INTRINSIC TIMESCALE PLACEMENT SHAPES TEMPORAL COMPUTATION AND LEARNING IN RECURRENT NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural systems comprise units with varying intrinsic timescales, but how this heterogeneity contributes to temporal computation and interacts with recurrent dynamics is not fully resolved. We show that intrinsic heterogeneity provides a richer temporal basis for a readout; however, in our temporal tasks, recurrence can largely reduce this advantage as diverse effective dynamics emerge. By fixing the recurrent graph, its weights, and the set of intrinsic timescales while varying only their assignment across nodes, we find that placement systematically changes the effective temporal repertoire. While degree and strength effects are topology dependent, alignment with slow recurrent modes shows more consistent effects across the tested topology families, indicating slow recurrent structure as a more general organizing coordinate. These dynamical effects also influence function: under fixed recurrent connectivity, slow mode alignment improves temporal computation in the tested tasks and topology families, with generally larger benefits at greater temporal demand, and serves as an inductive bias when recurrence is trainable. Finally, we test whether this principle is reflected in human cortical organization. Empirical magnetoencephalography (MEG) neural timescales are positively associated with slow modes of the structural connectome, and their observed spatial ordering produces a broader effective temporal repertoire than spatially rotated alternatives. Taken together, these results suggest that placement relative to slow recurrent structure is an organizing factor in how intrinsic timescale heterogeneity contributes to network dynamics, computation, and learning.

1 INTRODUCTION

Neural computation unfolds over timescales from rapid sensory processing to maintenance of information in memory. Neural activity exhibits temporal heterogeneity throughout the cortex and within local populations; sensory regions fluctuate faster than higher-order prefrontal systems, and neurons within a cortical region can display a range of intrinsic timescales (Murray et al., 2014; Bernacchia et al., 2011; Wasmuht et al., 2018; Abulikemu et al., 2026). Such variation in intrinsic timescales is functionally important, with longer timescales associated with stronger and more persistent task representations (Cavanagh et al., 2016; Wasmuht et al., 2018).

One source of this variation is heterogeneity in local neuronal and synaptic time constants, which can supply complementary fast and slow temporal filters. Distinct or adaptive time constants are shown to support temporal hierarchy in recurrent networks (Yamashita & Tani, 2008; Quax et al., 2020), and heterogeneous membrane and synaptic time constants improve learning on temporally structured tasks (Perez-Nieves et al., 2021). Effective empirical activity dynamics, however, do not necessarily correspond to a single local time constant (Zeraati et al., 2026).

Recurrence provides another source of network temporal structure. Recurrent state retains information of past inputs decodable by downstream readouts (Ganguli et al., 2008; Dambre et al., 2012), and recurrent interactions modulate the expression of local temporal processes at the network level (Beiran & Ostojic, 2019). In linear recurrent systems, eigenvalues determine the decay timescales of dynamical modes, and eigenvectors determine where those modes are spatially expressed in the network (Chaudhuri et al., 2014). Recurrent organization can correspondingly generate distinct and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

hierarchical activity timescales among assemblies and cortical regions (Chaudhuri et al., 2015; Stern et al., 2023; Zeraati et al., 2023). When local time constants and recurrent weights are jointly optimized, different training regimes can lead to local or network-mediated mechanisms for maintaining long timescales (Khajehabdollahi et al., 2024). The computational contribution of a local time constant therefore depends on its recurrent context.

Given this recurrent context dependency, an important question is whether the same set of local time constants produces different network dynamics based on their placement among nodes. Previous architectures have introduced temporal hierarchy by assigning different update rates to modules or different time constants to layers (Koutnik et al., 2014; Moro et al., 2026). Here, we directly examine timescale placement by altering which nodes receive which time constants and keeping the recurrent network and timescale set fixed. In-strength and degree are candidate structural coordinates for placement, as greater incoming connectivity is linked with slower dynamics (Sethi et al., 2017), and degree heterogeneity can facilitate slow hub dynamics (Zenari et al., 2026). These measures, however, summarize local connectivity, whereas recurrent modes provide a network-level summary of collective decay dynamics (Chaudhuri et al., 2014; Tang et al., 2020). We therefore hypothesize that aligning slower intrinsic timescales with stronger participation in slow recurrent modes provides a more direct placement coordinate. We compute this coordinate independently of the tested assignment, through a reference network whose units all share the same local time constant.

For such organization to be computationally meaningful, placement should have effects beyond spontaneous dynamics, and its effect should become more apparent when information must be retained for longer periods. We study this functional effect by comparing task performance across placements with fixed recurrent weights, then enable recurrence to adapt to test whether aligned placement provides an advantageous starting organization. Pre-learning dynamics can influence learning speed (Kurikawa & Kaneko, 2025), and recurrent learning can develop longer network-mediated dynamics without corresponding changes in local time constants (Khajehabdollahi et al., 2024). Thus, placement may act as a dynamical inductive bias, and recurrent plasticity may partly compensate for the initially unfavorable placement.

A complementary question is whether human cortical organization parallels this placement hypothesis. Resting functional MRI timescales are associated with structural connectivity strength (Fallon et al., 2020), and connectome-constrained models show that white matter connectivity shapes regional temporal processing (Triebkorn et al., 2025). Recent network control modeling links inferred heterogeneous regional decay rates to empirical activity timescales and slow connectome modes (Kim et al., 2025). Here, we investigate whether an independently measured cortical timescale map aligns with slow modes of the structural connectome and whether its empirical spatial arrangement leads to predicted dynamical consequences. We use intrinsic activity timescales estimated from magnetoencephalography (MEG) (Shafiei et al., 2023), which are treated as regional neural activity measures rather than direct measurements of local time constants (Zeraati et al., 2026).

Contributions We present four main findings. First, we isolate timescale placement from the recurrent network and timescale distribution, and show that slow mode alignment organizes the realized temporal repertoire more consistently compared to local structural coordinates. Second, we establish functional benefit of this alignment for temporal computation with increasing temporal demand. Third, we show that placement biases persist under trainable recurrence, while recurrent plasticity partly compensates for unfavorable placement and reorganizes slow recurrent structure. Lastly, we test the placement principle in human cortex, where empirical MEG timescale organization is associated with slow structural connectome modes and exhibits the predicted dynamical effect.

2 METHODS

2.1 EXPERIMENTAL OVERVIEW

Figure 1 summarizes the five-stage experimental design. We first characterize the contribution of intrinsic timescale heterogeneity with and without recurrence (Section 2.2; Result 3.1), then examine how timescale placement shapes effective dynamics and temporal computation under fixed recurrence (Section 2.3; Results 3.2–3.3). We next test placement during recurrent learning (Section 2.4; Result 3.4) and finally examine whether the same organizational principle is reflected in human cortical data (Section 2.5; Result 3.5).

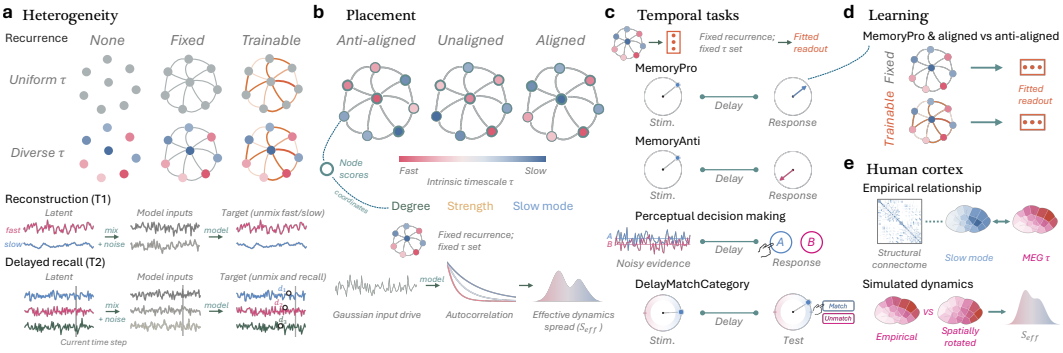


Figure 1: **Experimental overview. a** Intrinsic timescale heterogeneity is compared under no, fixed, and trainable recurrence with two diagnostic tasks, latent reconstruction (T1) and delayed recall (T2). **b** Keeping recurrent graph, weights, and timescale set fixed, timescales are permuted according to degree, strength, or slow mode coordinates. A Gaussian drive is used to estimate effective timescales and their spread S_{eff}. **c** Four cognitive tasks test functional consequences of placement with fixed recurrence and fitted readouts. **d** Learning is assessed using MemoryPro with aligned and anti-aligned placements under frozen or trainable recurrence; readouts fitted in both. **e** Human cortical analyses link empirical MEG timescales to slow connectome structure and compare the effective temporal repertoire in empirical and spatially rotated placements. Task signals, distributions, and cortical maps are shown schematically.

In model placement experiments, placement rules are defined from recurrent network properties before any task performance is evaluated, and effects are estimated by comparing different timescale assignments within the same independently generated network. The initial timescale heterogeneity and untasked placement experiments use 64-unit networks; the fixed recurrence behavioral study and recurrent learning study use 128-unit Erdős-Rényi (ER), heavy-tail, and modular networks. The cortical analysis uses a fixed 400-parcel human structural connectome with empirical MEG activity timescales mapped to the corresponding parcels.

2.2 Network model and temporal organization

We study leaky recurrent neural networks (RNNs) with neuron-level intrinsic timescales,

$$h_i(t) = (1-\alpha_i)h_i(t-1) + \alpha_i\,\phi((W_{\text{in}}x_t)_i + (W_{\text{rec}}h_{t-1})_i + b_i)\,, \qquad \alpha_i = 1 - e^{-\Delta t/\tau_i}, \quad (1)$$

where $h_i(t)$ is the state of unit i, x_t is the input, and ϕ is ReLU; larger τ_i results in slower updating. The timescale heterogeneity experiments compare non-recurrent leaky units ($W_{\text{rec}} = 0$), a reservoir with fixed recurrence, and an RNN with trainable recurrence (Fig. 1a). Homogeneous models learn one globally shared τ, and heterogeneous models learn separate τ_i for each unit. We use two diagnostic tasks; T1 recovers latent signals evolving at different speeds from noisy mixtures, and T2 recalls random latent inputs after fixed delays. Performance G is normalized relative to a memoryless baseline ($G = 0$) and a task-specific analytic reference ($G = 1$) (Appendix A).

Recurrence can make a unit's activity timescale differ from its intrinsic τ_i. Under a common Gaussian white noise drive, we estimate the effective timescale τ_i^{net} from the unit autocorrelation and define the realized temporal repertoire as $S_{\text{eff}} = \text{std}_i[\log \tau_i^{\text{net}}]$, using logarithms to capture proportional differences between timescales. Larger S_{eff} denotes a broader spread of effective timescales, which reflect the interaction between intrinsic and recurrent dynamics. Across placement studies, we use structural coordinates including degree and absolute weighted in-strength, and a spectral slow mode coordinate. To characterize where the recurrent network itself supports slow dynamics independently of heterogeneous intrinsic timescales, and motivated by eigenvector localization (Chaudhuri et al., 2014), we define the slow mode coordinate from a homogeneous reference system,

$$J_0 = (1-\alpha_0)I + \alpha_0 \bar{D} W_{\text{rec}}, \quad (2)$$

where $\alpha_0 = 1 - e^{-\Delta t/\tau_0}$ corresponds to a single reference timescale τ_0, and $\bar{D}_{ii}$ is the proportion of timesteps for which unit i has a nonzero ReLU activation under a fixed Gaussian reference drive.

Each node is scored by its loading onto the modes of J_0, where greater weights are given to modes decaying more slowly, so larger values indicate stronger participation in slow recurrent activity. As the reference system and resulting coordinate are computed before heterogeneous timescales are assigned, they do not depend on the placement being evaluated (Appendix C). The connectome analysis uses the same slow mode construction with a linear reference system defined in Section 2.5.

2.3 Timescale placement and temporal computation with fixed recurrence

In the untasked placement study, we vary the arrangement of intrinsic timescales using the structural and slow mode coordinates defined in the previous section and measure the resulting effective dynamics (Fig. 1b). We use 64-unit reservoirs and permute timescales to vary the Spearman correlation between τ_i and degree, in-strength, or the slow mode coordinate from anti-aligned to aligned configurations. All placements receive the same Gaussian probe, and the resulting temporal repertoire is quantified by S_{eff} (Appendices B–C).

To test whether the same placement intervention influences temporal computation, we use 128-unit ER, heavy-tail, and modular reservoirs with $\Delta t = 20$ ms and a shared set of 128 log-spaced timescales from 40 to 640 ms. Slow mode placement is the primary manipulation, with degree placement serving as a structural comparison. We evaluate 16 independent networks per topology at seven target alignment levels from -1 to 1. We also test the primary placement effect for size robustness in 64-, 128-, and 256-unit networks (Appendix D).

We vary temporal demand across seven delay levels (up to 2400 ms) in four cognitive tasks (Fig. 1c). MemoryPro requires maintaining a stimulus direction across a delay and responding in the same direction, whereas MemoryAnti requires responding in the opposite direction (Yang et al., 2019; Driscoll et al., 2024). The perceptual decision making (PDM) task requires integrating noisy stimulus evidence into a binary decision and retaining it through a delay (Yang et al., 2019; Inagaki et al., 2019). DelayMatchCategory requires retaining the category of a sample stimulus and deciding whether a later test stimulus belongs to the same category (Freedman & Assad, 2006). For each placement, we fix the reservoir and fit a ridge-regularized linear readout.

The ridge penalty, task difficulty, and intrinsic timescale range are calibrated using strictly random placements without inspecting aligned versus anti-aligned performance. The primary behavioral measure is the error contrast $B(d) = E_{\mathrm{anti}}(d) - E_{\mathrm{aligned}}(d)$, where positive values indicate an aligned placement advantage. At 2400 ms, we estimate 95% confidence intervals using paired bootstrap resampling across networks and assess statistical significance with paired sign-flip tests. As a secondary analysis, we examine the relationship between placement-induced S_{eff} and behavior.

2.4 Learning and recurrent plasticity

To test whether placement acts as an inductive bias when recurrence can adapt, we use 24 new 128-unit RNNs (eight per topology family) with MemoryPro at delays of 400, 800, 1600, and 2400 ms. Intrinsic timescales are placed at moderate target correlations of ± 0.5 with either the slow mode coordinate or degree; slow mode placements are constrained to remain nearly neutral with respect to degree. Each placement is evaluated with recurrence either frozen or trainable, with the output readout trained in both conditions. Paired frozen and trainable runs use matched initialization, training data, and evaluation trials (Appendix E).

We first ask whether aligned placement improves early learning and whether this advantage increases with temporal demand. Early learning is quantified as the cumulative improvement in held-out error over a predefined initial training phase relative to each run's initial error, with larger values indicating faster learning. The primary analysis tests whether the aligned versus anti-aligned learning advantage increases with delay within each network when recurrence is trainable. A secondary analysis tests whether the achieved slow mode alignment across trainable conditions predicts this demand dependent learning effect after controlling for degree alignment.

We next assess whether recurrent plasticity compensates for unfavorable placement. At the maximal 2400 ms delay, we compare the late aligned versus anti-aligned performance gap under frozen and trainable recurrence. Because intrinsic timescales remain fixed during training, any reduction in their mismatch with slow recurrent structure must arise from changes in W_{rec}. We therefore track dynamical compensation by measuring whether initially anti-aligned networks increase the align-

ment between their fixed intrinsic timescales and the evolving slow recurrent structure. Inference is performed at the network level using bootstrap confidence intervals and paired tests (Appendix E).

2.5 Human connectome and cortical timescale organization

The final analysis examines the consistency of human cortical timescale organization with slow mode placement and its predicted dynamics. We use a group-representative structural connectome from 327 Human Connectome Project (HCP) participants (Van Essen et al., 2013; Milisav et al., 2026) and a group-average MEG timescale map from 33 HCP participants (Shafiei et al., 2023), both in the 400-region Schaefer atlas (Schaefer et al., 2018). The structural connectivity matrix A is spectrally normalized as $W = gA/\lambda_{\max}(A)$, with near-critical gain $g = 0.9$. We define the homogeneous linear reference system

$$J_0 = (1 - \alpha_0)I + \alpha_0 W, \tag{3}$$

with $\alpha_0 = 1 - e^{-1/\tau_0}$ and no ReLU gating. Given that J_0 and W share eigenvectors, each parcel is mapped by its participation in connectome eigenmodes with slower modes weighted more strongly. We test the association between cortical timescales and slow mode score using Spearman correlation.

To test the dynamical consequence, the structural connectome and timescale set are fixed, and spatially rotated MEG maps provide null placements that preserve cortical organization but change their relation to slow mode structure. For each placement, the linear reservoir with recurrence W is driven by Gaussian white noise and its temporal repertoire quantified with S_{eff}. The empirical ordering is compared with the rotated null using a one-sided permutation test (Appendix F).

3 Results

3.1 Intrinsic heterogeneity enriches temporal representations; recurrence reduces its advantage

Without recurrence, intrinsic timescale heterogeneity improved temporal representations, particularly when the underlying temporal structure was more diverse. On T1, the heterogeneous advantage increased from $\Delta G = 0.013$ when the two latent signals had the same timescale to 0.256 when their timescales separated by a factor of 2^4 (Fig. 2a). T2 showed a complementary view of this benefit (Fig. 2b); heterogeneous timescales supplied distinct temporal kernels that a readout can combine, resulting in better representation of a delayed target (0.214 versus 0.043 at the delay target 8) and more recoverable past input (1.388 versus 0.816 summed over lags).

This intrinsic advantage diminished once recurrence could itself generate temporal diversity. Across non-recurrence, reservoir, and trained RNN, homogeneous performance on the hardest T1 condition increased from $G = 0.731$ to 0.934 to 0.994, while heterogeneous performance remained high

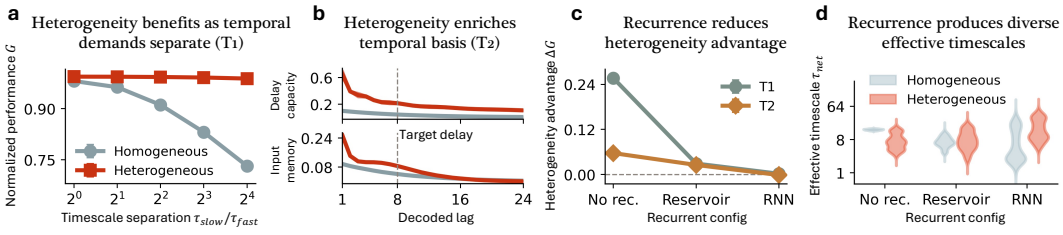


Figure 2: **Intrinsic heterogeneity enriches temporal representations and recurrence reduces its advantage in the diagnostic tasks.** **a** In non-recurrent leaky units, heterogeneous networks (red) exceed homogeneous networks (gray) on T1, and the difference increases with latent timescale separation. **b** In T2, heterogeneity provides a richer set of temporal kernels (top) and better recovery of past input (bottom); dashed line represents the target delay. Definitions of both diagnostics are given in Appendix A. **c** The heterogeneity advantage ΔG declines as recurrence is added on both tasks. **d** With recurrence, homogeneous networks develop a spread of τ_{net} approaching that of heterogeneous networks.

throughout. The advantage decreased from 0.257 95% CI $[0.252, 0.260]$ to 0.029 $[0.022, 0.039]$ and 0.002 $[0.001, 0.003]$, and T2 displayed the same effect (Fig. 2c). Moreover, homogeneous networks developed a spread of effective timescales despite sharing a single intrinsic timescale, with S_{eff} increasing from 0.046 $[0.042, 0.053]$ to 0.296 $[0.237, 0.381]$ and 1.076 $[0.856, 1.251]$ (Fig. 2d). Thus, in these task settings, recurrence can largely supply the effective temporal heterogeneity that intrinsic diversity otherwise provides. This shifts the focus from whether heterogeneous timescales are present to how a fixed set of intrinsic timescales is organized relative to recurrent structure.

3.2 Timescale placement shapes effective dynamics through slow modes

The realized temporal repertoire changed when the same intrinsic timescales were permuted across a fixed reservoir. Here, the graph structure, its connection weights, and the complete set of intrinsic timescales were retained, and only which node received which timescale was varied (Fig. 3a). Placing slow timescales onto high-degree nodes broadened S_{eff}, with within-network slopes of 0.048 $[0.038, 0.058]$ in heavy-tail networks and 0.019 $[0.008, 0.032]$ in ER networks (Fig. 3b). When comparing extremes, the aligned regime exceeded the anti-aligned regime by 0.092 $[0.071, 0.112]$ in heavy-tail versus 0.041 $[0.017, 0.070]$ in ER. Thus, identical intrinsic timescales can generate different effective dynamics depending on their arrangement, with a larger degree-based effect when node degree is more heterogeneous.

The topology dependence of degree points to a dynamical rather than purely structural explanation. Degree describes local network wiring, whereas the slow mode coordinate quantifies each node's loading onto slowly decaying recurrent modes. Among all three placement coordinates, slow mode alignment produced the largest effect in both topologies, increasing S_{eff} by 0.142 $[0.113, 0.169]$ in heavy-tail and 0.147 $[0.113, 0.187]$ in ER networks (Fig. 3c). Degree and strength effects were weaker in ER, in line with their weaker correspondence to the slow mode coordinate. Degree correlated more strongly with the slow mode coordinate in heavy-tail than ER networks ($\rho = 0.529$ versus 0.263), and similarly for in-strength (0.612 versus 0.371; Fig. 3d). Structural coordinates therefore manifest stronger placement effects when they track slow recurrent modes better, suggesting that slow mode organization is the more consistent coordinate across the tested topologies.

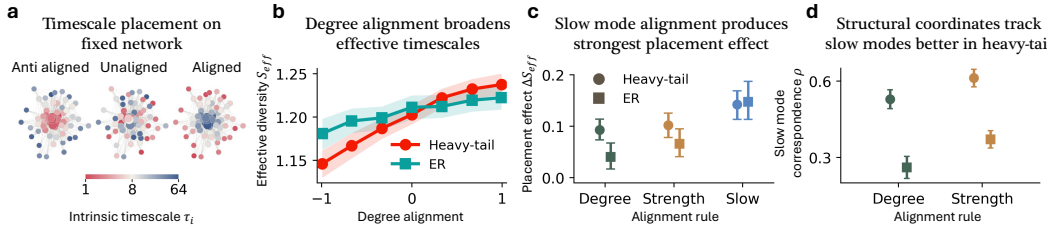


Figure 3: **Timescale placement and alignment coordinate shape the effective temporal repertoire.** **a** The same reservoir and intrinsic timescale set under different nodewise τ_i assignments; colors show intrinsic timescales. **b** Effective diversity S_{eff} increases as slower timescales move onto higher degree nodes, with a stronger effect in heavy-tail than ER networks. **c** Across placement coordinates, degree and strength show smaller effects in ER, whereas slow mode alignment shows the largest effect and is similar between topologies. **d** Spearman correlation between structural coordinates and the slow mode coordinate is stronger in heavy-tail networks.

3.3 Slow mode placement shapes temporal computation across demand

The dynamical consequences of slow mode placement were also reflected in temporal computation. In the reservoir functional experiment, aligned slow mode placement improved performance across all four tasks and all three network topologies at the maximal 2400 ms delay (Fig. 4a-d). Effects were coherent for MemoryPro ($d_z = 0.70, 0.70, 0.69$; all $p \leq .01$) and MemoryAnti ($d_z = 0.89, 0.79, 0.89$; all $p \leq .002$) across ER, heavy-tail, and modular networks, respectively. PDM showed the largest functional effects ($d_z = 1.12, 1.05, 1.15$; all $p < .001$), and DelayMatchCategory showed variable but positive effects ($d_z = 0.72, 1.11, 0.46$; $p = .007, < .001, .043$). The benefit of alignment was generally larger at higher temporal demand. Relative to the lowest de-

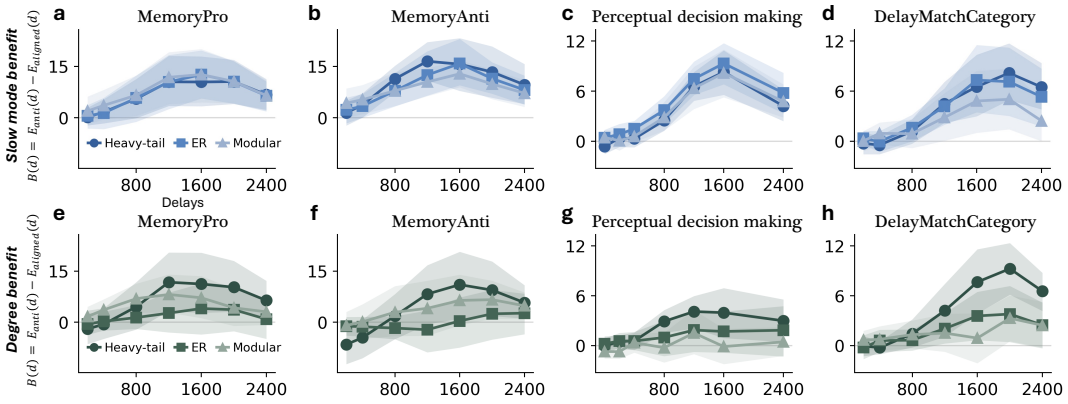


Figure 4: **Timescale placement shapes temporal computation across demand.** Aligned placement benefit $B(d) = E_{\text{anti}}(d) - E_{\text{aligned}}(d)$ with increasing temporal demand; positive values indicate better performance under aligned regime. **a-d**, slow mode placement for MemoryPro, MemoryAnti, PDM, and DelayMatchCategory, respectively. **e-h**, degree placement for the same four tasks. Slow mode benefits are generally larger at higher temporal demand and are coherent in ER, heavy-tail, and modular networks; degree placement effects are strongest in heavy-tail networks and weaker and mixed in the other topologies. Curves show topology-specific means across independently generated networks within each topology, and shaded regions show 95% confidence intervals.

mand condition, the aligned benefit was larger at 2400 ms in 9 of 12 task-by-topology comparisons at $p < .05$. This increase was observed in all three topologies for PDM and in ER and heavy-tail networks for MemoryPro, MemoryAnti, and DelayMatchCategory. Across the full trajectories, effects were generally small at low demand and increased as information had to be maintained over longer delays, peaking around 1600-2000 ms before attenuating at the hardest condition (Fig. 4a-d). Aligned slow mode placement thus became more influential as temporal demand increased, although the relationship was not strictly monotonic.

Degree placement provided a structural comparison to this pattern, where its effect was strongly topology dependent (Fig. 4e-h). In heavy-tail networks, all four effects at 2400 ms were positive ($d_z = 0.54$-1.36, all $p < .05$); in ER networks, effects were substantially weaker ($d_z = 0.08$-0.42, all $p > .05$). Modular networks were mixed, with evidence for MemoryAnti ($d_z = 0.58$, $p = .018$) but weaker effects in the remaining tasks ($p = .056$-$.316$). This coincides with the untasked dynamics; degree shows stronger effects where it better tracks slow recurrent modes, whereas slow mode placement shows a more consistent behavioral effect across network architectures.

A secondary analysis examined whether behavioral benefit was associated with the effective temporal repertoire induced by placement. Across the seven alignment levels, partial correlations between performance and S_{eff} became more positive at higher temporal demand controlling for mean effective timescale (Appendix D). At 2400 ms, all 12 estimates were positive ($\rho = 0.06$-0.71), with 10 of 12 confidence intervals entirely above zero. The effect was most consistent for MemoryPro and PDM, where all three topology specific intervals were positive, while the interval included zero for DelayMatchCategory in modular networks ($\rho = 0.30\,[-0.14, 0.63]$) and for MemoryAnti in heavy-tail ($\rho = 0.06\,[-0.31, 0.43]$). These observations align with a broader temporal repertoire contributing to the behavioral benefit, but do not indicate mediation.

3.4 Timescale placement shapes learning and recurrent plasticity compensates for unfavorable organization

When recurrent connectivity was allowed to adapt, slow mode placement shaped how efficiently temporal computation was learned. We observed that aligned slow mode placement accelerated early learning, with its advantage over anti-aligned placement increasing from 0.023 at 400 ms to 0.158 at 2400 ms (Fig. 5a). Within each network, this advantage increased with delay (slope $= 0.048$

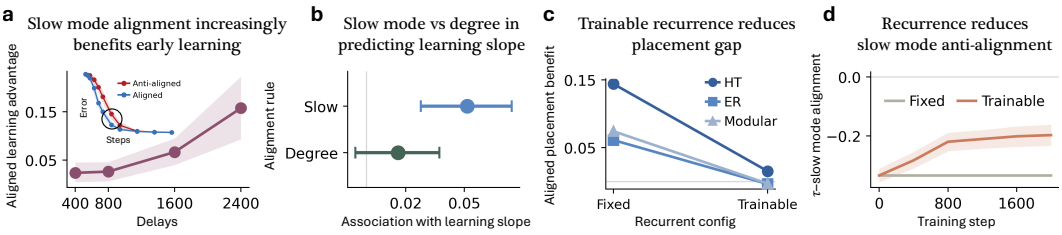


Figure 5: **Timescale placement shapes learning; recurrence compensates with reorganization.** **a** Aligned slow mode placement improves early learning, with the advantage increasing with temporal demand. **b** Across trainable conditions, slow mode alignment is more strongly associated with the demand dependent learning slope than degree alignment. **c** At 2400 ms, the aligned versus anti-aligned performance gap is smaller with trainable than frozen recurrence. **d** In initially anti-aligned networks, trainable recurrence increases alignment between fixed intrinsic timescales and slow recurrent structure. Bars and shaded regions show 95% confidence intervals.

$[0.025, 0.070]$); 20 of 24 networks showed a positive slope ($p < .001$). Thus, slow mode placement functions as a demand dependent inductive bias, with alignment becoming more beneficial as temporal demand grows.

When all trainable recurrence conditions were considered together, the demand dependent learning effect was more closely associated with slow mode alignment than degree alignment (Fig. 5b). After accounting for degree alignment and network identity, slow mode alignment was associated with the demand dependent learning slope ($\beta_{\text{slow}} = 0.051\,[0.028, 0.074]$, $p < .001$), whereas degree alignment showed a smaller independent association ($\beta_{\text{degree}} = 0.016\,[-0.006, 0.037]$, $p = .147$). The slow mode coefficient remained positive when each topology family was omitted in turn (Appendix E), indicating that it is a coordinate more coherently associated with the learning benefit.

Recurrent plasticity partly compensated for initial anti-alignment. At the maximal 2400 ms delay, we compared the aligned versus anti-aligned performance gap in matched networks with recurrence either frozen or trainable (Fig. 5c). This placement gap was smaller when recurrence was trainable, with a mean reduction of 0.090 [0.046, 0.136]; the reduction was positive in 20 of 24 networks ($p < .001$). Intrinsic timescales attached to each neuron remained fixed during training, so any increase in their alignment with slow recurrent structure had to arise from changes in the recurrent weights. In line with this, initially anti-aligned networks developed greater alignment between their fixed intrinsic timescales and the evolving slow recurrent structure under trainable recurrence (Fig. 5d). The pre- to post-training increase in slow mode alignment was $0.137\,[0.108, 0.166]$, positive in 23 of 24 networks ($p < .001$). Recurrent plasticity therefore reduced the behavioral cost of anti-aligned placement while the slow recurrent structure reorganized toward the fixed intrinsic timescale pattern.

3.5 Human cortical timescale organization aligns with the predicted dynamics of slow mode placement

The empirical MEG neural timescales were positively associated with the slow mode score of the structural connectome across 400 regions ($\rho = 0.565$, two-sided spatial rotation $p = .048$; Fig. 6a,b). Regions with longer cortical timescales therefore participated more strongly in slowly decaying connectome modes. Importantly, the empirical cortical ordering exhibited the dynamical consequence predicted by network experiments. With the structural connectome and timescale set fixed, the empirical ordering produced $S_{\text{eff}} = 1.299$, at the 97.6th percentile among spatially rotated placements ($p = .025$, one-sided; Fig. 6c). This indicates, in the connectome model, the spatial arrangement of cortical timescales contributes to the temporal repertoire beyond the timescale set itself. For rotated placements at $g = 0.90$, stronger slow mode alignment was associated with more diverse effective dynamics (Fig. 6d), in line with the relationship between placement and dynamics established in experiments. Hence, human cortical timescale organization aligns with the slow mode placement principle, and its ordering produces the predicted dynamical consequence in the connectome model.

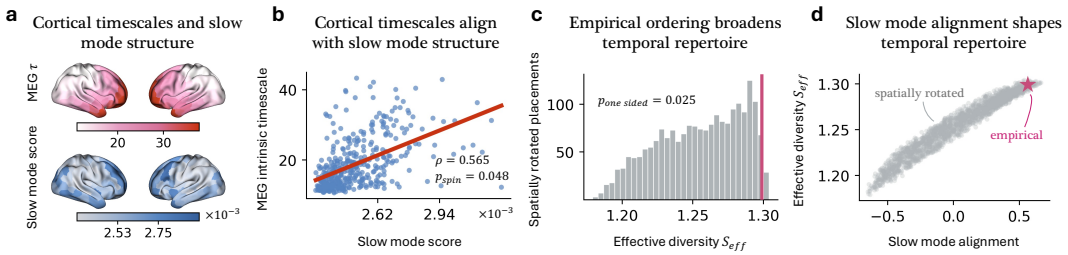


Figure 6: **Human cortical timescale organization is consistent with predicted dynamics of slow mode placement.** **a** Empirical MEG intrinsic timescales and structural connectome slow mode score cortical mapping. **b** Cortical timescales increase with the slow mode score; significance is assessed with spatial rotations. **c** The empirical cortical ordering leads to broader effective temporal repertoire S_{eff} compared to spatially rotated null placements. **d** Throughout rotated placements, stronger slow mode alignment is associated with a higher effective temporal diversity.

4 Discussion

Our study shows that rearranging an identical set of intrinsic time constants on a recurrent graph with fixed weights changes effective dynamics and computation. This indicates that effective temporal heterogeneity depends not only on intrinsic timescale diversity, but also on how these properties align with recurrent dynamics. Previous work established computational benefits of heterogeneous timescales, effects of spatially structured local heterogeneity, and relationships between inferred regional decay rates and slow network modes (Tanaka et al., 2022; Demirtaş et al., 2019; Kim et al., 2025). The present findings connect these results by showing that the correspondence between local temporal properties and recurrent structure shapes network dynamics.

The comparison between structural and slow mode coordinates suggests that placement effects are better captured by distributed recurrent dynamics than local connectivity. Degree can carry meaningful dynamical information when heterogeneity gives rise to slow hub dynamics (Zenari et al., 2026), but its relevance depends on architecture. Cortical timescale hierarchies can likewise depend on local and long-range recurrence (Chaudhuri et al., 2015; Stern et al., 2023). Local graph measures may therefore transfer less reliably between architectures. In the tested leaky-ReLU setting, across topologies and tasks, slow mode alignment consistently shaped effective temporal dynamics and computation, without implying that the reference modes fully determine nonlinear dynamics.

The functional effects show that placement has effects beyond spontaneous temporal statistics. With separate fitted readouts, performance differences reflect temporal information made linearly accessible from the recurrent state, and generally larger effects at longer delays link this organization to memory retention. Placement also biased early learning in MemoryPro, in line with evidence that pre-learning dynamics can shape subsequent learning (Kurikawa & Kaneko, 2025). Recurrent plasticity partly reduced this dependence; similarly, long timescale computation can rely on different balances of local and recurrent mechanisms (Khajehabdollahi et al., 2024). Broader effective timescales and recurrent realignment accompanied these effects, though they were not shown to be causal; placement thus acts as an initial temporal bias that recurrent learning can partly offset.

The cortical results provide supportive but more limited evidence. MEG neural timescales were modestly associated with slow connectome structure, and their ordering broadened the temporal repertoire of the connectome model. This complements links between activity timescales and connectivity strength (Fallon et al., 2020) and timescale heterogeneity in homogeneous connectome reservoirs (Milisav et al., 2026). Whereas Kim et al. (2025) inferred regional decay rates and related them to slow modes, we directly manipulate placement of a fixed timescale set and test its dynamical consequences. However, MEG timescales reflect regional activity rather than local decay constants (Shafiei et al., 2023; Zeraati et al., 2026), mapping their ordering onto model time constants is therefore an assumption, and group-average maps limit individual-level inference. Taken together, our experiments distinguish which timescales are available, how they are placed relative to recurrent structure, and how recurrent plasticity adapts to that placement. The cortical findings are compatible with this organization, and the model experiments establish its computational and learning effects.

AI USE STATEMENT

In this work, we used generative AI tools to design or provide feedback on research methodology or experiments, implement methods, and interpret results. We have not used generative AI tools to propose or refine hypotheses, generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, assist with translation, clean and reformat dataset, or support qualitative and thematic data analysis. Additionally, we used generative AI tools to create or edit software code, identify relevant literature, and edit a research paper to improve readability. All AI-assisted work was reviewed by the authors. Literature identified with AI assistance was checked against the original sources before citation, and AI-generated or modified code was reviewed and tested. The authors made all final scientific and methodological decisions and take responsibility for the content of this work.

REPRODUCIBILITY STATEMENT

Definitions of the initial diagnostic tasks and training details are given in Appendix A; the placement protocol and slow mode coordinate are detailed in Appendices B-C; the fixed recurrence behavioral and learning experiments are described in Appendices D and E; and the human connectome analyses are detailed in Appendix F. The accompanying code is included in the supplementary material as `timescale_placement.zip`.

REFERENCES

Subati Abulikemu, Puria Radmard, Michail Mamalakis, and John Suckling. Association-sensory spatiotemporal hierarchy and functional gradient-regularised recurrent neural network with implications for schizophrenia. *npj Systems Biology and Applications*, 12(1):100, 2026. doi: 10.1038/s41540-026-00727-x. URL https://www.nature.com/articles/s41540-026-00727-x.

Aaron F. Alexander-Bloch, Haochang Shou, Siyuan Liu, Theodore D. Satterthwaite, David C. Glahn, Russell T. Shinohara, Simon N. Vandekar, and Armin Raznahan. On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage*, 178:540–551, 2018. doi: 10.1016/j.neuroimage.2018.05.070. URL https://www.sciencedirect.com/science/article/pii/S1053811918304968.

Manuel Beiran and Srdjan Ostojic. Contrasting the effects of adaptation and synaptic filtering on the timescales of dynamics in recurrent networks. *PLoS Computational Biology*, 15(3): e1006893, 2019. doi: 10.1371/journal.pcbi.1006893. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006893.

Alberto Bernacchia, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, 14(3):366–372, 2011. doi: 10.1038/nn.2752. URL https://www.nature.com/articles/nn.2752.

Richard F. Betzel, Alessandra Griffa, Patric Hagmann, and Bratislav Mišić. Distance-dependent consensus thresholds for generating group-representative structural brain networks. *Network Neuroscience*, 3(2):475–496, 2019. doi: 10.1162/netn_a_00075. URL https://doi.org/10.1162/netn_a_00075.

Sean E. Cavanagh, Joni D. Wallis, Steven W. Kennerley, and Laurence T. Hunt. Autocorrelation structure at rest predicts value correlates of single neurons during reward-guided choice. *eLife*, 5:e18937, 2016. doi: 10.7554/eLife.18937. URL https://elifesciences.org/articles/18937.

Rishidev Chaudhuri, Alberto Bernacchia, and Xiao-Jing Wang. A diversity of localized timescales in network activity. *eLife*, 3:e01239, 2014. doi: 10.7554/eLife.01239. URL https://elifesciences.org/articles/01239.

Rishidev Chaudhuri, Kenneth Knoblauch, Marie-Alice Gariel, Henry Kennedy, and Xiao-Jing Wang. A large-scale circuit mechanism for hierarchical dynamical processing in the primate

cortex. *Neuron*, 88(2):419–431, 2015. doi: 10.1016/j.neuron.2015.09.008. URL https://www.sciencedirect.com/science/article/pii/S0896627315007655.

Joni Dambre, David Verstraeten, Benjamin Schrauwen, and Serge Massar. Information processing capacity of dynamical systems. *Scientific Reports*, 2:514, 2012. doi: 10.1038/srep00514. URL https://www.nature.com/articles/srep00514.

Murat Demirtaş, Joshua B. Burt, Markus Helmer, Jie Lisa Ji, Brendan D. Adkinson, Matthew F. Glasser, David C. Van Essen, Stamatios N. Sotiropoulos, Alan Anticevic, and John D. Murray. Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron*, 101 (6):1181–1194.e13, 2019. doi: 10.1016/j.neuron.2019.01.017. URL https://doi.org/10.1016/j.neuron.2019.01.017.

Laura N. Driscoll, Krishna Shenoy, and David Sussillo. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nature Neuroscience*, 27:1349–1363, 2024. doi: 10.1038/s41593-024-01668-6. URL https://www.nature.com/articles/s41593-024-01668-6.

John Fallon, Phillip G. D. Ward, Linden Parkes, Stuart Oldham, Aurina Arnatkevičiūtė, Alex Fornito, and Ben D. Fulcher. Timescales of spontaneous fMRI fluctuations relate to structural connectivity in the brain. *Network Neuroscience*, 4(3):788–806, 2020. doi: 10.1162/netn_a_00151. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7888482/.

David J. Freedman and John A. Assad. Experience-dependent representation of visual categories in parietal cortex. *Nature*, 443:85–88, 2006. doi: 10.1038/nature05078. URL https://www.nature.com/articles/nature05078.

Surya Ganguli, Dongsung Huh, and Haim Sompolinsky. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48):18970–18975, 2008. doi: 10.1073/pnas.0804451105. URL https://www.pnas.org/doi/10.1073/pnas.0804451105.

Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, Mark Jenkinson, and WU-Minn HCP Consortium. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124, 2013. doi: 10.1016/j.neuroimage.2013.04.127. URL https://doi.org/10.1016/j.neuroimage.2013.04.127.

Hidehiko K. Inagaki, Lorenzo Fontolan, Sandro Romani, and Karel Svoboda. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature*, 566:212–217, 2019. doi: 10.1038/s41586-019-0919-7. URL https://www.nature.com/articles/s41586-019-0919-7.

Sina Khajehabdollahi, Roxana Zeraati, Emmanouil Giannakakis, Tim Jakob Schäfer, Georg Martius, and Anna Levina. Emergent mechanisms for long timescales depend on training curriculum and affect performance in memory tasks. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=xwKt6bUkXj.

Jason Z. Kim, Richard F. Betzel, Ahmad Beyh, Amber Howell, Amy Kuceyeski, Bart Larsen, Caio Seguin, Xi-Han Zhang, Avram Holmes, and Linden Parkes. Inferring intrinsic neural timescales using optimal control theory. *Nature Communications*, 16(1):11639, 2025. doi: 10.1038/s41467-025-66542-w. URL https://www.nature.com/articles/s41467-025-66542-w.

Jan Koutnik, Klaus Greff, Faustino Gomez, and Juergen Schmidhuber. A Clockwork RNN. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 1863–1871. PMLR, 2014. URL https://proceedings.mlr.press/v32/koutnik14.html.

Tomoki Kurikawa and Kunihiko Kaneko. Fluctuation-learning relationship in recurrent neural networks. *Nature Communications*, 16(1):9663, 2025. doi: 10.1038/s41467-025-64976-w. URL https://www.nature.com/articles/s41467-025-64976-w.

Filip Milisav, Andrea I. Luppi, Laura E. Suárez, Guillaume Lajoie, and Bratislav Misic. Neuromorphic hierarchical modular reservoirs. *Nature Communications*, 17:7962, 2026. doi: 10.1038/s41467-026-74466-2. URL https://www.nature.com/articles/s41467-026-74466-2.

Filippo Moro, Pau Vilimelis Aceituno, Laura Kriener, and Melika Payvand. Temporal hierarchy in spiking neural networks. *Neuromorphic Computing and Engineering*, 6:024023, 2026. doi: 10.1088/2634-4386/ae7ab5. URL https://doi.org/10.1088/2634-4386/ae7ab5.

John D. Murray, Alberto Bernacchia, David J. Freedman, Ranulfo Romo, Jonathan D. Wallis, Xinying Cai, Camillo Padoa-Schioppa, Tatiana Pasternak, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, 17(12): 1661–1663, 2014. doi: 10.1038/nn.3862. URL https://www.nature.com/articles/nn.3862.

Bo-yong Park, Reinder Vos de Wael, Casey Paquola, Sara Larivière, Oualid Benkarim, Jessica Royer, Shahin Tavakol, Raul R. Cruces, Qiongling Li, Sofie L. Valk, Daniel S. Margulies, Bratislav Mišić, Danilo Bzdok, Jonathan Smallwood, and Boris C. Bernhardt. Signal diffusion along connectome gradients and inter-hub routing differentially contribute to dynamic human brain function. *NeuroImage*, 224:117429, 2021. doi: 10.1016/j.neuroimage.2020.117429. URL https://doi.org/10.1016/j.neuroimage.2020.117429.

Nicolas Perez-Nieves, Vincent C. H. Leung, Pier Luigi Dragotti, and Dan F. M. Goodman. Neural heterogeneity promotes robust learning. *Nature Communications*, 12(1):5791, 2021. doi: 10.1038/s41467-021-26022-3. URL https://www.nature.com/articles/s41467-021-26022-3.

Silvan C. Quax, Michele D'Asaro, and Marcel A. J. van Gerven. Adaptive time scales in recurrent neural networks. *Scientific Reports*, 10(1):11360, 2020. doi: 10.1038/s41598-020-68169-x. URL https://www.nature.com/articles/s41598-020-68169-x.

Alexander Schaefer, Ru Kong, Evan M. Gordon, Timothy O. Laumann, Xi-Nian Zuo, Avram J. Holmes, Simon B. Eickhoff, and B. T. Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9):3095–3114, 2018. doi: 10.1093/cercor/bhx179. URL https://doi.org/10.1093/cercor/bhx179.

Sarab S. Sethi, Valerio Zerbi, Nicole Wenderoth, Alex Fornito, and Ben D. Fulcher. Structural connectome topology relates to regional BOLD signal dynamics in the mouse brain. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(4):047405, 2017. doi: 10.1063/1.4979281. URL https://pubs.aip.org/aip/cha/article/27/4/047405/322545/Structural-connectome-topology-relates-to-regional.

Golia Shafiei, Ben D. Fulcher, Bradley Voytek, Theodore D. Satterthwaite, Sylvain Baillet, and Bratislav Misic. Neurophysiological signatures of cortical micro-architecture. *Nature Communications*, 14:6000, 2023. doi: 10.1038/s41467-023-41689-6. URL https://www.nature.com/articles/s41467-023-41689-6.

Merav Stern, Nicolae Istrate, and Luca Mazzucato. A reservoir of timescales emerges in recurrent circuits with heterogeneous neural assemblies. *eLife*, 12:e86552, 2023. doi: 10.7554/eLife.86552. URL https://elifesciences.org/articles/86552.

Gouhei Tanaka, Tadayoshi Matsumori, Hiroaki Yoshida, and Kazuyuki Aihara. Reservoir computing with diverse timescales for prediction of multiscale dynamics. *Physical Review Research*, 4 (3):L032014, 2022. doi: 10.1103/PhysRevResearch.4.L032014. URL https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.4.L032014.

Evelyn Tang, Harang Ju, Graham L. Baum, David R. Roalf, Theodore D. Satterthwaite, Fabio Pasqualetti, and Danielle S. Bassett. Control of brain network dynamics across diverse scales of space and time. *Physical Review E*, 101(6):062301, 2020. doi: 10.1103/PhysRevE.101.062301. URL https://doi.org/10.1103/PhysRevE.101.062301.

Paul Triebkorn, Viktor Jirsa, and Peter Ford Dominey. Simulating the impact of white matter connectivity on processing time scales using brain network models. *Communications Biology*, 8(1):197, 2025. doi: 10.1038/s42003-025-07587-x. URL https://www.nature.com/articles/s42003-025-07587-x.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E. J. Behrens, Essa Yacoub, Kamil Ugurbil, and WU-Minn HCP Consortium. The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80:62–79, 2013. doi: 10.1016/j.neuroimage.2013.05.041. URL https://doi.org/10.1016/j.neuroimage.2013.05.041.

D. F. Wasmuht, E. Spaak, T. J. Buschman, E. K. Miller, and M. G. Stokes. Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nature Communications*, 9 (1):3499, 2018. doi: 10.1038/s41467-018-05961-4. URL https://doi.org/10.1038/s41467-018-05961-4.

Yuichi Yamashita and Jun Tani. Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. *PLoS Computational Biology*, 4(11): e1000220, 2008. doi: 10.1371/journal.pcbi.1000220. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000220.

Guangyu Robert Yang, Madhura R. Joglekar, H. Francis Song, William T. Newsome, and Xiao-Jing Wang. Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, 22:297–306, 2019. doi: 10.1038/s41593-018-0310-2. URL https://www.nature.com/articles/s41593-018-0310-2.

Marco Zenari, Luca Taffarello, Luca Mazzucato, Amos Maritan, and Samir Suweis. Topological origin of the diversity of timescales in recurrent neural circuits. *ArXiv*, abs/2603.04149, 2026. URL https://arxiv.org/abs/2603.04149.

Roxana Zeraati, Yan-Liang Shi, Nicholas A. Steinmetz, Marc A. Gieselmann, Alexander Thiele, Tirin Moore, Anna Levina, and Tatiana A. Engel. Intrinsic timescales in the visual cortex change with selective attention and reflect spatial connectivity. *Nature Communications*, 14:1858, 2023. doi: 10.1038/s41467-023-37613-7. URL https://www.nature.com/articles/s41467-023-37613-7.

Roxana Zeraati, Anna Levina, Jakob H. Macke, and Richard Gao. Neural timescales from a computational perspective. *Nature Neuroscience*, 29(7):1534–1547, 2026. doi: 10.1038/s41593-026-02343-8. URL https://www.nature.com/articles/s41593-026-02343-8.

A DIAGNOSTIC TASKS AND TRAINING DETAILS

Diagnostic tasks The first experiment uses two diagnostic tasks, T1 and T2. In T1, two independent latent components each follow a stationary AR(1) process,

$$z_k(t) = \rho_k z_k(t-1) + \sqrt{1-\rho_k^2}\,\epsilon_k(t), \qquad \rho_k = e^{-1/T_k}, \tag{4}$$

where $\epsilon_k(t) \sim \mathcal{N}(0,1)$ and $z_t = (z_1(t), z_2(t))^\top$. The latent state z_t is not observed directly; the network receives

$$x_t = M z_t + \eta_t, \tag{5}$$

where M is orthogonal and $\eta_t \sim \mathcal{N}(0, I_2)$ is independent observation noise. The network is trained to reconstruct z_t. We keep the geometric mean of the two latent timescales fixed at 8 and vary their separation, so that $T_{\text{fast}} = 8/\sqrt{s}$ and $T_{\text{slow}} = 8\sqrt{s}$, where s is the timescale ratio. We test $s \in \{1, 2, 4, 8, 16\}$, and the comparison across model configurations (non-recurrent leaky units, reservoir, and RNN) uses the hardest condition, $s = 16$. Performance is evaluated after a 32-step burn-in.

For T2, the latent signals are three independent white noise sequences,

$$u_t \sim \mathcal{N}(0, I_3), \tag{6}$$

with $u_t = (u_1(t), u_2(t), u_3(t))^\top$. The network observes only the noisy orthogonal mixture,

$$x_t = M u_t + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, I_3), \tag{7}$$

and is trained to recover each latent signal after a specified delay,

$$y_k(t) = u_k(t - d_k). \tag{8}$$

Thus, T1 tests whether the network can unmix latent signals with different timescales, and T2 tests whether it can retain multiple latent signals over a delay. The comparison across model configurations uses $(d_1, d_2, d_3) = (4, 8, 16)$, and T2 performance is evaluated after a burn-in of $\max_k d_k + 16$ steps.

Training and model configuration Throughout these diagnostic experiments, time is expressed in model steps ($\Delta t = 1$). Models here use 64 hidden units and sequence length 256, and are trained for 2000 steps with batch size 128 using Adam with learning rate 10^{-3}, no weight decay, and gradient norm clipping at 1, with five paired model seeds. Homogeneous models learn one globally shared τ, and heterogeneous models learn separate τ_i for each unit. Learned timescales are initialized around 8 and constrained to $[0.5, 128]$, and the checkpoint with the lowest validation error is used for held-out evaluation. Training uses mean squared error and all target channels are weighted equally.

The reservoir and RNN configurations use the same heavy-tail graph construction with density 0.15, and the initial masked recurrent matrix is rescaled to spectral radius 0.9. Recurrent weights are fixed in the reservoir and optimized under the same fixed graph mask in the RNN; the readout and intrinsic timescales are learned in both.

Performance normalization Performance is normalized relative to a memoryless baseline and an analytic causal reference,

$$G = \frac{E_{\text{base}} - E_{\text{model}}}{E_{\text{base}} - E_{\text{ref}}}, \tag{9}$$

so that $G = 0$ corresponds to memoryless performance and $G = 1$ to the analytic reference. For T1, E_{base} is the error of the Bayes estimate $\mathbb{E}[z_t \mid x_t]$, which uses just the current observation, and E_{ref} is the error of the exact causal Kalman estimate $\mathbb{E}[z_t \mid x_{0:t}]$, which uses the full observation history. For T2, E_{base} is the error of predicting the prior mean of zero, and E_{ref} is the error of the minimum mean squared error estimate obtained from the corresponding delayed observation. Heterogeneity effects are shown as $\Delta G = G_{\text{heterogeneous}} - G_{\text{homogeneous}}$. Reported 95% confidence intervals are percentile bootstrap intervals from $20{,}000$ resamples over the five paired model seeds.

To test if the T2 heterogeneity advantage depends on having different delay demands across channels, we hold the maximum delay at 8 and compare $(8, 8, 8)$, $(3, 5, 8)$, $(2, 4, 8)$, and $(1, 3, 8)$. Heterogeneous leaky units remain better than homogeneous leaky units even in the equal delay condition

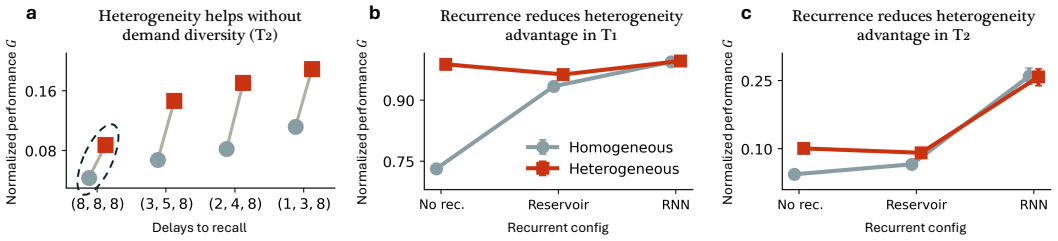


Figure 7: **Heterogeneity remains beneficial without demand diversity; recurrence reduces the heterogeneity advantage.** **a** T2 performance for different target delay combinations. Heterogeneous leaky units outperform homogeneous leaky units, including when all three channels have the same delay, $(8, 8, 8)$. **b** Normalized performance across leaky units, reservoir, and RNN configurations on T1 at $s = 16$. **c** The same comparison on T2 with $(d_1, d_2, d_3) = (4, 8, 16)$. In both tasks, homogeneous performance approaches heterogeneous performance with recurrence added.

(Fig. 7a). Normalized performance across model configurations shows that homogeneous performance approaches heterogeneous performance with recurrence added in both T1 and T2 (Fig. 7b,c).

Temporal representation diagnostics We examine temporal representations in the trained leaky units in T2 with $d_1 = d_2 = d_3 = 8$. For a unit with intrinsic timescale τ_i, its temporal kernel is

$$k_i(\ell) = \alpha_i(1 - \alpha_i)^\ell, \qquad \alpha_i = 1 - e^{-1/\tau_i}. \tag{10}$$

Collecting these kernels as columns of K, we define the delay capacity $C_K(d)$ as how well their span represents an input at delay d, obtained by projecting the corresponding unit impulse e_d,

$$\hat{e}_d = KK^+e_d, \qquad C_K(d) = \|\hat{e}_d\|_2^2. \tag{11}$$

Because $\|e_d\|_2 = 1$, larger $C_K(d)$ indicates better representation of that delay.

We also quantify how much past input remains recoverable from the state of the frozen trained network. Each trained model is driven by independent Gaussian white noise $r_t \sim \mathcal{N}(0, I_3)$. After a 64 step burn-in, a readout is fitted to recover each input channel at lags 0 through 64 and evaluated on independent sequences. The input memory at lag d is

$$m(d) = \frac{1}{3}\sum_{c=1}^{3} \mathrm{corr}^2(\hat{r}_{t-d,c}, r_{t-d,c})\,. \tag{12}$$

The readout is fitted on 64 sequences of length 320 and evaluated on 48 independent sequences. Three white noise realizations are averaged within each trained model, with the five paired model seeds serving as the unit of replication. Thus, the delay capacity $C_K(d)$ measures delay representation by the learned temporal kernels, and the input memory $m(d)$ measures recoverable information about past inputs.

Effective timescale diagnostic For the effective timescale comparison in Fig. 2d, the T2 $(4, 8, 16)$ checkpoints from each model configuration are driven by the same three Gaussian white noise input batches of 48 sequences each. After the 32 step T2 burn-in, each unit's effective timescale τ_i^{net} is estimated from an exponential fit to its hidden state autocorrelation. Per-unit estimates are geometrically averaged among the input batches before computing $S_{\mathrm{eff}} = \mathrm{std}_i[\log \tau_i^{\mathrm{net}}]$.

B Placement Experiment Details

Placement experiments use 64 unit reservoirs with fixed recurrent weights and step time units ($\Delta t = 1$). Erdős-Rényi (ER) and heavy-tail graphs have exactly the same off-diagonal edge budget, corresponding to density 0.15. ER edges are sampled uniformly, and heavy-tail graphs use a bounded heterogeneous degree distribution. Recurrent weights are obtained by masking an orthogonally initialized matrix and rescaling the masked recurrent matrix to spectral radius 0.9. We use 20 independently generated graph realizations and two recurrent weight initializations per graph, which

gives 40 fixed recurrent network instances per topology. Every instance receives the same fixed set of 64 intrinsic timescales, log-spaced from 1 to 64.

Placement alignment is quantified by Spearman correlation between intrinsic timescale and a node coordinate (degree, strength, or slow mode coordinate; App. C). We test target alignments $-1,\ -2/3,\ -1/3,\ 0,\ 1/3,\ 2/3,\ 1$. At the aligned regime ($\rho_s = 1$), the slowest intrinsic timescales are assigned to the highest scoring nodes; the assignment is reversed at the anti-aligned regime. Intermediate placements are generated by permuting the same timescale set and, from 256 candidate permutations, selecting the assignment whose achieved Spearman correlation is closest to the target. Two placement replicates are evaluated at each alignment level. Placement construction depends only on the fixed node coordinate and achieved alignment; effective dynamics are not used to select an assignment.

After each placement, the reservoir is driven by the same fixed four channel Gaussian white noise. This contains 8 sequences of length 384, with a 96 step burn-in. Effective timescales are estimated from unit activity autocorrelations using lags up to 64, and the resulting temporal repertoire is quantified by S_{eff} as defined in the main text. Thus, graph, recurrent weights, intrinsic timescale set, and external drive are all fixed within each network instance and only the nodal assignment of timescales changes. Confidence intervals for the placement effects use 4000 hierarchical bootstrap resamples over graph realizations and recurrent weight initializations.

C SLOW MODE COORDINATE

We compare two local structural coordinates with a recurrent dynamics based coordinate. The structural coordinates are off-diagonal binary degree and off-diagonal in-strength (sum of absolute incoming weights). Motivated by the eigenvector localization view of Chaudhuri et al. (2014), the slow mode coordinate quantifies where slowly decaying recurrent modes are expressed across nodes.

For each fixed recurrent substrate, we construct a homogeneous reference system in which every unit has the same intrinsic timescale $\tau_0 = 8$. The reference system retains the recurrent substrate but replaces the heterogeneous timescale assignment by this common value. It is driven by a fixed four channel Gaussian white noise reference input that is independent of the input used to measure S_{eff}. Using 8 sequences of length 384 and a 96 step burn-in, we estimate how often each ReLU unit is active and form the reference Jacobian

$$J_0 = (1 - \alpha_0)I + \alpha_0 \bar{D} W_{\text{rec}}, \qquad \alpha_0 = 1 - e^{-1/\tau_0}, \tag{13}$$

where $\bar{D}_{ii}$ is the fraction of post burn-in samples for which unit i has a positive ReLU preactivation.

We then decompose J_0 into recurrent modes. Modes with eigenvalues closer to the unit circle decay more slowly. Each node is assigned a score from its normalized squared loading on each right eigenvector, with slower modes receiving greater weight,

$$s_i = \sum_k w_k \frac{|v_{ik}|^2}{\|v_k\|_2^2}, \qquad w_k \propto -\frac{1}{\log|\lambda_k|}, \tag{14}$$

where λ_k and v_k are the eigenvalue and right eigenvector of mode k. The resulting slow mode coordinate measures the spatial expression of slow recurrent modes across nodes. It is computed once from the homogeneous reference system before heterogeneous intrinsic timescales are assigned and is then kept fixed across all tested placements.

D TASKED PLACEMENT ACROSS TEMPORAL DEMAND

Network construction and calibration The fixed recurrence behavioral study uses 128-unit leaky-ReLU reservoirs with $\Delta t = 20$ ms. All topology families have the same off-diagonal edge budget and density 0.15. ER and heavy-tail graph constructions are as above; modular graphs contain four modules, with 0.10 of off-diagonal edges assigned between modules while preserving the same total edge count. Recurrent weights are obtained by masking an orthogonally initialized matrix with the graph and rescaling the resulting matrix to spectral radius 0.9. Input weights use Xavier initialization

and biases are zero. Every network receives the same fixed set of 128 intrinsic timescales, log-spaced from 40 to 640 ms with geometric mean 160 ms.

Calibration and confirmation were separated. The intrinsic timescale center and span were selected using behavioral viability under random placement. Ridge regularization and task difficulty were also calibrated using random placements only. All timescale, readout, and task settings were frozen before the aligned versus anti-aligned confirmation analyses.

Cognitive tasks MemoryPro and MemoryAnti use continuous circular stimuli and responses following delay-go and delay-anti task families used in multitask RNN studies (Yang et al., 2019; Driscoll et al., 2024). In MemoryPro, the network maintains a stimulus direction over a delay and responds in that direction; in MemoryAnti, the response is displaced by π. We test delays of $200, 400, 800, 1200, 1600, 2000$, and 2400 ms. Performance is measured as mean angular error.

Perceptual decision making (PDM) requires integrating noisy binary evidence, maintaining the resulting decision through a variable delay, and reporting a binary choice (Inagaki et al., 2019). Each trial contains a 1200 ms evidence period followed by a delay of $0, 200, 400, 800, 1200, 1600$, or 2400 ms. DelayMatchCategory requires retaining the category of a sample direction over a delay and reporting whether a subsequent test direction belongs to the same category (Freedman & Assad, 2006). Sample and test directions are drawn relative to two circular categories, with delays of $200, 400, 800, 1200, 1600, 2000$, and 2400 ms. Performance on both binary tasks is reported as percentage error, $100(1 - \text{balanced accuracy})$.

For each network, latent task variables and stochastic input streams are generated once and reused across placement conditions. Demand levels are paired; longer delays thus extend the same stochastic input stream instead of introducing independent noise. A separate ridge readout is fitted for each placement using the same regularization and matched training data, and evaluated on matched held-out trials. The graph, recurrent and input weights, biases, timescale set, trial variables, and input streams are otherwise fixed.

Placement confirmation and inference Confirmation uses 16 independently generated networks from each of the ER, heavy-tail, and modular families. For each network, placement is varied with respect to either the slow mode coordinate or degree. We test seven target Spearman alignments, $-1,\ -2/3,\ -1/3,\ 0,\ 1/3,\ 2/3,\ 1$. The aligned and anti-aligned placements use the exact rank ordering of the timescale set. At each intermediate alignment, three independently generated permutations are selected to lie within 0.03 of the target correlation and averaged before inference.

The primary placement contrast is

$$B(d) = E_{\text{anti}}(d) - E_{\text{aligned}}(d), \tag{15}$$

where E is angular error for MemoryPro and MemoryAnti or percentage error for PDM and DelayMatchCategory, so $B(d) > 0$ indicates an aligned placement benefit. Means and 95% confidence intervals use 5000 paired bootstrap resamples over networks, and significance is assessed with paired sign-flip tests. Effects are reported separately for each topology family.

Effective temporal repertoire and behavioral benefit For each network and demand level, we computed the partial Spearman correlation between S_{eff} and task performance across the seven slow

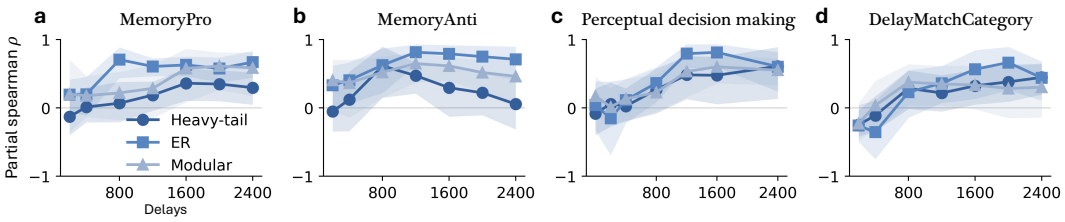


Figure 8: **Association between effective temporal repertoire and behavioral performance across temporal demand.** Partial Spearman correlations between S_{eff} and task performance across the seven slow mode alignment levels for MemoryPro, MemoryAnti, PDM, and DelayMatchCategory, respectively, controlling for mean effective network timescale.

mode alignment levels, controlling for mean effective network timescale. Network-level correlations were summarized within each topology family, with 95% confidence intervals obtained by bootstrap resampling over networks. The full demand-dependent trajectories are shown in Fig. 8.

Network size robustness We tested whether the slow mode placement effect depended on reservoir size using 64, 128, and 256 unit networks, with the network construction and task, readout, and placement protocols unchanged. Size robustness was evaluated at 1600 ms rather than the maximal 2400 ms delay because the 64-unit networks approached chance performance at 2400 ms under both placements. Aligned placement produced a positive behavioral benefit in all 36 task, topology, and network size combinations (Fig. 9). Confidence intervals were entirely above zero in 8 of 12 comparisons for 64-unit networks and in all comparisons for 128 and 256 unit networks, suggesting that the placement effect was not specific to the 128-unit setting.

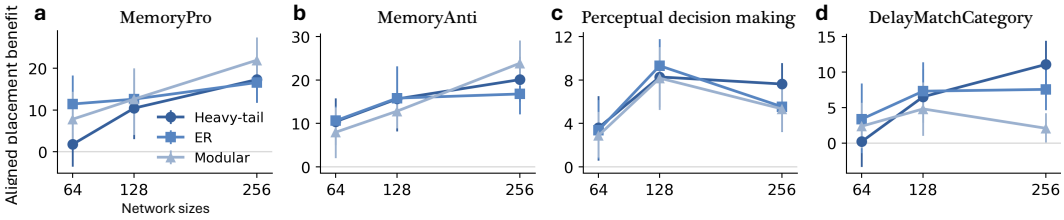


Figure 9: **Slow mode placement benefit across network size.** Aligned placement benefit $B(1600)$ for 64, 128, and 256 unit networks across MemoryPro, MemoryAnti, PDM, and DelayMatchCategory. Points show mean paired placement effects across networks and error bars show 95% bootstrap confidence intervals.

E Learning, recurrent plasticity, and dynamical reorganization

Experimental design and placement The learning study uses the same 128-unit network construction, topology families, timestep, and 40-640 ms intrinsic timescale set as the fixed recurrence task study. We generated eight independent networks from each topology family (24 networks total) and trained MemoryPro at delays of 400, 800, 1600, and 2400 ms. Each network is evaluated with two placement rules (slow mode and binary degree), two target alignments (-0.5 and $+0.5$), and two recurrence conditions (fixed and trainable), which gives 768 runs. A placement is constructed once for each network, rule, and sign and reused across all four delays.

Slow mode placements target the prescribed Spearman correlation while constraining absolute correlations with degree and weighted in-strength to at most 0.10, along with a rank quintile balance constraint. Degree placements target the same ± 0.5 correlations with degree but are not constrained with respect to the slow mode coordinate, as the two coordinates covary in some topology families. Achieved correlations with slow mode, degree, and weighted in-strength are retained for the secondary coordinate analysis. Placement construction depends only on the recurrent network and fixed intrinsic timescales, not on task performance.

Training and pairing The output readout is trained in every condition. In trainable recurrence, recurrent weights are optimized along with the readout (retaining the original graph mask); in fixed recurrence, recurrent weights are unchanged. No spectral radius projection or recurrent gain renormalization is applied during learning. Optimization uses Adam for 2000 steps with batch size 32, recurrent learning rate 3×10^{-4}, readout learning rate 3×10^{-2}, no weight decay, and gradient norm clipping at 100. Performance is evaluated on 160 held-out trials at steps 0, 50, 100, 200, 300, 400, 600, 800, 1200, 1600, and 2000. Within each network and delay, compared placement and recurrence conditions use matched readout initialization, training batches, and evaluation trials.

Learning efficiency and coordinate analysis Early learning efficiency is quantified from held-out error over the first 600 optimization steps. Baseline adjusted early learning gain A summarizes the average reduction in held-out error from its pre-training value across this window, integrated over the recorded checkpoints. The first 600 steps capture the early, pre-convergence phase of learning; larger values indicate faster improvement. For the primary analysis, we consider slow mode placements

with trainable recurrence. Within each network and delay, $\Delta A(d) = A_{+0.5}(d) - A_{-0.5}(d)$. A least-squares slope of ΔA against $\log_2 d$ is computed across the four delays within each network, so each network contributes one estimate of how the aligned learning advantage changes with temporal demand.

We perform a secondary analysis using the achieved slow mode and degree correlations across all trainable recurrence conditions. For each of the 96 conditions (24 networks $\times$ four rule/sign combinations), we compute the slope of early learning gain A across $\log_2 d$, and fit

$$S_{na} = \gamma_n + \beta_{\text{slow}}\rho_{na}^{\text{slow}} + \beta_{\text{deg}}\rho_{na}^{\text{deg}} + \epsilon_{na}, \tag{16}$$

where S_{na} is the demand slope for condition a in network n, γ_n is a network fixed effect, and the predictors are the achieved slow mode and degree correlations. Confidence intervals and two-sided tests use 5000 topology stratified cluster bootstrap resamples over networks. As a topology sensitivity analysis, we refit the model after omitting each topology family in turn. The slow mode coefficient remained positive when omitting heavy-tail ($\beta_{\text{slow}} = 0.045\,[0.018, 0.072]$), modular ($0.045\,[0.014, 0.072]$), or ER networks ($0.059\,[0.031, 0.088]$).

Plastic compensation and recurrent reorganization Compensation is assessed at the maximal 2400 ms delay using slow mode placement. Late error is the mean held-out error over checkpoints 1200, 1600, and 2000. Within each network, we compare the aligned versus anti-aligned late performance gap under fixed and trainable recurrence; a smaller gap with trainable recurrence indicates compensation for unfavorable initial placement.

For monitoring the evolving slow recurrent structure during learning, we construct a homogeneous timescale reference Jacobian from the current W_{rec} at steps 0, 400, 800, 1200, 1600, and 2000. For longitudinal comparison, we track the slow invariant subspace rather than individual modes, avoiding the need to match modes across training checkpoints. Complex Schur decomposition is used to identify the invariant subspace associated with the 20% of recurrent modes having the largest eigenvalue magnitudes. Each node is scored by its mean squared participation in this slow subspace, and its Spearman correlation with the fixed heterogeneous τ_i assignment defines slow subspace alignment. In initially anti-aligned networks at 2400 ms, reorganization is quantified as the change in slow subspace alignment from step 0 to step 2000. This quantity is unchanged in fixed recurrence by construction, providing a reference for the trainable recurrence condition. Confidence intervals are obtained from 5000 bootstrap resamples over networks, and directional comparisons use one-sided Wilcoxon signed-rank tests.

F HUMAN CONNECTOME ANALYSIS DETAILS

Data and connectome construction The structural connectome and MEG timescale map were obtained from the processed data release from Milisav et al. (2026). The connectome was derived from diffusion-weighted MRI in 327 unrelated Human Connectome Project (HCP) participants using the HCP minimal preprocessing pipeline (Glasser et al., 2013) and tractography-based structural network reconstruction; further details are described in Park et al. (2021). Networks were represented in the 400-parcel Schaefer atlas (Schaefer et al., 2018). A distance-dependent consensus procedure was used to construct the group network, preserving mean density and the edge length distribution (Betzel et al., 2019); retained edge weights were averaged across participants, log transformed, and rescaled to $[0, 1]$ (Milisav et al., 2026). The group-average MEG timescale map was from 33 unrelated HCP participants (Shafiei et al., 2023) and was parcellated to the same Schaefer atlas by Milisav et al. (2026). The structural and MEG data are therefore group-level representations, not subject-matched measurements.

Slow mode score and spatial association The connectome is spectrally normalized with gain $g = 0.9$, and the homogeneous reference timescale is fixed at $\tau_0 = 8$. For the reference system, the eigenvalues are

$$\mu_k = (1 - \alpha_0) + \alpha_0\lambda_k(W), \tag{17}$$

with mode decay times

$$T_k = -\frac{1}{\log|\mu_k|}. \tag{18}$$

The parcelwise slow mode score is

$$s_i = \sum_k q_k v_{ik}^2, \qquad q_k = \frac{T_k}{\sum_j T_j}, \tag{19}$$

where v_{ik} is the loading of parcel i on mode k. Association between the MEG timescale map and s_i is quantified using Spearman correlation. Spatial significance is assessed using $10{,}000$ spherical spin permutations of the cortical map, which preserve its spatial autocorrelation (Alexander-Bloch et al., 2018). The two-sided spatial p-value is computed from the absolute null correlations using a $+1$ permutation correction.

Dynamical placement analysis For the dynamical placement analysis, only the rank ordering of the empirical MEG map is retained and mapped onto a fixed set of 400 model timescales log-spaced from 1 to 64. We use $2{,}000$ of the same spatial rotations to define alternative placements, so that the structural connectome and complete timescale set remain fixed and only their cortical assignment changes. For each placement,

$$J = I - D_\alpha + D_\alpha W, \tag{20}$$

where $D_\alpha = \mathrm{diag}(\alpha_i)$ and $\alpha_i = 1 - e^{-1/\tau_i}$. Under Gaussian white noise input, stationary node autocorrelations are computed analytically for this linear system, and effective timescales are estimated from these autocorrelations using lags up to 64. The resulting temporal repertoire is quantified by S_{eff} as defined in the main text. The empirical ordering is evaluated relative to the rotated null using its percentile and a one-sided upper tail permutation test with a $+1$ correction. Across rotated placements, slow mode alignment is quantified by the Spearman correlation between assigned model timescales and the slow mode score, and its association with S_{eff} is assessed using Spearman correlation.