, \mathbf{w}_{\text{eff}}^\top \mathbf{R}_{\mathbf{eff}}\left(\left(\mathbf{L}_{\mathbf{eff}}\, \mathbf{L}_{\mathbf{eff}}^\top\right) \odot \mathbf{G}\right)\mathbf{R}_{\mathbf{eff}}^\top \mathbf{w}_{\text{eff}}$$

where $\mathbf{A}_{ij} = \frac{-1}{[\lambda_{\text{eff}}]_i + [\lambda_{\text{eff}}]_j}$, $G_{ij} = [\lambda_{\text{eff}}]_i [\lambda_{\text{eff}}]_j A_{ij}$, N_{thal} is the number of thalamic units in the preparatory network, and $\mathbf{z}_0$ are the random initial conditions which are assumed to be centered and i.i.d. Note that $\mathbf{M}_{\mathbf{eff}} = \mathbf{R}_{\mathbf{eff}}\, \text{diag}(\lambda_{\text{eff}})\, \mathbf{L}_{\mathbf{eff}}$ now relates to the preparatory circuit.

Biases able to enforce rate positivity in the effective preparatory network despite initial rates variability

We then consider which biases $\bar{\mathbf{z}}_{\mathbf{prep}}$ to add to the preparatory dynamics (Equation 35) to get positive rates $\mathbf{z}_{\mathbf{abs}} = \mathbf{z} - \mathbf{M}_{\mathbf{eff}}^{-1}\, \mathbf{V}_{\mathbf{eff}}\, \bar{\mathbf{z}}_{\mathbf{prep}}$.

We should note that at the beginning of the preparatory period, the cortical and thalamic rates will be non-negative if the just-completed motif was started at its correct initial condition and driven by the input defined in Equation 38. Similarly, we can posit that the cortical rates at the start of the upcoming motif will be positive as they too will be defined by Equation 38. Additionally, to guarantee that, during the preparatory period, the cortical rates converge to the necessary initial configuration for the upcoming motif, we will need to set their inputs to explicit values. This therefore leaves the inputs to the thalamic units as our only degrees of freedom for ensuring rate positivity during the preparatory period. Here we describe a process for determining those thalamic biases in a way that is agnostic to either the rates at the end of the prior motif or at the beginning of the upcoming motif. These thalamic biases can then be used for all preparatory periods independent of the upcoming motif. We will proceed in two steps.

1. To find biases for the thalamic units, first we proceed in this step by finding biases for both the cortical and thalamic units that would ensure positivity *on average.* In step 2, we will then save the thalamic biases we find and replace the cortical biases with the specific values needed for the upcoming motif.
Hence, we here treat the non-centered rates at the beginning of the preparatory period $\mathbf{z}^{\mathbf{prep}}_{\mathbf{abs}\ 0}$ as an i.i.d. random vector with mean $\boldsymbol{\mu}^{\mathbf{prep}}_{\mathbf{z}_{\mathbf{abs}\ 0}}$ and standard deviation $\boldsymbol{\sigma}^{\mathbf{prep}}_{\mathbf{z}_{\mathbf{abs}\ 0}}$. For the cortical units, we can simply estimate these statistics from the mean and standard deviation of $\mathbf{z_{abs}}$ over all units and all motifs at the end of six example motifs (the sinc function as in Figure 2 and slow oscillating motifs as in Figure 4). In addition, we assume that the rates of the thalamic units are close to zero at the beginning of the preparatory period, which would happen if, for example, at the end of the previous motif, the basal ganglia signals just compensate for the cortical inputs. (Note again that the exact value of the initial thalamic rates is not critical for the circuit's dynamics as thalamic units have a faster timescale ensuring that they quickly follow their synaptic inputs.) We then design a cost function C_p to find bias terms $\mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}} = -\ \mathbf{M}^{-1}_{\mathbf{eff}}\ \mathbf{V}_{\mathbf{eff}}\ \bar{\mathbf{z}}^{\mathbf{prep}}$ such that the random variable:

$$\mathbf{z}^{\mathbf{prep}}_{\mathbf{abs}}(t) = \mathbf{R}_{\mathbf{eff}}\ \mathrm{diag}\left(e^{\lambda_{\mathrm{eff}}\ t}\right)\ \mathbf{L}_{\mathbf{eff}}\ \mathbf{z}^{\mathbf{prep}}_{0} + \mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}}$$

is likely to be positive at all times. The stochasticity of $\mathbf{z}^{\mathbf{prep}}_{\mathbf{abs}}(t)$ emerges because $\mathbf{z}^{\mathbf{prep}}_{0} = \mathbf{z}^{\mathbf{prep}}_{\mathbf{abs}\ 0} - \mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}}$ is a random vector with mean $\mu^{\mathbf{prep}}_{\mathbf{z}_0} = \mu^{prep}_{z_{abs0}} - \mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}}$ and standard deviation $\boldsymbol{\sigma}^{\mathbf{prep}}_{\mathbf{z}_0} = \boldsymbol{\sigma}^{\mathbf{prep}}_{\mathbf{z}_{\mathbf{abs}\ 0}}$. As a step to define the cost function C_p, we can express the statistics of $\mathbf{z}^{\mathbf{prep}}_{\mathbf{abs}}(t)$ at all times to find the vector of trajectories $\mathbf{h}_{\mathbf{low}}(t)$ which we define as the values taken at three standard deviations below the mean for each unit:

$$\begin{cases} E_{\mathbf{z}^{\mathbf{prep}}_0}\left[\mathbf{z}^{\mathbf{prep}}(t)\right] = \mathbf{R}_{\mathbf{eff}}\ \mathrm{diag}\left(e^{\lambda_{\mathrm{eff}}\ t}\right)\ \mathbf{L}_{\mathbf{eff}}\ \left(\mu^{\mathbf{prep}}_{\mathbf{z}_{\mathbf{abs}0}} - \mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}}\right) = \mathbf{f}(t) \\ E_{\mathbf{z}^{\mathbf{prep}}_0}\left[\mathbf{z}^{\mathbf{prep}}(t) - E_{\mathbf{z}^{\mathbf{prep}}_0}\left[\mathbf{z}^{\mathbf{prep}}(t)\right]\right]^2 = \\ \quad \mathbf{R}_{\mathbf{eff}}\ \mathrm{diag}(e^{\lambda_{\mathrm{eff}}\ t})\ \mathbf{L}_{\mathbf{eff}}\ \mathrm{diag}\left(\left(\boldsymbol{\sigma}^{\mathbf{prep}}_{\mathbf{z}_{\mathbf{abs}\ 0}}\right)^2\right)\ \mathbf{L}^{\top}_{\mathbf{eff}}\ \mathrm{diag}(e^{\lambda_{\mathrm{eff}}\ t})\ \mathbf{R}^{\top}_{\mathbf{eff}} = \mathbf{g}(t) \\ \mathbf{h}_{\mathbf{low}}(t) = \mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}} + \mathbf{f}(t) - 3 * \sqrt{\mathbf{g}(t)} \end{cases}$$

To ensure that, under linear dynamics during the preparatory period, the rates stay as positive as possible, we therefore optimize $\mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}}$ to minimize the sum of the square of the negative values of the vector of element-wise minima $\min_t(\mathbf{h}_{\mathbf{low}}(t)) = \mathbf{h}^{\mathbf{min}}_{\mathbf{low}}$:

$$C_\mathrm{p} = \sum_{i\ \ s.\ t.\ \ \left[\mathbf{h}^{\mathbf{min}}_{\mathbf{low}}\right]_i < 0} \left(\left[\mathbf{h}^{\mathbf{min}}_{\mathbf{low}}\right]_i\right)^2$$

where the dynamics of $\mathbf{h}_{\mathbf{low}}(t)$ were computed over a duration of five cortical timescales.

2. In this second step, we will now use the thalamic biases we just found (the elements $[\mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}}]_i$ for $i > N$) and combine them with the cortical biases needed to ensure that the preparatory period converges to the correct rates for the upcoming motif.
We require that the cortical rates converge toward the following positive pattern to start upcoming motif μ: $[\mathbf{z}_\mu + \mathbf{b}_\mu + \varepsilon]_{i \le N}$ (as defined in Equation 38 and above). Similarly to the methods used in the main text, it is easy to see that this relies on designing an additive input $\bar{\mathbf{z}}^{\mathbf{prep}}_{\mu}$ specific to the dynamics leading to a particular motif μ:(Equation 39)

$$\dot{\mathbf{z}}^{\mathbf{prep}}_{\mathbf{abs}} = \mathbf{M}_{\mathbf{eff}}\ \mathbf{z}^{\mathbf{prep}}_{\mathbf{abs}} + \mathbf{V}_{\mathbf{eff}}\ \bar{\mathbf{z}}^{\mathbf{prep}}_{\mu} \Leftrightarrow \lim_{t \to \infty} \mathbf{z}^{\mathbf{prep}}_{\mathbf{abs}}(t) = -\ \mathbf{M}^{-1}_{\mathbf{eff}}\ \mathbf{V}_{\mathbf{eff}}\ \bar{\mathbf{z}}^{\mathbf{prep}}_{\mu} = \mathbf{z}_{\mathbf{abs}\ \mu} \qquad \text{(Equation 39)}$$

We then match the cortical steady-state rates expressed in Equation 36 to the next motif's ideal initial pattern: we set, for $i \le N$, $[\mathbf{z}_{\mathbf{abs}\ \mu}]_i = [\mathbf{c}_\mu]_i + [\mathbf{b}_\mu + \varepsilon]_i$. Adding the thalamic biases that favor positive dynamics across motif transitions, we therefore set, for $N_{\mathrm{prep}} \ge i > N$, $[\mathbf{z}_{\mathbf{abs}\ \mu}]_i = [\mathbf{b}^{\mathbf{prep}}_{\mathbf{eff}}]_i$, where N_{prep} is the number of units in the preparatory network.
Hence, for each transition toward a particular motif μ, we determined a vector $\bar{\mathbf{z}}^{\mathbf{prep}}_{\mu} = -\ \mathbf{V}^{-1}_{\mathbf{eff}}\ \mathbf{M}_{\mathbf{eff}}\ \mathbf{z}_{\mathbf{abs}\ \mu}$ such that we can write an efficient and mostly positive preparatory dynamics according to Equation 39.

4.2.6.3 Numerical implementation

The methodology developed above can be used to show that two major simplifications made in Equation 1: infinitely fast thalamus and linear centered dynamics, can be relaxed while preserving the insights from the theoretical analysis that we present in the main text.

More specifically, we can produce sequential dynamics using a reasonable dynamic timescale and positive rectified rates according to Equation 28, by applying the non-linearity on top of the mostly positive linear rate dynamics described above in Equation 34 that is designed to approximately match the readout of a network with an infinitely fast thalamus. Note that we did not explicitly model the details of the basal ganglia input to thalamus: instead, we simply assumed that thalamic units can be turned off very fast by their basal ganglia inputs, while when these units start interacting with cortex their initial rates are assumed to be close to zero.

In addition, when simulating Equations 28 and 34, we have to think about the effect of the biases on the readout. For any motif, adding the biases $\bar{\mathbf{c}}$ and $\overline{\mathbf{t}_{\mathbf{act}}}$ – regrouped into the vector $\bar{\mathbf{z}}$ – to the differential equations transforms the centered dynamics $\mathbf{z}$ into offset

dynamics. In order to recover the desired motifs, we simply remove the offset from the cortical readout at each time (Figures S7A and S7B). By inducing a minimal modification of our theoretical framework, this strategy is in line with our aim to show that our simplified centered dynamics – which we designed for the production of desired motifs – can be well-approximated in a particular regime of the more biologically plausible dynamics. Other strategies involving more modifications of the dynamics would however be possible to recover the desired network output, as we will discuss at the end of this section.

In Figure S7B, top, we show that adding the above-mentioned constraints: linear offset simulations with a plausible thalamic timescale (just 10 times faster than the cortical timescale in Equation 34), induces very small deviations (gray curves) compared to the case of instantaneous thalamus (black curves) when starting motifs from their ideal initial rate patterns (Equation 16). Also, the finite response timescale of thalamus does not prevent fast convergence to appropriate cortical initial conditions during the motifs' preparatory period using a small preparatory network (10% of the cortical network's size with plausible synaptic weight magnitudes set as for the green line of Figure S6E; Figure S7C, green curve).

We now turn to examining the effect of the nonlinear rectification of the rates (Equation 28). We show in the top panel of Figure S7C that introducing a rectification in the positively biased linear rate dynamics of Equation 34 leads to minimal deviations during motifs (cyan curve versus green line). Finally, in of Figures S7C and S7D, we show how introducing the rectifying non-linearity in the dynamics usually speeds the preparatory process.

These results successfully show that more plausible neural dynamics with rectified rates and a finite timescale in thalamus have a dynamical regime that is successfully described by our simplified theoretical framework.

Note that the need for the readout offset correction may be avoided if, during motif production, the eigenvalue control would include – in addition to eigenmodes for fitting the desired 'centered' motifs – an additional eigenvalue close to zero real and imaginary part (after accounting for the leak – i.e., this eigenvalue would be on the same vertical line as the desired eigenvalues of Figure 1E). This would permit the intrinsic production of a constant value added to the readout (controlled by the initial rates of the motif, which would need to be optimized along with the biases as defined in Equation 38). This would also make the motif transitions smoother, as the change of the readout offset when preparing a new motif introduces a discontinuity in the readout. In addition, the rectifying non-linearity does introduce some additional sways during transition times (compare the green and cyan lines in the top panel of Figure S7B). These limitations are likely to resolve – or at least to become much less visible – if adjusting the synaptic weights with an online learning rule that would account for the constraints of the more plausible dynamics (indeed smooth transitions have been shown in a nonlinear network constrained to have similar dynamics as described here but trained with gradient descent (Logiaco and Escola, 2020)). The design of a biologically plausible learning rule for this circuit is however beyond the scope of this article.

4.2.7 Numerics

The simulations were performed using MATLAB and Python. We numerically optimized Equation 22 and Equation 27 (we have experimented with either the square roots of these cost functions or the cost functions themselves, both versions give good results). As discussed in STAR Methods Sec. 4.2.3, these equations can have multiple local minima. Thus, starting from different random initializations for the optimization, we selected the best among the solutions we found to present in our figures.

These optimization procedures are relatively efficient:

- The cost function $\boldsymbol{C}(\mathbf{u})$ given by Equation 22 requires the eigenvalues $\tilde{\lambda}$ and eigenvectors $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{L}}$ of $\tilde{\mathbf{J}}_\mu$. As we will clarify below, these can be computed efficiently avoiding explicit eigendecomposition on each iteration of optimization.

  First, because $\mathbf{d} = \mathbf{P}^+ \mathbb{1}$, $\mathbf{d}$ is fully determined by $\mathbf{P}$ which depends on the eigenvalues of $\mathbf{J}_{cc}$ and the target eigenvalues, but not on $\mathbf{u}$. As we explained in STAR Methods Sec. 4.2.2 (Equation 13), this implies that all N eigenvalues of $\tilde{\mathbf{J}}_\mu$ are also fully determined and independent of $\mathbf{u}$. Thus $\boldsymbol{\Lambda}$ can be computed once prior to optimization of Equation 22.

  Second, we are able to explicitly express the eigenvectors of the dynamics (Equation 19). These expressions depend on $\mathbf{R}$, $\mathbf{L}$, and $\mathbf{A}$, all of which can be computed prior to optimization, and on $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ which can be computed from $\mathbf{u}$ (and $\mathbf{L}$ and $\mathbf{d}$) on each iteration of optimization. Implementing the formulae given in Equation 18 requires one matrix multiplication each per iteration (in addition to other much cheaper operations). With $\tilde{\mathbf{R}}$, $\tilde{\mathbf{L}}$, and $\boldsymbol{\Lambda}$ in hand, the remaining operations to compute C are dominated by two additional matrix multiplications (in Equations 22 and 23). Without the need for more computationally intensive operations (e.g., eigendecomposition or matrix inversion), the cost function C is efficient to optimize.
- For Equation 27, though we do not have explicit expressions for the eigenvalues or the eigenvectors, it is then possible to compute the gradient of these expressions through the loop weight vectors analytically (Boeddeker et al., 2017). This considerably speeds up the optimization.

When matching the target trajectory y_μ for one motif μ to its approximation with a small number of eigenmodes $\hat{y}_\mu = \sum_{i=1}^{K} \alpha_i^{\text{des}} e^{(\lambda_i^{\text{des}}-1)t}$, we tried a few different numbers of non-zero eigenmodes K (typically between four and twenty, see Figures S1A–S1E). For each value of K, we used MATLAB's fmincon function with 50 different random starts to optimize the real and imaginary part of each amplitude α_i and each eigenvalue λ_i^{des} to minimize the mean square difference between y_μ and $\hat{y}_\mu$. Constraints were added to discourage very large amplitudes α_i^{des} (here, we impose $\|\boldsymbol{\alpha}^{\text{des}\,2}\| < 18$ because this bound approximately corresponds to

the square of the maximum desired magnitude of the readout), very large real or imaginary parts of the eigenvalues λ_i^{des} (here, by imposing that the maximal distance between these eigenvalues would be 2) or to find very similar eigenvalues which would tend to lead to high eigenvector correlations (Sec. 4.2.3; here we imposed a conservative minimum distance of about 0.05 between eigenvalues).

Unless otherwise stated, we used cortical matrices $\mathbf{J_{cc}}$ of size $N = 500$, and we made sure that $\mathbf{J_{cc}} - \mathbf{I}$ was stable. More specifically, we discarded the matrix if $\mathbf{J_{cc}} - \mathbf{I}$ had positive eigenvalues, and we added a small positive ε to the leak during numerical simulations of the dynamics to ensure that no instability would arise due to numerical approximations during eigenvalue control with target eigenvalues very close to, or on, the line $\mathrm{Re}\lambda = 1$.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using MATLAB. Sample sizes and p values are indicated.

Cell Reports, Volume *35*

Supplemental information

Thalamic control of cortical dynamics in a model of flexible motor sequencing

Laureline Logiaco, L.F. Abbott, and Sean Escola

Supplementary Figures

Figure S1: Tuned basis functions vs. linear reservoir for generating motifs

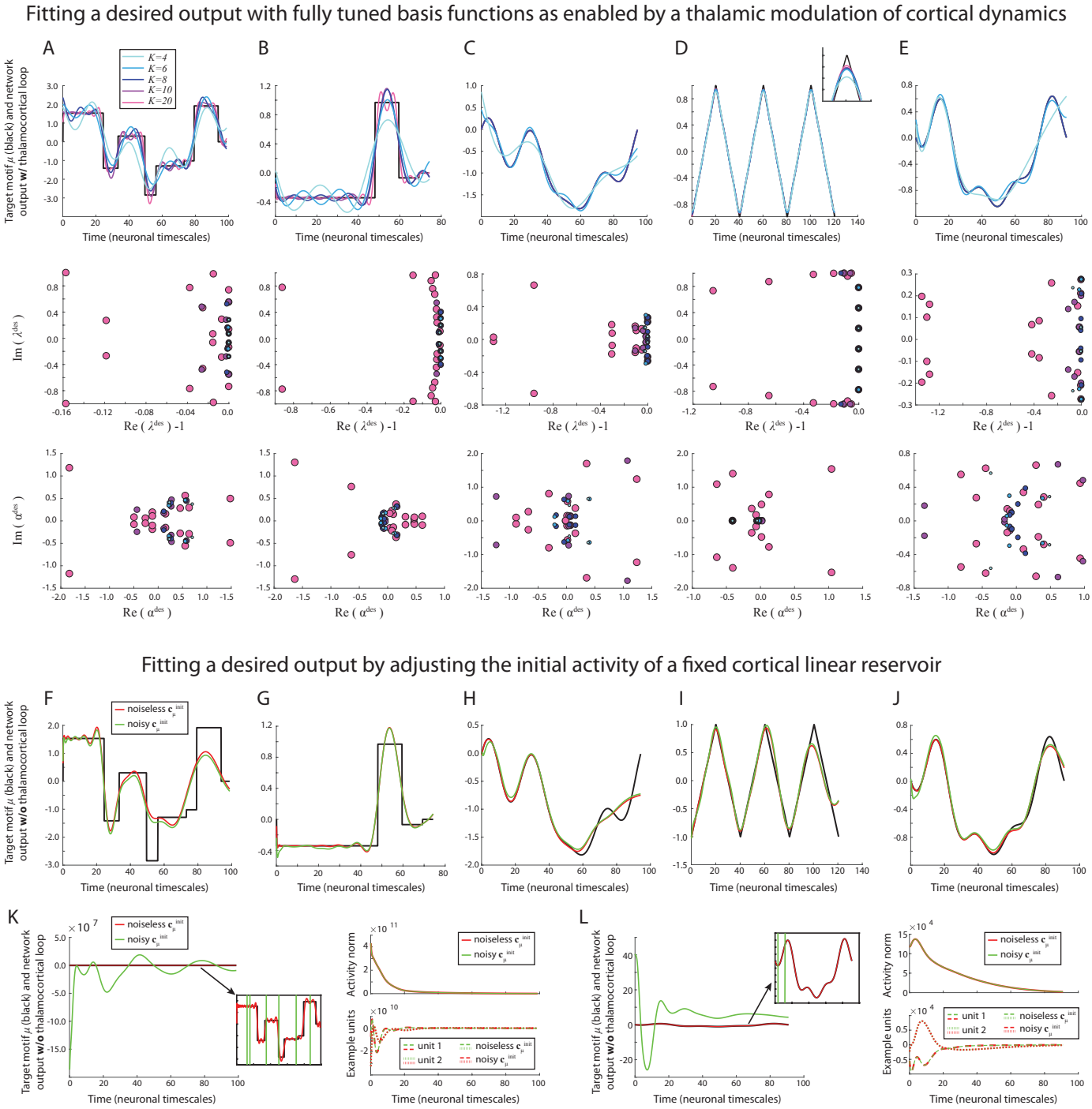


Figure S.1: See next page.

Figure S.1: tuned basis functions vs. linear reservoir for generating motifs, related to Figure 1. **A–E)** Approximating each target motif by the sum of a small number of fully-tuned basis functions of the network dynamics: the eigenmodes (see Figures 1 C-D for a visualization of these basis functions). Here, the exponential and oscillation timescales as well as the initial amplitudes and phases of these eigenmodes are adjusted to each desired output. Such a sum of fully-tuned eigenmodes corresponds to the network output that is generated when cortex interacts with a thalamocortical loop whose weights are tuned for the production of a particular motif (i.e., as per Eq. 2 and Eq. 22). **Top:** Example target motifs y_μ (black) and their approximations $\hat{y}_\mu$ which are sums of eigenmodes (i.e., sums of conjugate pairs of complex exponentials: $\hat{y}_\mu(t) = \sum_{i=1}^{K} \alpha_i^{\text{des}} e^{\left(\lambda_i^{\text{des}}-1\right)t}$, with $K = 4$, $K = 6$, $K = 8$, $K = 10$ and $K = 20$ corresponding to light blue to pink curves). Here, the parameters α_i^{des} and λ_i^{des} are numerically optimized to minimize the mean square error between the y_μ and $\hat{y}_\mu$ (see Methods section 4.2.7 for details). Note that in Figures 1 and 2 these parameters could instead be determined analytically by sampling the Fourier transform of the mathematically well-defined sinc function (which typically requires many more eigenmodes than the numerical procedure presented here because the analytical solution is a valid approximation over all time, while the numerical solution is identified only for the desired motif duration). Note also that, in **C** and **E**, the curves for the larger values of K – as well as the target motif – are hidden under the $K = 8$ curve. In **D**, the inset shows a magnified part of the first upper triangle. Finally, note that in **E**, we show motif 'C' from Figure 4 – it is fit without any discernible error for $K \geq 8$. **Middle:** Real and imaginary parts of the parameters $\lambda_1^{\text{des}} - 1, \ldots, \lambda_K^{\text{des}} - 1$ used to match $\hat{y}_\mu$ to the corresponding target motif y_μ, for different values of K. Note that increasingly large symbols are used for larger values of K, in order to make all symbols visible in cases when the λ_i^{des} overlap for different values of K. **Bottom:** Real and imaginary part of the α_i^{des} (conventions as in the middle panels). Note that even if these values were constrained to have a relatively small norm during the optimization (here, $\left\|\boldsymbol{\alpha}^{\text{des}}\right\|^2 < 18$), the solution that is found to minimize the error between desired output and weighted sum of tuned eigenmodes is characterized by magnitudes of the α_i^{des} that are most often 'naturally' smaller than the bound, especially for small values of K and if the maximum target output is not very large. This illustrates how the tuning of the timescales of a few eigenmodes can lead to a powerful and robust approximation of a wide variety target outputs. **F–J)** Approximating the same target motifs when only tuning the initial activity pattern $\mathbf{c}_\mu^{\text{init}}$ of the isolated cortical network (whose eigenspectrum is shown in Figure 1E, and which is a random reservoir network). This amounts to tuning the initial amplitude and phase of each eigenmode but not their timescales. If no limitations are set on the norm of the initial activities, the least squared error solution between the network output and the target motifs settles on extremely large norms for $\mathbf{c}_\mu^{\text{init}}$ which are unstable with respect to noise in $\mathbf{c}_\mu^{\text{init}}$ (shown in panels **K** and **L** below). Here we show solutions for which the elements of $\mathbf{c}_\mu^{\text{init}}$ were loosely bounded between $[-50, 50]$ (i.e., a maximal initial activity norm of $\approx 10^3$; compare to the much smaller activities and norms for the thalamocortical network shown in Figure 4B, Figures S.6G–I, Figures 3E–F and Figure S.3E). These solutions have some level of robustness to 1% noise in $\mathbf{c}_\mu^{\text{init}}$ (green curve), though they still are not as accurate or robust as in the thalamocortical network where both eigenvalues and eigenvectors are controlled (compare panel **J** to panel **E** for $K \geq 8$ and to Figures S.6G–I). In order to find the appropriate $\mathbf{c}_\mu^{\text{init}}$, we used a constrained linear least square algorithm and selected the best solutions among 100 different random initializations of the optimization. **K, L)** Least squared error solution – relying on the Moore–Penrose pseudoinverse – for two example motifs (**K** is the difficult square-wave motif shown in **A**&**F** and **L** is motif 'C' shown in **E**&**J**). Left: network readout, starting from $\mathbf{c}_\mu^{\text{init}}$ (red) or from a noisy version of $\mathbf{c}_\mu^{\text{init}}$ (green, adding a Gaussian vector whose expected norm is 1% of the norm of $\mathbf{c}_\mu^{\text{init}}$). Right: Activity norm (top) and example units (bottom) with or without noise in $\mathbf{c}_\mu^{\text{init}}$. Notice that the activities are extremely large, and that their orientation to the readout is fine-tuned to get a relatively accurate output if starting exactly from $\mathbf{c}_\mu^{\text{init}}$. However, the output deviates wildly if some small amount of noise is added to $\mathbf{c}_\mu^{\text{init}}$.

Figure S2: Eigenvalue control through a rank-one perturbation

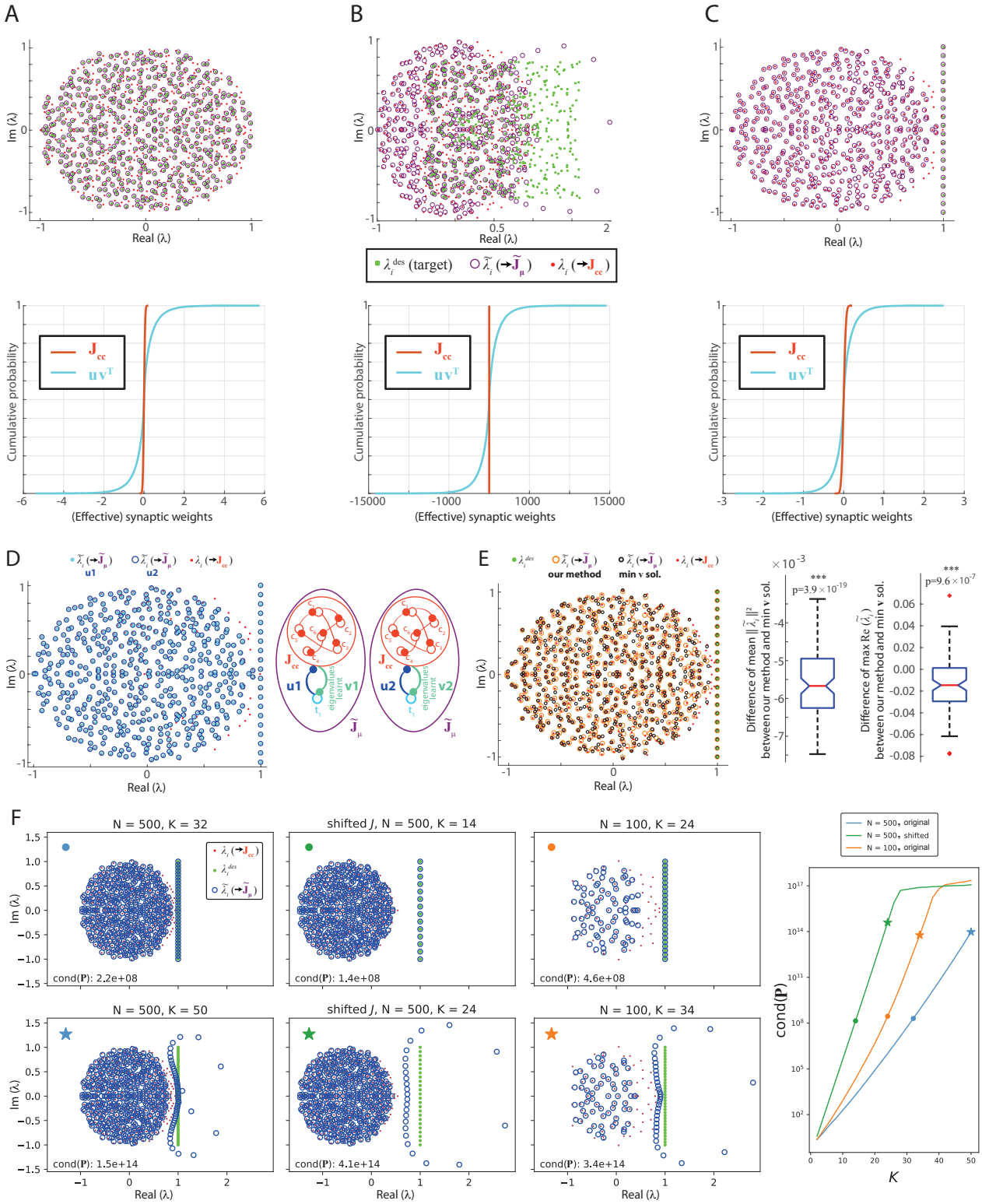


Figure S.2
See next page.

Figure S.2: eigenvalue control through a rank-one perturbation, related to Figure 1. **A,B,C)** Top: eigenspectra of the effective connectivity matrices consisting of a fixed recurrent cortical connectivity $\mathbf{J}_{cc}$ and a rank one perturbation $\mathbf{uv}^\intercal$ where $\mathbf{u}$ is a random vector and $\mathbf{v}$ is computed according to Eq. 2 ($\tilde{\lambda}_i$, purple circles), compared with the desired eigenvalues λ_i^{des} (green dots). The eigenspectrum of the unperturbed cortical connectivity matrix $\mathbf{J}_{cc}$ is given for comparison (λ_i, red dots). Bottom: respective cumulative distribution functions for *(i)* the effective loop weights $\mathbf{uv}^\intercal$ and *(ii)* the recurrent cortical connectivity weights $\mathbf{J}_{cc}$. Note that the three graphs have different scales for their x-axis. **A)** The λ_i^{des} are the eigenvalues of a different random matrix from $\mathbf{J}_{cc}$ but with the same distribution of entries. **B)** There are again N desired eigenvalues λ_i^{des} but many of them are very different from the eigenvalues of $\mathbf{J}_{cc}$. **C)** Only 20 desired eigenvalues λ_i^{des} are specified, all outside of the spectrum of $\mathbf{J}_{cc}$. **D)** Left: eigenspectra of the effective connectivity matrices for the same desired eigenvalues as in panel **C**, but starting from two different vectors **u1** and **u2** and finding two different vectors **v1** and **v2** using Eq. 2 (cyan dots and blue circles). We also show the eigenvalues of the unperturbed cortical matrix (red dots). Right: schematic illustrating the network configurations giving these (identical) eigenspectra of the perturbed matrices. **E)** Left: eigenspectra of the effective connectivity matrices for the same desired eigenvalues as in panel **C** (green dots), where the eigenvalue control is implemented by tuning $\mathbf{v}$ either using 'our method' (Eq. 2, orange circles) or using the minimal norm solution for $\mathbf{v}$ (black circles, as in (Schuessler et al. 2019)). We also show the eigenvalues of the initial recurrent cortical matrix (red dots). Middle: boxplot showing the distribution of differences in mean squared norm of the (non-desired) eigenvalues between the two methods over 100 networks (each of them with a random $\mathbf{J}_{cc}$ and $\mathbf{u}$). Eq. 2 yields significantly smaller norms (the p-value of a signed rank test is indicated). Right: boxplot showing the distribution of between-method differences in maximum real part for the non-desired eigenvalues over the same 100 networks. Eq. 2 yields a distribution of the maximum real part of the eigenvalues that is significantly shifted towards lower values (the p-value of a signed rank test is indicated). **F)** Stability of eigenvalue control with different sizes N of the recurrent cortical matrix $\mathbf{J}_{cc}$ relative to the number K of desired eigenvalues λ_i^{des}, and different distances between the eigenspectrum of $\mathbf{J}_{cc}$ and the desired eigenvalues. **Left**: example eigenspectra (eigenvalues λ_i of the unperturbed cortical matrix $\mathbf{J}_{cc}$: red dots; desired eigenvalues λ_i^{des}: green dots; and eigenvalues $\tilde{\lambda}_i$ of the effective connectivity matrix $\tilde{\mathbf{J}}_\mu$: blue circles) illustrating the outcome of eigenvalue control for relatively larger (top row) *vs.* relatively smaller (bottom row) numbers of desired eigenvalues. Left column: $N = 500$, original eigenspectrum of $\mathbf{J}_{cc}$ (as for other panels of this figure, and as in the main text). Middle column: $N = 500$, shifting the eigenspectrum of $\mathbf{J}_{cc}$ by 0.5 to the left – i.e., away from the desired eigenvalues – makes eigenvalue control harder such that fewer eigenvalues can be successfully controlled. Right column: $N = 100$, original eigenspectrum of $\mathbf{J}_{cc}$ – reducing the size of $\mathbf{J}_{cc}$ also makes eigenvalue control more difficult. The condition numbers of the matrix $\mathbf{P}$ with elements $P_{ij} = 1/(\lambda_i^{des} - \lambda_j)$ – which needs to be inverted during eigenvalue control (Eq. 2) – are indicated. **Right**: plot summarizing the condition numbers of $\mathbf{P}$ as a function of the number of desired eigenvalues K, for different sizes N and different shifts of the eigenspectrum of $\mathbf{J}_{cc}$. For a given K, the condition number is larger for smaller N or for larger shift of the eigenvalues of $\mathbf{J}_{cc}$ away from the desired eigenvalues. The condition number of $\mathbf{P}$ predicts whether eigenvalue control will be successful. Circles and stars indicate the examples illustrated on the left.

Figure S3: Non-normal properties of the motif production network

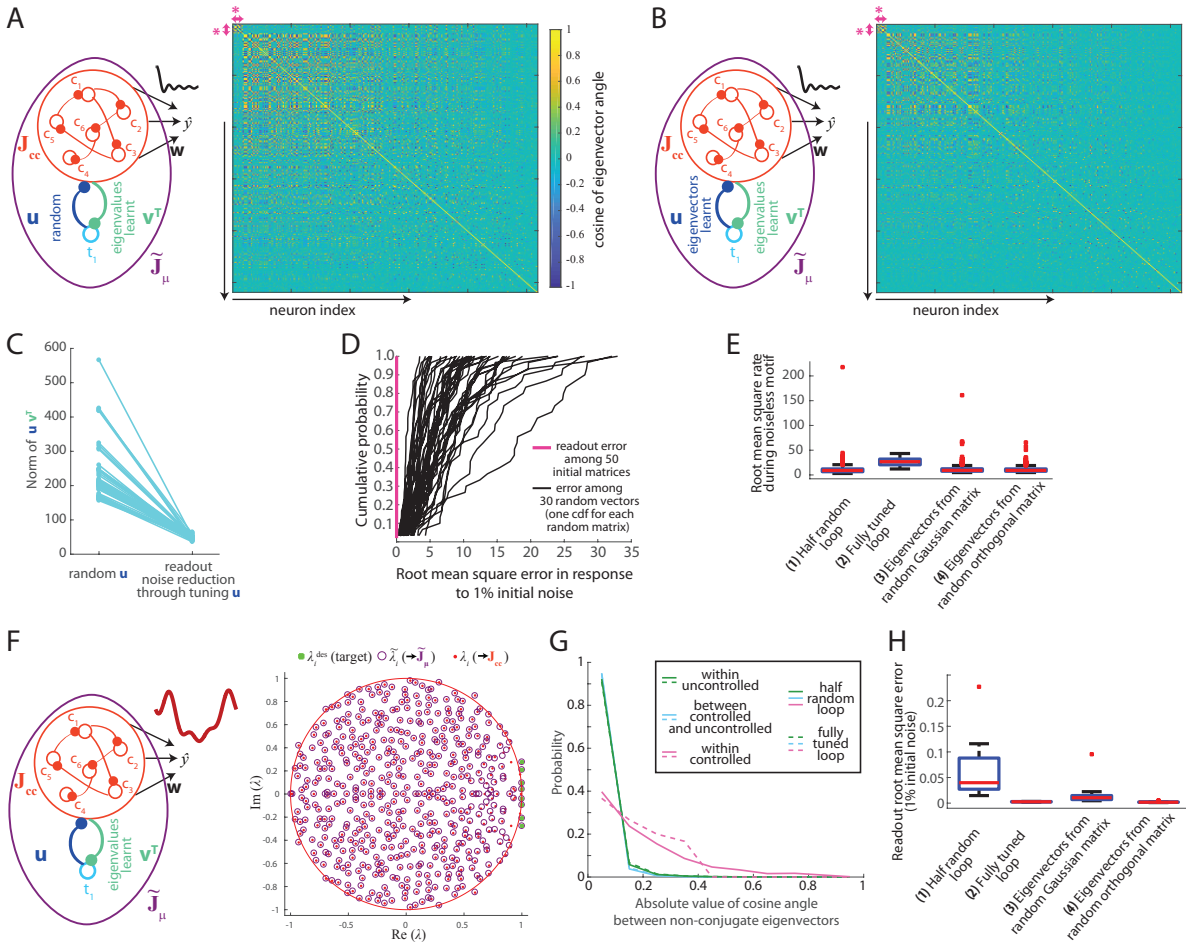


Figure S.3: non-normal properties of the motif production network, related to Figure 2. **A-E)** Networks producing the sinc function (main text Figures 1E–F and Figure 2). **A)** For a half-random thalamocortical perturbation (**left**), cosine of angle between pairs of non-complex conjugate eigenvectors (**right**). Bright yellow and deep blue indicate larger correlations. Eigenvectors are ordered such that the first ten (outlined by the pink stars) correspond to the controlled eigenvalues, while subsequent indices correspond to non-controlled eigenvalues ordered from smallest to largest eigenvalue norm. Note that the non-controlled eigenvalues with smallest norm are all at the center of the bulk of the eigenspectrum. **B)** Same as **A** but in the case of a fully learned loop. **C)** Decrease of the Frobenius norm of $\mathbf{uv}^\intercal$ between half-random and fully-tuned loops (50 different $\mathbf{J}_{\rm cc}$ and $\mathbf{w}$). **D)** Cumulative distributions of mean squared deviation due to initial noise in networks with fully tuned loops, separately along the readout direction (pink) or along thirty different random directions (black, each line is the distribution over these random directions for a given $\mathbf{J}_{\rm cc}$ and $\mathbf{w}$). **E)** Root mean square rate during noiseless motif production (starting from $\mathbf{c}_\mu^{\rm init}$) in networks with a half-random loop (1), after full optimization of the loop (2), or in matrices with the same eigenspectrum as $\tilde{\mathbf{J}}_\mu$ but with either eigenvectors from random matrices (3) or orthogonal eigenvectors (4). **F)** For a thalamocortical loop tuned to produce the motif 'C' from Figure 4 (inset), eigenspectrum of the effective matrix $\tilde{\mathbf{J}}_\mu$ (purple circles) and target desired eigenvalues (green dots). The eigenvalues of the unperturbed cortical matrix $\mathbf{J}_{\rm cc}$ are also shown (red dots) as well as the limits of the eigenspectrum that $\mathbf{J}_{\rm cc}$ would have if its size was infinite (red circle). **G,H)** Properties of networks optimized for five oscillatory motifs (including the three motifs of Figure 4). **G)** Distributions of the absolute values of the cosines between eigenvector angles, separately if *(i)* both eigenvectors correspond to controlled eigenvalues (magenta), *(ii)* only one of the two eigenvectors of the pair corresponds to a controlled eigenvalue (cyan), *(iii)* both eigenvectors correspond to uncontrolled eigenvalues (green). We further differentiate between a half-random loop and a fully tuned loop. **H)** Root mean square error of the output in the presence of 1% noise. Groups defined as in **E**.

Figure S4: Synaptic weight distributions of the motif production thalamocortical circuits

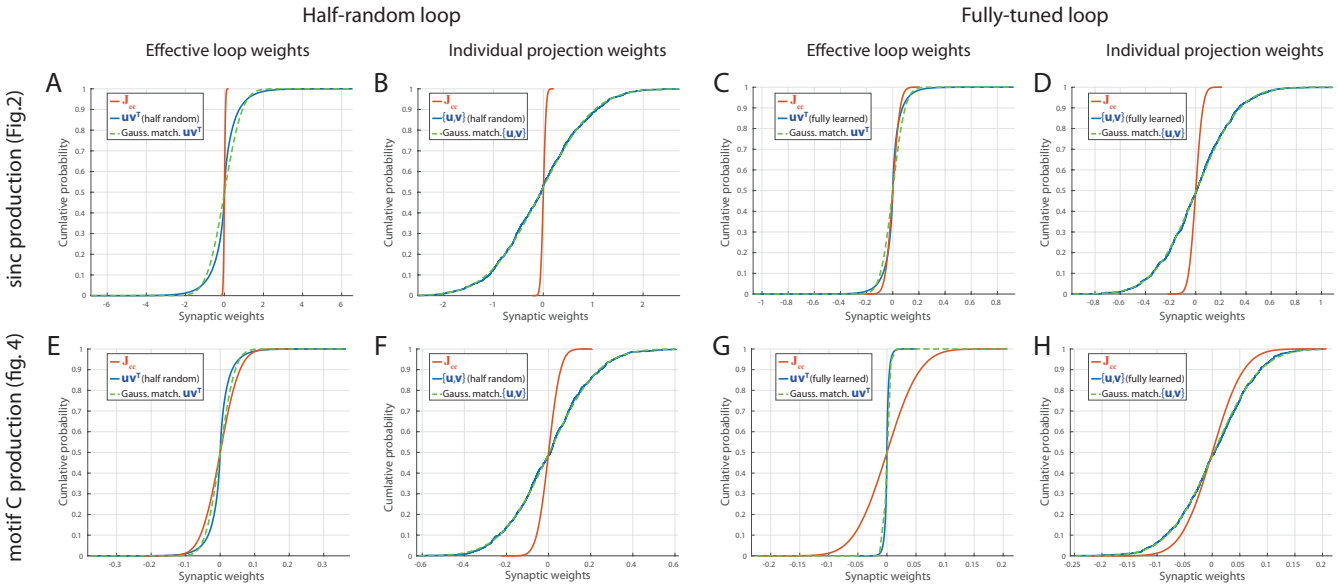


Figure S.4: synaptic weight distributions of the motif production thalamocortical circuits, related to Figure 2. Cumulative distribution functions (cdfs) for the thalamocortical networks' weights, during *(i)* the production of the sinc function as in Figure 2 (panels **A**–**D**), *(ii)* the production of motif 'C' as in Figure 4 (panels **E**–**H**). As a reference, the cdf for the weights in $\mathbf{J}_{\text{cc}}$ is shown in red in all panels (i.e., a particular drawing from a centered Gaussian distribution with standard deviation $1/\sqrt{N}$ where $N = 500$ is the cortical network's size). In all panels, the blue curve is the cdf for the weights of a rank-one perturbation tuned for motif production: either half-tuned weights which set the target eigenvalues (panels **A** &**B** and **E** &**F**) or fully-tuned weights which are also optimized for noise robustness (panels **C** &**D** and **G** &**H**). In addition, we show both the effective perturbation weights (i.e., the entries of $\mathbf{uv}^\intercal$; panels **A**, **C**, **E** and **G**) or the individual weights between units (i.e., the mixture distribution gathering the entries of $\mathbf{u}$ and $\mathbf{v}$, symbolized by $\{\mathbf{u}, \mathbf{v}\}$; panels **B**, **D**, **F** and **H**). Note that there is a certain degree of freedom for choosing the norm of the individual vectors $\mathbf{u}$ and $\mathbf{v}$ as multiplying one vector by a value and dividing the other by the same value leaves the perturbation $\mathbf{uv}^\intercal$ unchanged. Here, we chose a normalization factor such that the vectors $\mathbf{u}$ and $\mathbf{v}$ have the same norm.

Figure S5: Robustness relative to small levels of noise in the weights

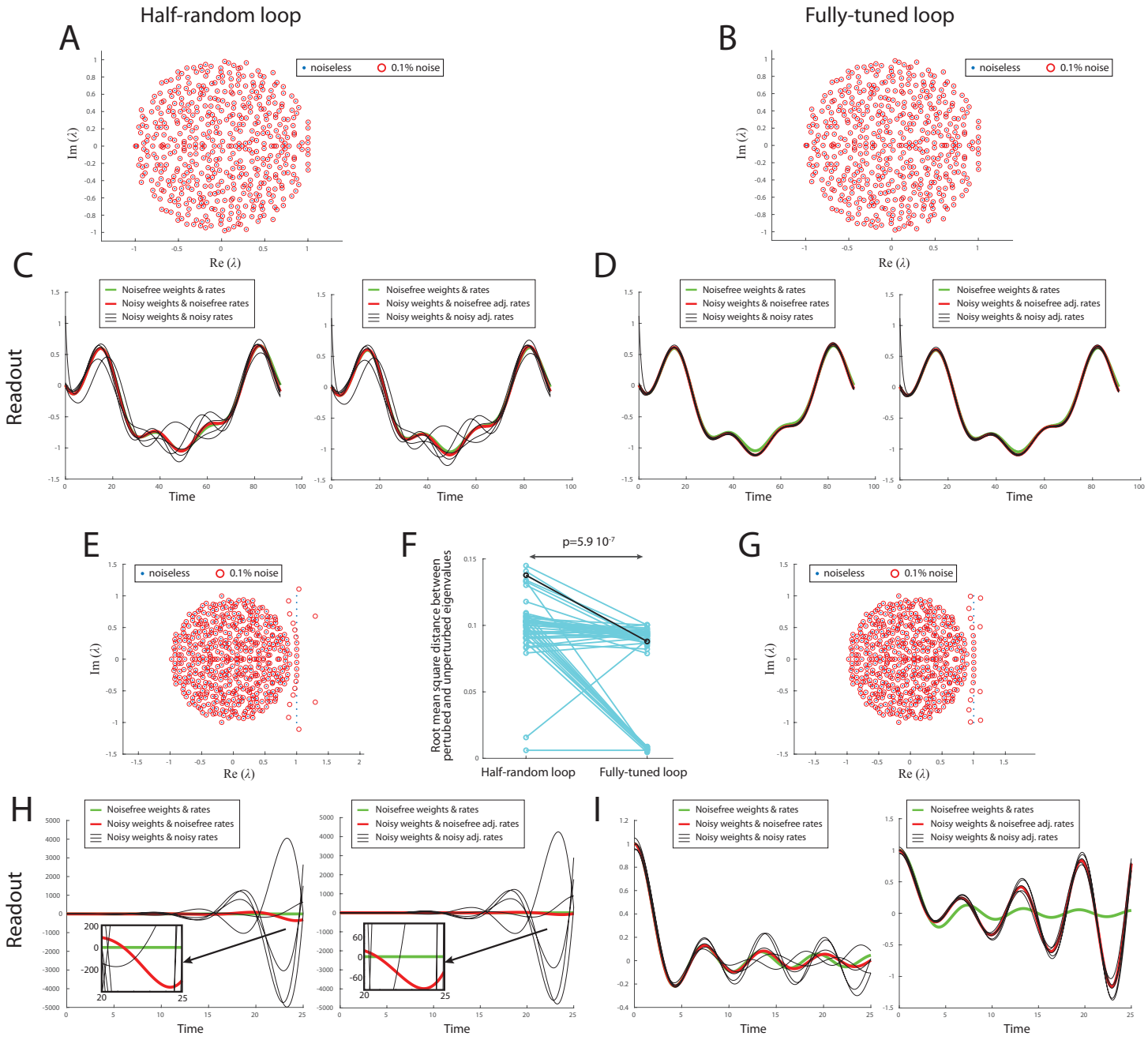


Figure S.5
See next page.

Figure S.5: robustness relative to small levels of noise in the weights, related to Figure 2. Dynamics of the thalamocortical circuit in the presence of small levels of synaptic noise (noise matrix with Frobenius norm 0.1% of $\mathbf{J}_{\text{cc}}$'s Frobenius norm). **A-D**) Production of an 'easier' motif relying on $K = 8$ eigenmodes (motif 'C' from Figure 4); **E-I**) production of a 'harder' motif relying on $K = 20$ eigenmodes (the motif from Figure 2). **A,C,E,H**) Half random thalamocortical loop (eigenvalues controlled by tuning $\mathbf{v}$ using Eq. 2); **B,D,G,I**) fully tuned loop (additionally tuning $\mathbf{u}$ to improve the robustness to noise in the activities by optimizing $C(\mathbf{u})$; Eq. 22). **A,B,E,G**) Eigenspectra of $\tilde{\mathbf{J}}_\mu$ in presence of noise *vs.* without noise. Note that we use the same noise between the half-random and fully-tuned loop cases. **F**) In the case of the larger modulation of $\tilde{\mathbf{J}}_\mu$ (20 controlled eigenvalues), root mean square norm of the difference between the eigenvalues between the noisy and noiseless matrices, for 50 draws of $\mathbf{J}_{\text{cc}}$, their associated half-random and fully-tuned loops, and the noise. The difference is significantly smaller in the fully-tuned case. The black line corresponds to the eigenspectra in panels **E** and **G**. **C,D,H,I**) Left: network output, using the original initial activities $\mathbf{c}_\mu^{\text{init}}$, in the noiseless case (green), when adding noise in the weights (red), and for noisy weights in five instances of noisy $\mathbf{c}_\mu^{\text{init}}$ (black). Right: same as left, with the difference that for the red and black curves the initial activities are remapped to adapt to the change in the eigenvectors (using Eq. 16 with the eigenvectors of the noisy $\tilde{\mathbf{J}}_\mu$) to solely visualize the effect of the modifications of the eigenspectrum. The level of noise in the initial rates was 5% in panels **C**&**D**; and 1% in **H**&**I**.

Figure S6: Dynamics and loop weights of the motor preparation network

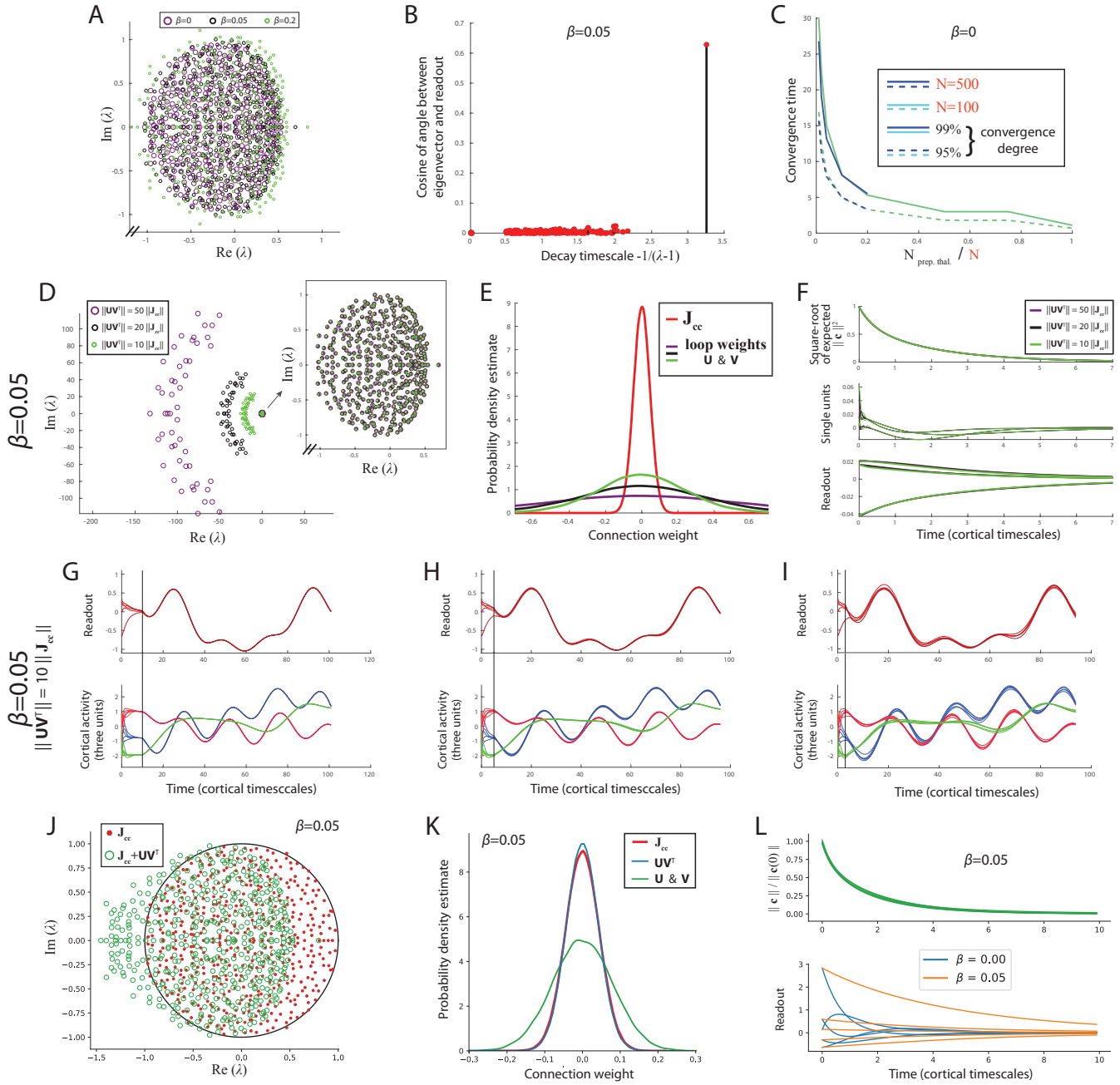


Figure S.6
See next page.

Figure S.6: dynamics and loop weights of the motor preparation network, related to Figure 3. Analysis of the thalamic preparatory network (for all but panel **C**: 50 loops for 500 cortical units). **A–C)** Effect of the parameter β after optimization of Eq. 27. **A)** Eigenspectra of $\mathbf{J}_{cc} + \mathbf{U}\mathbf{V}^\intercal$; increasing β mostly increases the real part of the rightmost (real) eigenvalue. **B)** Cosine of angle between the eigenvector and the readout, as a function of the exponential decay timescale. The slowest eigenmode is aligned with the readout. **C)** For $\beta = 0$, convergence time (number of cortical timescales such that the square root of the expected square norm of the activity decays to the indicated fraction of its initial value). These times are similar to those for $\beta = 0.05$ shown in Figure 3D. **D–F)** Effect of scaling $||\mathbf{U}\mathbf{V}^\intercal||$ reached after optimizing Eq. 27 with $\beta = 0.5$. **D)** The left outliers of the eigenspectrum – which are equal in number to the rank of $\mathbf{U}\mathbf{V}^\intercal$ – become smaller when reducing the scale of the perturbation compared to their original large norm after optimization; however the rightmost part of the spectrum is almost unchanged (inset) and, importantly, remains shifted to the left of the stability line. **E)** Distribution of the perturbation weights (entries of the matrices $\mathbf{U}$ and $\mathbf{V}$ – whose corresponding columns are chosen to have the same norm without loss of generalization) for the same scalings of $||\mathbf{U}\mathbf{V}^\intercal||$ as in panel **D**. **F)** Square root of expected squared activity norms, single unit examples, and readout when starting with random initial activities of norm one. The readout is smoother and slower to decay than the activity norm or the single units. **G–I)** Readout and activities when preparing and executing motif 'C' from Figure 4 following a preparatory duration of 10, 5 and 3 timescales for **G**, **H**, and **I** respectively (as indicated by a vertical black line). Different lines of the same color indicate different initial conditions. **J–L)** Optimizing Eq. 27 while constraining the solution to obey $||\mathbf{U}\mathbf{V}^\intercal||_{\text{fro}} = ||\mathbf{J}_{cc}||_{\text{fro}}$, for $\beta = 0.5$ (except for the bottom panel of **L** which also shows $\beta = 0$). **J)** Eigenspectra of the perturbed matrix $\mathbf{J}_{cc} + \mathbf{U}\mathbf{V}^\intercal$ (notice the absence of eigenvalues with very negative real part as in panel **D**) and of $\mathbf{J}_{cc}$. **K)** Corresponding synaptic weight distributions. **L)** Normalized cortical activity norm and readout trajectories (which, for the latter, contrast $\beta = 0.05$ which is smoother with $\beta = 0$ which is not) over different random initial conditions.

Figure S7: More biologically plausible dynamics (positive rates and non-instantaneous thalamus)

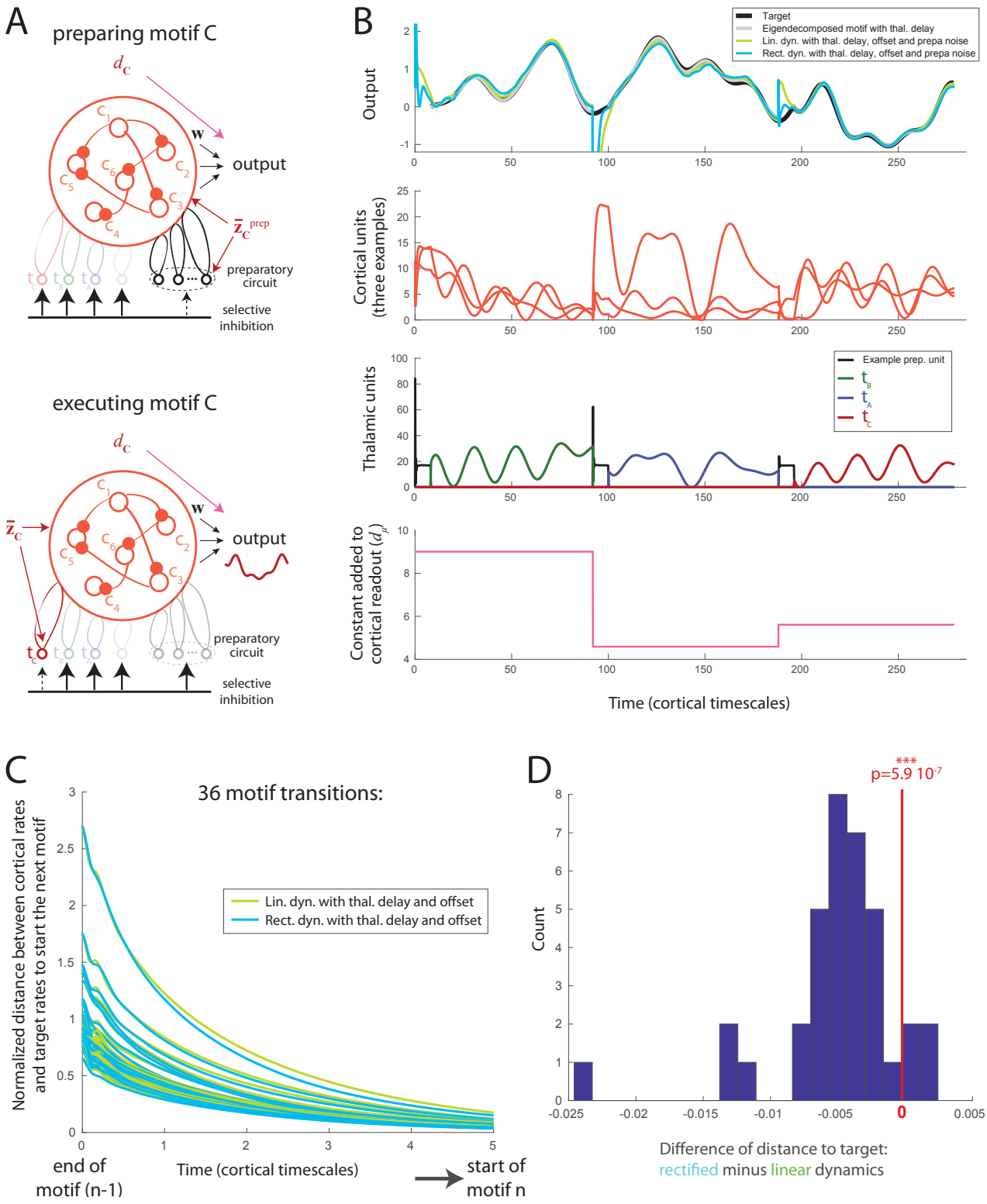


Figure S.7: See next page.

Figure S.7: more biologically plausible dynamics (positive rates and non-instantaneous thalamus), related to Figure 4. **A)** Schematic showing the added inputs to the dynamics (for motif μ,$\bar{\mathbf{z}}^{\text{prep}}_{m\,u}$ during preparation – top – and $\bar{\mathbf{z}}_{m\,u}$ during execution – bottom) to shift it such that the activities stay close to the linear regime, as well as the bias (d_μ) that is added to the cortical readout to form the total output. This bias removes the offset of the dynamics induced by shifting the activities to positive values. Note that the necessity of this bias added to the readout is a consequence of strongly constraining the dynamics to have very similar eigenspectra as in the main text, rather than a strict limitation of the framework (see 4.2.6.3). **B)** First row: output for a network trained to produce different sequences, while accounting for important biological constraints, as compared to the desired sequence (black). Second and third rows: rates of the cortical and thalamic units, respectively. Fourth row: constant d_μ added to the cortical readout $\mathbf{w}^\intercal\mathbf{c}$ at all times to form the total network output. First, thalamic units are constrained to be realistically fast (10 times faster than the cortical units). The eigenvalues are still constrained using the same theoretical framework as in the main text (Eq. 2) which assumes an idealized network with infinitely fast thalamic responses, but the deviations induced by the thalamic timescale are interpreted as a biased noise whose effects on the readout are minimized along with the effect of initial rates' noise (Eq. 37). The effect of introducing the thalamic timescale on motif production is visible by comparing the black target readout with the grey lines, which show the cortical readout when exactly adjusting the rates to their ideal values at the beginning of each motif. In addition, the rates are constrained to be mostly positive in the linear regime by adding motif-specific constant inputs to the dynamics, and we set the cortical rate pattern at the beginning of each motif using a preparatory thalamic network. The green line is the readout of these offset linear dynamics, while the cyan line is the result when additionally rectifying the rates below zero. Note that we used a restricted number of thalamic units to optimize the preparatory network (10% of the cortical size or 50 units), and the norm of the thalamic preparatory perturbation weights was set to ten times the norm of the cortical weight matrix (as in Figure S.6D–E). See Methods Sec. 4.2.6.3 for details. **C)** Details of the effectiveness of the preparatory network constrained by the presence of a thalamic timescale and of constant inputs setting the rates to mostly positive values during linear dynamics. The x-axis is restricted to the start of the preparatory period which links two motifs, and 36 motif transitions are illustrated. Each trace shows the euclidean distance between the cortical rates and the target cortical rate pattern for starting the next motif, relative to the norm of the target cortical pattern. The green lines correspond to offset linear dynamics while the rectified dynamics are shown in cyan. **D)** Difference of final distance from the target rates between rectified and linear dynamics in **C**, showing that the rectified dynamics often cause the cortical rates to converge faster than in the case of linear dynamics.