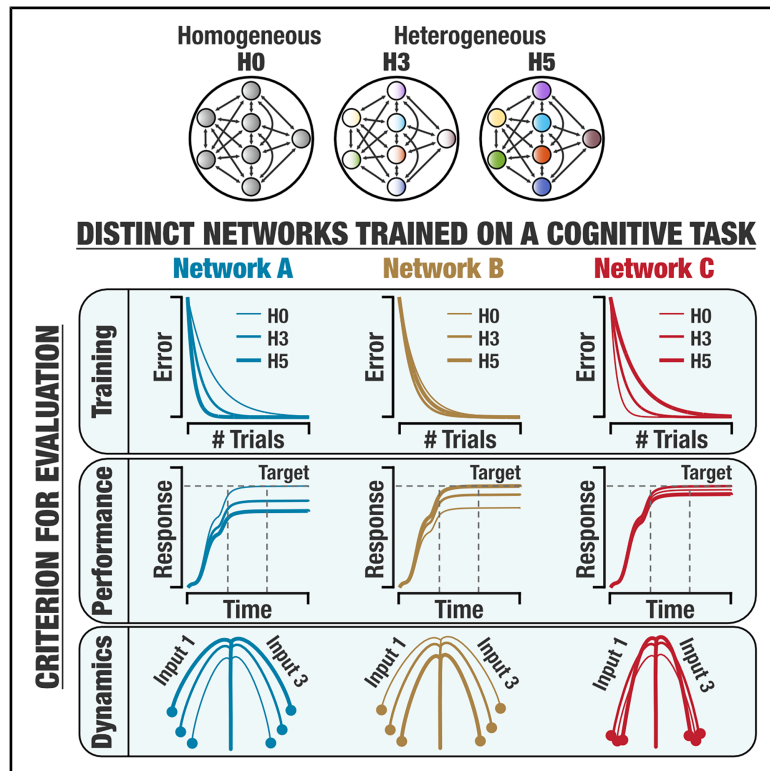


Patterns

Diversity in the impact of heterogeneities on recurrent networks performing a cognitive task

Graphical abstract



Authors

Anjana Santhosh,
Rishikesh Narayanan

Correspondence

rishi@iisc.ac.in

In brief

No two neurons in the brain are identical. In this study, Santhosh and Narayanan show that differences between neurons affect learning, task performance, activity patterns, and resistance to disruption in highly varied ways. Using many recurrent neural networks trained on the same cognitive task, they found that the effects depended on network hyperparameters, the amount and type of variation, and how performance was measured. Their results highlight the importance of studying neural networks as complete, interacting systems across many examples, rather than focusing on individual differences in isolation.

Highlights

- Heterogeneities in neural circuits are not universally beneficial or detrimental
- Form and degree of heterogeneity, hyperparameters, and metrics define their impact
- Performance robustness depends on network, heterogeneities, and perturbation type
- Studying neural-circuit heterogeneities requires a complex systems framework

Article

Diversity in the impact of heterogeneities on recurrent networks performing a cognitive task

Anjana Santhosh^{1,2} and Rishikesh Narayanan^{1,2,3,4,*}

¹Cellular Neurophysiology Laboratory, Molecular Biophysics Unit, Indian Institute of Science, Bengaluru 560012, India

²Interdisciplinary Doctoral Programme on Brain, Computation, and Data Science, Indian Institute of Science, Bengaluru 560012, India

³Senior author

⁴Lead contact

*Correspondence: rishi@iisc.ac.in

<https://doi.org/10.1016/j.patter.2026.101685>

THE BIGGER PICTURE Neurons in the brain have different intrinsic properties, connectivity, and responses to experience. Yet most neural network models used to study cognition are built from repeating, identical units. This study looks at whether differences between neurons are beneficial or detrimental to the brain's ability to learn and process information. This is done by systematically introducing different degrees of unit variation into several recurrent neural networks trained to perform a cognitive task. The same type of variation improved learning in one network but reduced performance in another. The outcome also depended on the network's properties and on the way performance was measured. The variation observed indicates that a network can solve a task in different ways and helps explain why a system can continue to function even when its individual parts differ or are disturbed. It also means that conclusions drawn from a single network may not apply more broadly. The findings in this study suggest that differences between neurons should not be studied as isolated features or reduced to a simple comparison between "different" and "identical" networks. Instead, they should be considered as part of a larger system in which several forms of variations interact. Learning allows a network to settle on one of many possible combinations of components that can achieve the same goal. To understand how brains and brain-inspired machines learn and perform tasks, researchers need to carefully study how different components work together, adapt through learning, and produce reliable behavior as a group.

SUMMARY

Artificial recurrent networks are widely used to model the complex dynamics of biological neural circuits performing cognitive tasks. Yet, most studies assume homogeneous units, ignoring functionally important neural heterogeneities. We present a systematic framework to quantify the impact of heterogeneities on recurrent networks performing a cognitive task. We trained a population of networks, differing in hyperparameters, with six graded levels of intrinsic heterogeneities. Our analyses unveiled a surprising diversity in the impact of intrinsic heterogeneities on training, task-execution dynamics, and robustness, despite identical training conditions. First, the dependence of training performance on heterogeneity levels showed pronounced network-to-network variability. Second, task-execution dynamics exhibited marked variability in their dependence on heterogeneities. Finally, post-training perturbations strongly affected robustness, manifesting substantial network-specific variability. This striking diversity in the impact of heterogeneities underscores the need to consider global structure, interactions, and adaptations across multiple heterogeneities to understand collective function in complex neural circuits.

INTRODUCTION

Artificial recurrent networks are widely used as models to study the complex dynamics underlying biological neural networks during execution of cognitive tasks.^{1–7} However, most studies as-

sume individual units in the recurrent network to be homogeneous repeating units, an assumption that is fundamentally in conflict with biological neural networks where individual neurons manifest several forms of heterogeneities.^{8–19} There are several lines of evidence that these heterogeneities are not just chance occurrences

but have diverse functional impact on different aspects of neural and network function.¹⁹ A diversity of such impact has been reported, in providing alternate routes to implement specific functional outcomes,^{20–28} efficient coding,^{29–36} spatial navigation,^{37–40} decorrelation^{23,41,42} and pattern separation,⁴³ disruption of function in certain types of networks,^{44,45} introducing plasticity heterogeneity toward efficient resource allocation,^{46–51} modulating synchrony in networks,^{44,52–56} promoting stability of learning in spiking neural networks,⁵⁷ and mediating robustness to perturbations.^{25,58–65} Despite the several lines of evidence on the diverse roles of heterogeneities on brain function, the impact of neural heterogeneities on recurrent networks trained to perform a cognitive task has not been systematically explored.

In this study, we develop and employ a framework to systematically study the impact of neural heterogeneities on circuit function. Our unified framework specifically identifies several lacunae in existing methodologies to study the impact of heterogeneities on neural learning systems and presents a systematic process to address each of these lacunae.

1. **Observation:** learning in biological systems occurs in the presence of heterogeneities.

Lacuna: in many approaches, heterogeneities are not introduced during the training process, but instead are studied in hand-tuned networks.

Our framework: neural heterogeneities must be introduced into artificial networks during training, when the network learns the specific task structure to achieve effective performance.⁵⁷

2. **Observation:** biological heterogeneities manifest in different forms at different degrees.^{8–19,66}

Lacuna: in many approaches, the expression of heterogeneities is treated to be binary, assessing heterogeneous vs. homogeneous networks. They do not explicitly account for graded levels of different forms of heterogeneities.

Our framework: different forms of heterogeneities must be introduced into the network at multiple levels during the training process, specifically designed to assess the graded impact of different heterogeneities on task performance.

3. **Observation:** biological networks manifest network-to-network variability.^{20,25,67–72}

Lacuna: many approaches use a single hand-tuned network with specific hyperparameters to study the impact of heterogeneities. This approach biases outcomes to the manual choice of hyperparameters made for that network, both in terms of network performance and its dependence on heterogeneities.

Our framework: a population of networks with very different hyperparameters^{22,25,43,73} must be assessed across all levels of heterogeneities.

4. **Observation:** biological networks and their several components are in a continuous state of flux owing to external perturbations, neuromodulation, homeostatic computations, or the need to learn additional tasks.^{25,58,63,67,74}

Lacuna: the impact of post-training perturbations on network performance is rarely tested in the context of studying the impact of neural-circuit heterogeneities. Even if done, the analyses do not account for different levels of disparate perturbations.

Our framework: robustness of network performance to different forms and several levels of post-training perturbations must be assessed across all networks.

5. **Observation:** biological systems typically manifest trade-offs between multiple competing functional objectives, characterized as Pareto optimality.^{75–77}

Lacuna: many approaches assess the impact of heterogeneities focusing on a single functional outcome.

Our framework: several quantitative metrics must be used for evaluating different aspects of network performance: training performance, task-execution dynamics, and robustness to post-training perturbations. The impact of heterogeneities on all these quantitative metrics must be collectively studied across different levels of heterogeneities and several graded post-training perturbations.

In implementing this broad framework, we explored the impact of intrinsic heterogeneities in artificial recurrent networks trained to perform a cognitive task. In explicitly adhering to the framework above, our experimental design incorporated the following (presented in the same sequence as above):

1. We trained recurrent networks using a modified reward-modulated Hebbian learning mechanism⁷⁸ to perform a cognitive task in the presence of intrinsic heterogeneities in unit time constants. We focus on the only intrinsic component in the units in our network, the unit time constant. Therefore, synaptic heterogeneities emerged through the training process in networks endowed with intrinsic heterogeneities.
2. All networks were trained with 5 different gradations of intrinsic heterogeneities (H1–H5) and were compared to their homogeneous counterparts (H0) trained in the absence of intrinsic heterogeneities.
3. A population of 53 artificial recurrent networks endowed with distinct hyperparameters was trained for all 6 gradations of intrinsic heterogeneities, yielding a total of 318 networks. Additionally, in understanding the impact of individual hyperparameters, networks with distinct combinations of 3 instances of 5 different hyperparameters were trained across 6 different intrinsic heterogeneities, providing an additional 1,458 networks.
4. The impact of 7 post-training perturbations (mean of unit time constants, range of unit time constants, mean of initial activity, range of initial activity, recurrent synaptic weight jitter, stimulus epoch duration, and frequency of exploratory activity impulses) on network performance was tested. The robustness to each of the 7 forms of perturbations was tested at 6 different grades (P0–P5) across all 318 networks.
5. Multiple metrics were used to assess the impact of neural heterogeneities on different aspects of network performance. We assessed training performance using the number of training trials required for convergence and the error dynamics through the training process. Network performance was assessed using errors incurred in performing the task across multiple trials of different input patterns. Task-execution dynamics was assessed using distances of network trajectories in the latent space. Robustness

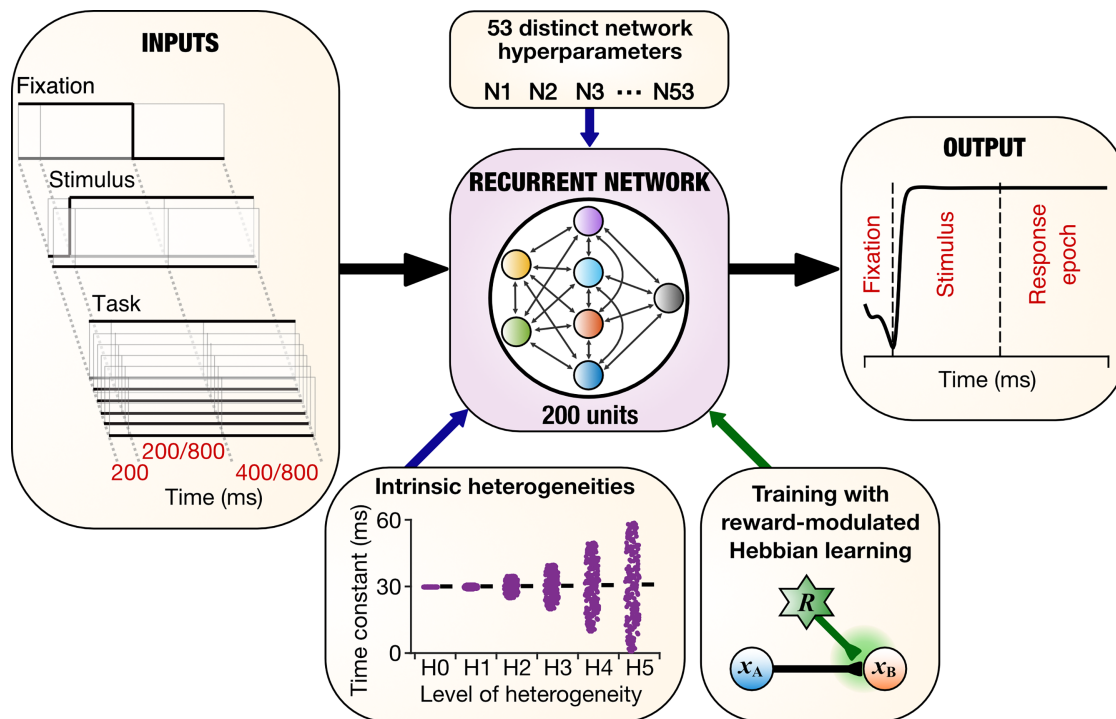


Figure 1. A population of recurrent networks with distinct hyperparameters trained and tested with different levels of heterogeneities to perform a cognitive task

A rate-based fully connected recurrent network of 200 units was initialized with 53 different parametric configurations, each trained and tested with six levels (H0–H5; H0 refers to the homogeneous network) of intrinsic heterogeneities. The total number of networks assessed is $53 \times 6 = 318$. Progressively higher levels of intrinsic heterogeneities (H1–H5) were achieved by increasing the range of randomly initialized time constants of the units in the network. The network under consideration receives three types of inputs: a fixation input that turns off when the network is expected to provide its output, two inputs corresponding to the stimuli for the task being performed, and 6 distinct inputs that indicate the specific task the network is expected to perform. All networks in this study were trained and tested for the memoryless Go task, which is indicated on the “Inputs” tab. Each trial of the Go task consists of three epochs with variable duration (fixation epoch, fixed at 200 ms; stimulus epoch, switching between 200 and 800 ms; and response epoch, alternating between 400 and 800 ms). The fixation input stays on through the fixation and the stimulus epochs, while the stimulus input stays on through the stimulus and the response epochs. A single unit was designated as the output unit of the network whose response was considered as the network’s output. The network was trained using a modified reward-modulated Hebbian learning rule. For the Go task, the expected output is to reflect the active input value (either +1 or –1) assigned to either of the two stimuli (which are never simultaneously active). One of the two stimuli is always inactive and is set to 0.

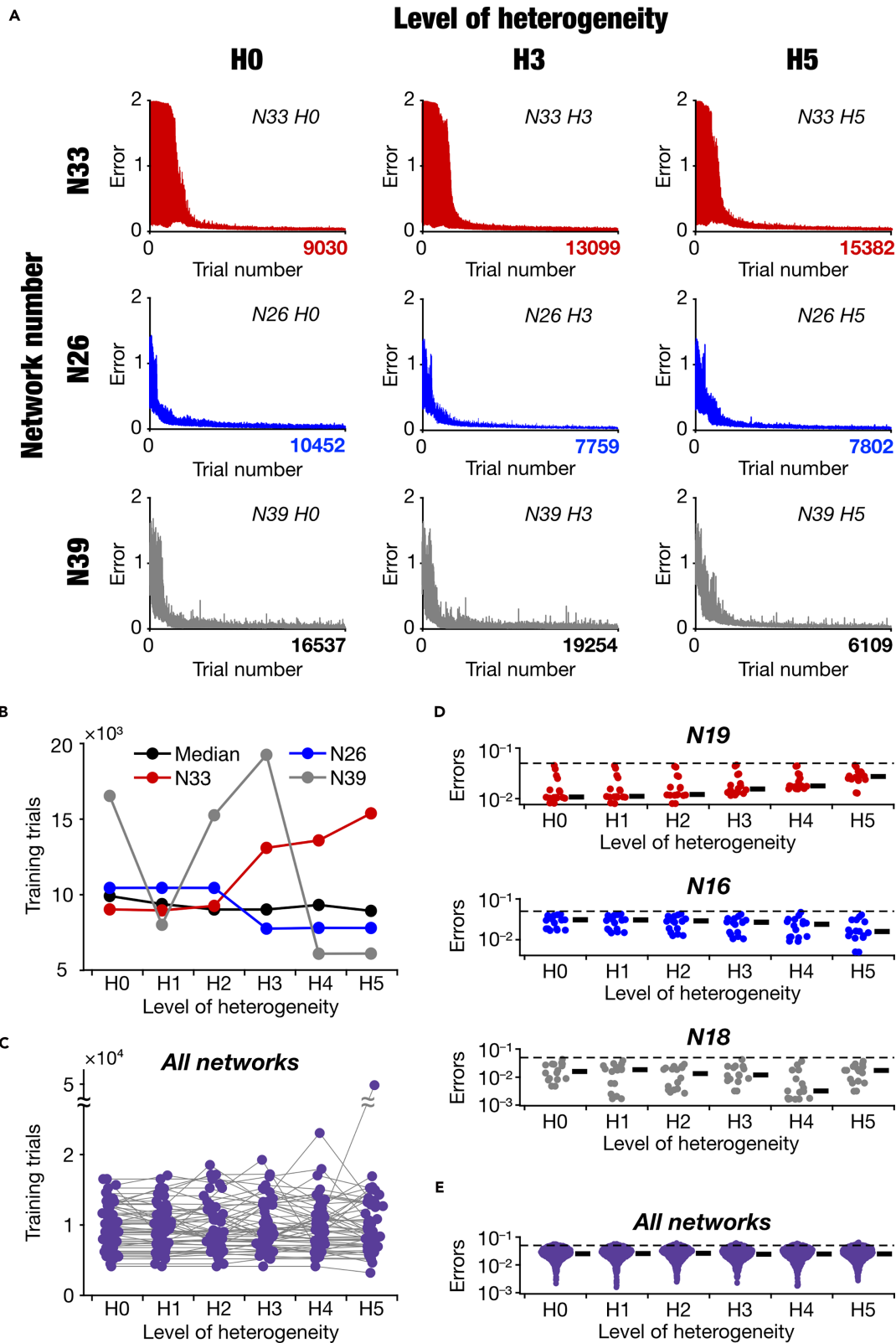
to 7 different forms of graded post-training perturbations in each of the trained networks was assessed using task performance errors and latent-space distances.

Our analyses unveiled pronounced network-to-network variability in the impact of heterogeneities on training, task performance, network dynamics during task execution, and robustness to post-training perturbations. Based on our analyses, we propose a complex systems viewpoint to study the impact of neural heterogeneities on circuit function, where the focus is not on heterogeneities in individual components. Instead, we emphasize the need to focus on the global structure of different forms and degrees of heterogeneities across different components and how they interact and adapt toward achieving collective function. Our results emphasize that disparate non-random combinations of components, hyperparameters, and forms/degrees of heterogeneities could yield similar circuit function, together highlighting functional degeneracy as a design principle that defines network function. As each network picks one of several such possible solutions toward learning-specific tasks, different networks manifest substantial network-to-network vari-

ability in their performance measures and in their robustness to different forms and degrees of perturbations.

RESULTS

The principal goal of this study was to test the impact of neural heterogeneities on a recurrent network trained to perform a cognitive task. We used a fully connected, continuous-time recurrent network with 200 units (Figure 1). The recurrent network was trained using a modified reward-modulated Hebbian learning algorithm⁷⁸ to perform a simple memoryless cognitive task Go,¹ where the network output is expected to reflect the input value assigned to either of the two stimuli (Figure 1). Our experimental design toward assessing the impact of heterogeneities on recurrent networks performing this cognitive task involved important methodological constraints. First, we did not binarize heterogeneous vs. homogeneous networks. Instead, we introduced heterogeneities in a graded fashion in six different ranges in unit time constants (Figure 1). Such gradation accounts for variability in the extent of biological heterogeneities observed under different



(legend on next page)

conditions. More importantly, this experimental design involving different levels of heterogeneities (H0–H5) allowed us to assess progressive changes in network performance and dynamics with increasing levels of intrinsic heterogeneities.

Second, to avoid biases introduced by the use of a single network for all analyses, we used a population-of-networks approach throughout our analyses to quantify potential network-to-network variability in training performance and response dynamics. Specifically, we employed 53 instantiations of networks, which were different in terms of all hyperparameters that defined the initialization and training of these networks. Together, this yielded 318 different networks (53 parameter initializations, each with 6 levels of heterogeneity) that were trained individually to perform the cognitive task. The impact of intrinsic heterogeneities on training performance and task-execution dynamics was tested for all 53 networks, spanning all 6 levels of heterogeneities.

Third and finally, biological networks are sequentially learning new representations and undergo changes due to random perturbations. The changes introduced post-learning could alter the intrinsic properties of the neurons, the synaptic connections, the network activity, or the level of noise in the system, in varying degrees. To account for these post-learning changes, we introduce several forms of perturbations to our trained network. Our experimental design involved six graded levels, each of several forms of post-training perturbations (referred to as P0–P5, to distinguish from H0–H5 introduced during training). For each form of post-training perturbation, we studied the impact of the 6 levels of post-training perturbations on all 318 networks. The experimental design and the associated analyses allowed us to quantitatively assess (1) the extent to which networks were robust to each kind of post-training perturbations, (2) network-to-network variability in how networks react to different forms of post-training perturbations, (3) the impact of pre-training heterogeneities (H0–H5) on how robust the network was to each form of post-training perturbations (P0–P5), and (4) network-to-network variability in the impact of pre-training heterogeneities on network robustness to post-training perturbations.

In implementing this broad experimental design, we first trained each of the 318 networks to perform the Go task till each network reached the convergence criterion, whereby the absolute error must be below 0.05 for 100 consecutive training trials (more details on the task, the heterogeneities, and the training and testing process are provided in the [methods](#) section). The number of trials required for convergence and the errors across testing trials were noted for each network.

Pronounced network-to-network variability in the impact of heterogeneities on training performance

All 318 networks successfully reached the set convergence criterion toward executing the Go task. However, the training trajectories manifested pronounced variability across different networks and across heterogeneity levels ([Figures 2A–2C](#)). First, even in the absence of intrinsic heterogeneities (H0 group), there were large differences in the number of trials required for convergence. Whereas some networks converged with as low as 4,000 trials, others required close to 17,000 trials for achieving convergence ([Figure 2C](#)). Second, the impact of progressively increasing the level of heterogeneities (from H0 to H5) increased the number of required training trials in some networks (e.g., N33), decreased in others (e.g., N26), while had no such structure (e.g., N39) in a third set of networks ([Figures 2A and 2B](#)). Similarly, the error values observed in test trials showed pronounced trial-to-trial variability for any given level of heterogeneity ([Figure 2D](#)), pronounced network-to-network variability in the error values, as well as dependency on heterogeneities ([Figures 2D and 2E](#)). Specifically, for some networks, the median error increased with the level of heterogeneity (e.g., N19), for others the median error decreased (e.g., N16), while a third class of networks did not exhibit any monotonic trend (e.g., N18) ([Figure 2D](#)).

For one of the networks (N1), we also tested the impact of intrinsic heterogeneities on networks that were trained with different parameters associated with the network and the training process ([Figure S1](#)). The network was trained in the presence of 6 levels of heterogeneities, with a lower learning rate ($\eta = 0.05$), lower connection density (connection density = 50%), or a more stringent convergence criterion (error below 0.05 for 400 consecutive training trials). With increasing level of heterogeneities, the number of training trials showed an increasing trend in networks trained with a reduced learning rate as well as in the network with a reduced connection density, but the trend reversed in the network that was trained with a stringent convergence criterion ([Figure S1A](#)). Similarly, we observed variable dependencies of test trial errors on the level of heterogeneities, depending on which parameter was changed in the network ([Figure S1B](#)). These results demonstrated that the variability in dependence of training performance on the level of heterogeneities extended to other training hyperparameters as well.

Together, these results demonstrated pronounced variability across networks in terms of the number of training trials required for learning ([Figures 2A–2C](#)), the error values associated with specific trials ([Figures 2D and 2E](#)), connectivity/training

Figure 2. Pronounced network-to-network variability in training performance of heterogeneous recurrent networks trained to perform a cognitive task

- (A) The evolution of absolute error during training for three different initializations (rows) and different levels of intrinsic heterogeneity (columns). The x axis ranges from 0 to N_T , the total number of trials required for the specific network to converge. It may be noted that the number of training trials required for convergence increased (N33), decreased (N26), or varied randomly (N39) with increasing levels of heterogeneity.
- (B) Number of trials taken to train the three example networks shown in (A) plotted as a function of the level of training heterogeneity, depicting network-to-network variability in the impact of heterogeneities on training. The black line depicts the median number of trials across all 53 networks.
- (C) Pronounced network-to-network variability in the number of trials required for convergence in all 53 networks, each trained across the 6 heterogeneity levels (H0–H5).
- (D) Distribution of errors in the 16 different testing trials (of the Go task, with different stimuli as well as different durations for the stimulus and the response epochs) plotted across H0–H5 for three example trained networks (dotted lines represent the 0.05 level used as a convergence requirement). Network N19 showed an increase in error with increasing heterogeneity, network N16 manifested a reduction, while network N18 did not show any monotonic pattern. Black dashes indicate median errors for each heterogeneity level.
- (E) Distribution of errors in all trials for all 318 trained networks. The black dashes represent the median error for each level of heterogeneity.

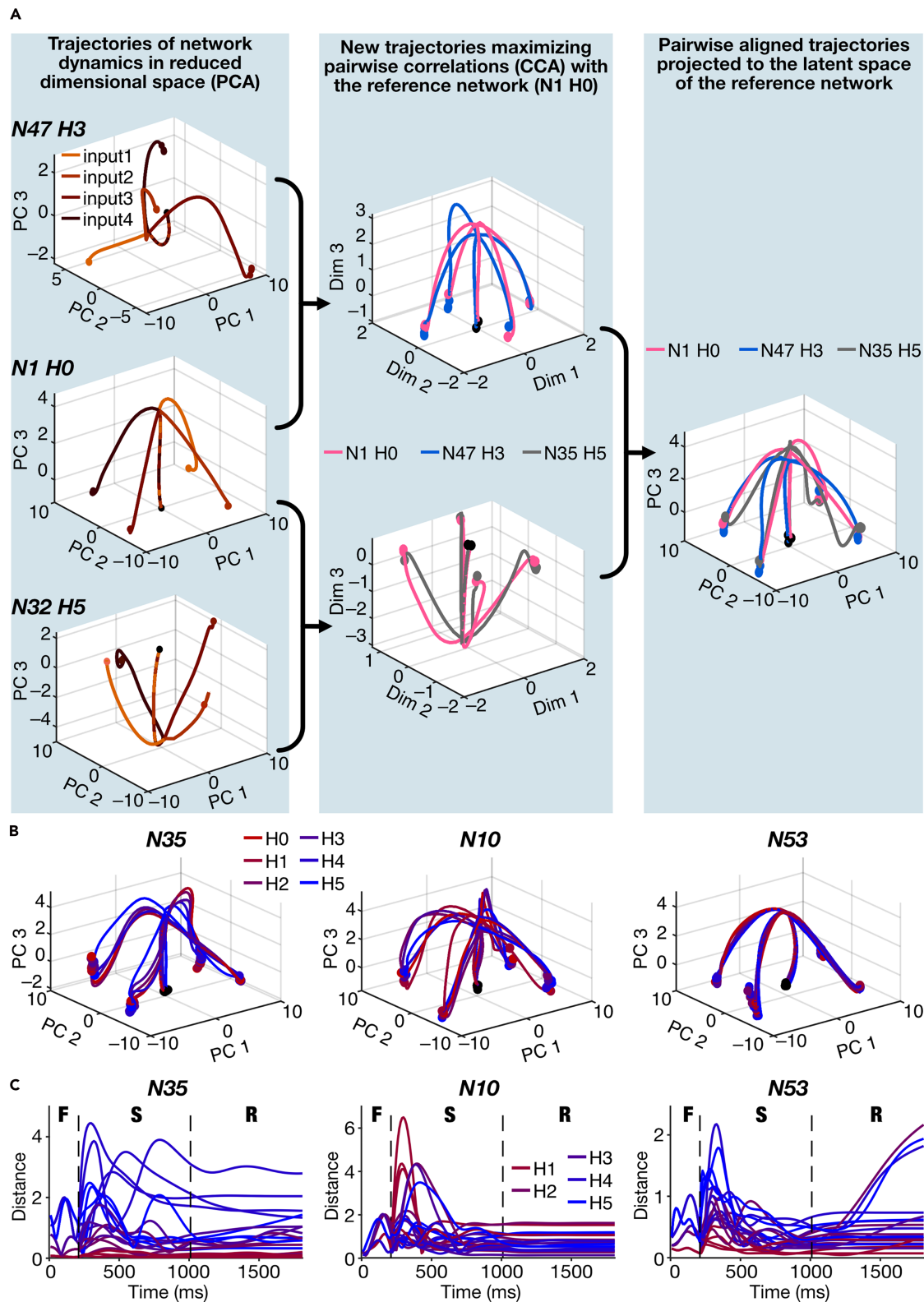


Figure 3. Pronounced network-to-network variability in latent dynamics of heterogeneous recurrent networks trained to perform a cognitive task

(A) Left: a step-by-step representation of the process of aligning latent-space dynamics of multiple networks. The dynamics of all 200 units in the network are represented by a trajectory in the respective reduced dimensional spaces computed for each network using principal component analysis (PCA). Center: using

(legend continued on next page)

hyperparameters (Figure S1), and the dependency of trial numbers (Figures 2A–2C) and error values (Figures 2D and 2E) on the level of intrinsic heterogeneities introduced during training.

Dynamics of heterogeneous networks during task performance varied across network initializations

We performed principal component analysis (PCA) on the activity dynamics of all recurrent units to visualize and assess the task-associated dynamics of the trained recurrent networks. We repeated this for all 53 networks for each of the 6 different levels of heterogeneities. As the dimensionality of the activity dynamics (Equation 10) was around 3 for this task across all networks, we used a 3-dimensional representation for visualizing the dynamics of these networks (Figure 3A). In this visualization, the initial time point ($t = 0$) is represented by a black dot. The trajectory in the reduced dimensional space followed an identical route until the stimulus arrived (Figure 3A). Upon stimulus arrival at the beginning of the stimulus epoch, the trajectory diverged based on the specific input stimuli and eventually converged to one of the four stable points, each defining a specific input pattern ($\mathbf{u}_{stim} \in \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$). We found the latent-space dynamics to be distinct across different levels of heterogeneity for the same network (Figure S2) as well as across different networks for the same level of heterogeneity (Figure S3). To compare trajectories across different networks and across different levels of heterogeneities, we first aligned the latent spaces of the different networks using canonical correlation analysis (CCA) in a pairwise manner using a reference network N1H0 (Figure 3A). Finally, the aligned trajectories were transformed along the principal components of N1H0, thus representing all the network trajectories in the same latent space (Figure 3A).

When the same network with different levels of heterogeneities was compared, we found variability in task-related trajectories as well as the specific location of the input-specific fixed points in the aligned latent space (Figure 3B). While some networks manifested pronounced variability in the dynamics across different levels of heterogeneities (e.g., N35 and N10), there were networks (e.g., N53) where the trajectories were less diverse across different levels of heterogeneities (Figure 3B). We computed the Euclidean distance between the latent-space trajectories of the heterogeneous H1–H5 networks and that of the homogeneous H0 network for each of the 53 network initializations. We plotted these distances as functions of time for different networks with different levels of heterogeneities (Figure 3C). These distance computations quantitatively confirmed the visual observation that the variability in trajectories was higher in N35 and N10, while the distances between trajectories were comparatively lower for N53 (Figure 3C).

Importantly, in certain networks, the distances from the H0 trajectories increased with increasing level of heterogeneities (e.g.,

N35), while in others the sequence reversed (e.g., N10). In addition, the deviation in the fixed points (the final distance value at the end of the trial) depended on the level of heterogeneities and the type of input presented (e.g., N35 and N53). Together, these trajectories provided examples for the pronounced network-to-network variability observed in the impact of training heterogeneities on network dynamics during task execution (Figure 3C). These examples unveiling network-to-network variability in trajectories and their dependence on the levels of heterogeneities further emphasize the need for using multiple networks and multiple levels of heterogeneities in assessing the impact of heterogeneities on neural circuits.

We assessed the distances of all heterogeneous networks (H1–H5) with their homogeneous counterpart (H0) for each of the different durations of the stimulus and response epochs across different input configurations (Figure 4). Across different durations, we found network-to-network variability in latent-space trajectories to initially increase and then decrease during the fixation epoch. We observed maximum variability in trajectories during the period that followed stimulus presentation (Figure 4). Eventually, depending on how long the stimulus epoch was, variability in trajectories across heterogeneities reduced either during the response epoch (Figures 4A and 4B) or during the stimulus epoch (Figures 4C and 4D). Across durations, the variability was highest during a ~ 300 – 400 ms duration, after which the trajectories converged to their respective fixed points with relatively minimal variability. Importantly, despite the pronounced network-to-network variability in how the distance values depended on the level of training heterogeneity (Figures 3C and 4E), we found monotonic increases in the median values of the maximum distance, of the average distance, and of the variance of distances computed across the 53 networks analyzed here (Figure 4E).

Together, these observations demonstrated that training heterogeneities in intrinsic properties impacted network dynamics in diverse ways even if they were trained with the same training algorithms, convergence criteria, and task specifications. The impact of training heterogeneities on network dynamics during task execution manifested pronounced variability in training performance (Figure 2) as well as in task-execution dynamics (Figures 3 and 4) across networks initialized with distinct sets of hyperparameters.

Interdependencies of network hyperparameters on heterogeneous network training performance and dynamics

The 53 instantiations of networks analyzed thus far were generated by independently sampling all of the five major hyperparameters (seed values defining the initial recurrent weights $\mathbf{J}$, the initial input weights $\mathbf{B}$, the initial activity patterns $\mathbf{x}_0$, the

canonical correlation analysis (CCA), the latent-space dynamics of each network and that of a reference network were transformed to a new coordinate space by maximizing the correlation between the trajectories in the new space in a pairwise manner. Right: the aligned trajectories in each CCA space were then projected back along the latent PCA space of the reference network bringing all the networks to a common reduced dimensional space, which enabled distance comparisons.

(B) The aligned latent-space dynamics of three example networks, each shown for all 6 different levels of heterogeneity.

(C) Distance between trajectories, plotted as a function of time, of heterogeneous networks (H1–H5) with respect to the homogeneous network (H0) trajectory in the latent space.

F, fixation epoch; S, stimulus epoch; R, response epoch.

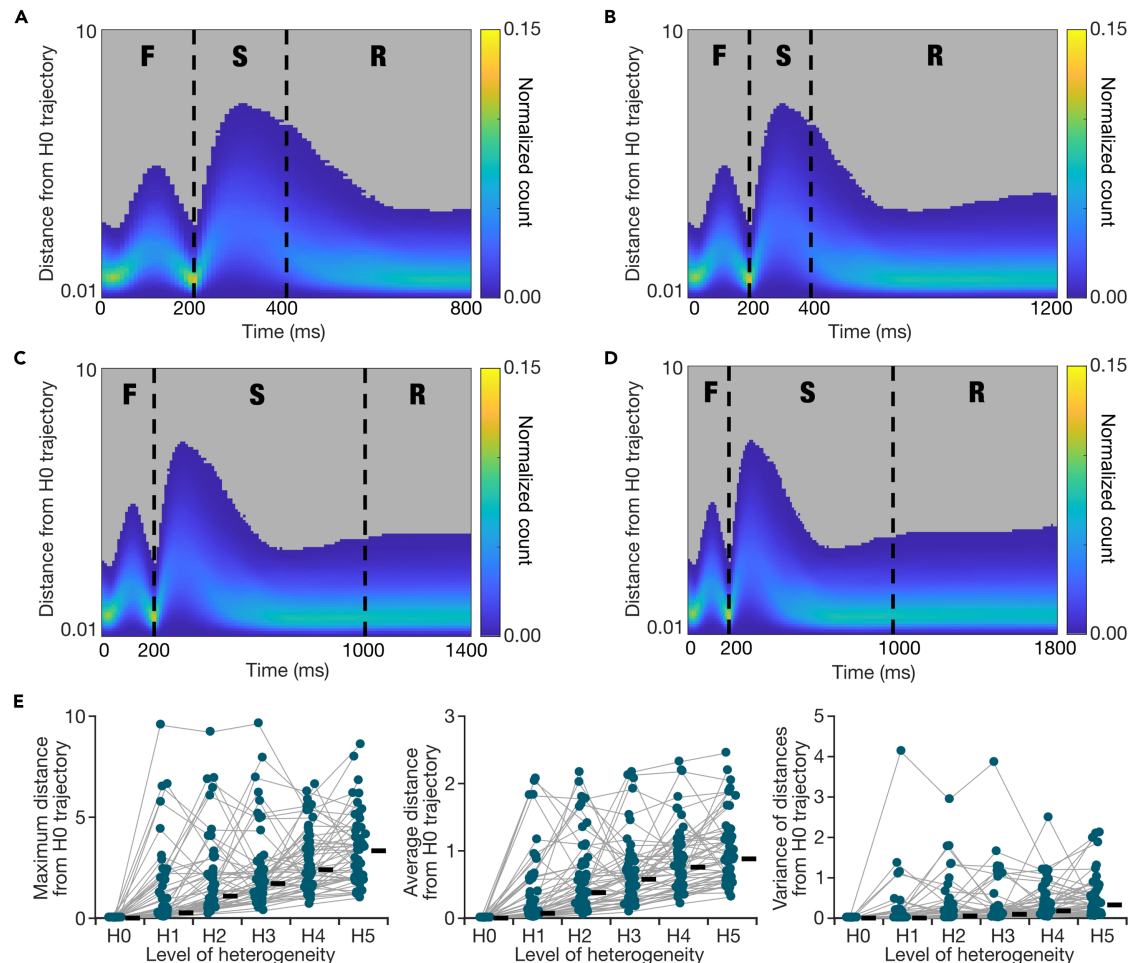


Figure 4. Network-to-network variability in latent dynamics of heterogeneous recurrent networks trained to perform a cognitive task across different epochs

(A–D) Heatmaps of the distribution of Euclidean distances between all pairs of aligned latent-space trajectories for all 318 networks. The fixation epoch remained constant at 200 ms for all panels. The stimulus epoch was 200 ms (A and B) or 800 ms (C and D). The response epoch was either 400 ms (A and C) or 800 ms (B and D). The variability in trajectories across networks especially during the stimulus period might be noted. F, fixation epoch; S, stimulus epoch; R, response epoch. (E) Distribution of maximum (left), average (middle), and variance (right) of latent-space distances with respect to the corresponding H0 network trajectory, calculated across all the 16 trials and plotted against the level of heterogeneity. Gray lines connect the distance values for the same network across different heterogeneity levels. Black bars represent the median values.

exploratory activity impulses Δ , and time constant values τ). Thus, no two networks shared any of the five major hyperparameters. The absence of common hyperparameters implies that the impact of individual hyperparameters on network performance cannot be assessed from these networks. In addition, the interdependencies of the effects of individual hyperparameters across each other and on the training level of heterogeneities remains unknown. To understand cross-dependencies across different levels of training heterogeneities and individual network hyperparameters, we picked three independent seed values for each of the five major hyperparameters (Table S1). We initialized networks with all $3^5 = 243$ possible combinations of these seed values and trained each of these networks with six levels of heterogeneity. Together, this resulted in a set of $3^5 \times 6 = 1,458$ networks spanning all hyperparameters and heterogeneity levels (Figure 5). Notations-wise, the J3-B2- x_0 3- Δ 1- τ 3-H2 represented the network where J, B, x_0 , Δ , and τ were

initialized with seed values J3, B2, x_0 3, Δ 1, and τ 3 (Table S1), respectively, and trained with H2 level of heterogeneity.

Among the 1,458 networks, the number of training trials (required for convergence) showed pronounced network-to-network variability, spanning values as low as 4,086 to as high as 58,531 (Figure 5A). Among the different hyperparameters, we found strong dependencies of the median number of trials on the hyperparameters governing recurrent and input weights, with relatively minor dependencies on the other hyperparameters. Specific hyperparameter values triggered larger variability in the distribution of required training trials. For instance, some (but not all) networks initialized with J3, x_0 2, and B2 required longer training periods than the others (Figures 5A, 5D, and 5E), maximizing network-to-network variability and emphasizing cross-dependencies across different hyperparameters and heterogeneities. As with the 53 distinct networks tested earlier (Figures 2B and 2C), the median value of the

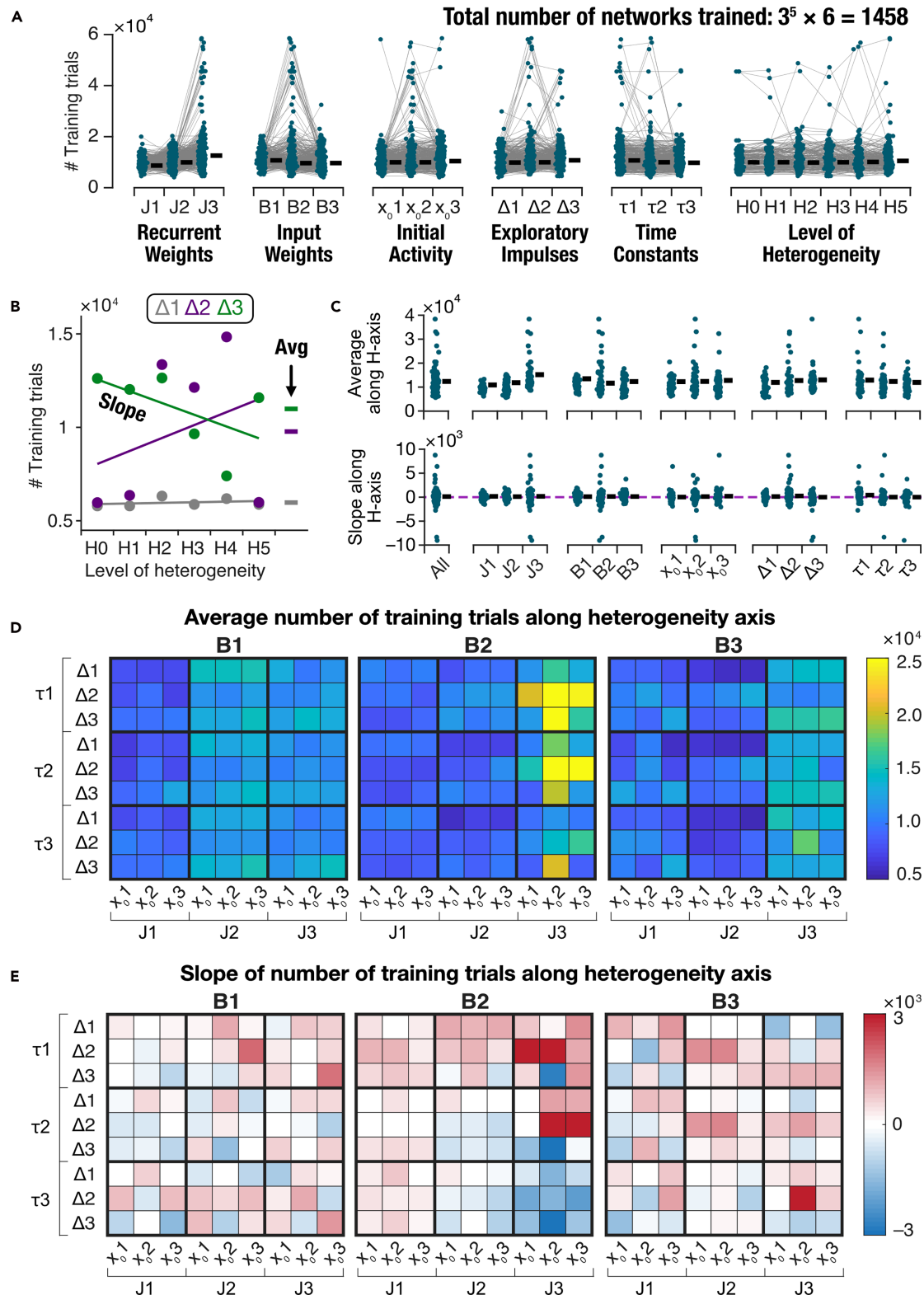


Figure 5. Variability in the impact of different network hyperparameters on the number of training trials

(A) Variability in the number of trials required for convergence plotted across each value of the five hyperparameters and heterogeneity levels. The black lines depict the median number of trials across all the networks with the corresponding value of hyperparameter ($N = 486$ for each of the five hyperparameters and 243 for each heterogeneity level).

(legend continued on next page)

number of training trials remained similar across all values of heterogeneities despite pronounced network-to-network heterogeneities across the different networks (Figure 5A; H0–H5 plots).

We computed the average value of training trials across levels of heterogeneity for each of the hyperparameter values to assess the dependence on individual hyperparameters (Figures 5B and 5C). We plotted these average values for all networks that were built with that specific hyperparameter value (Figure 5C). The average number of training trials across heterogeneity levels were distributed differently for different values of hyperparameters, with larger ranges for networks initialized with J3 or B2 (Figure 5C). In addition, to quantify the dependence on the level of heterogeneities, we performed a linear fit on the number of training trials plotted as a function of heterogeneity levels. We used the slope value to determine the dependence of training performance on heterogeneity levels for individual networks (Figure 5B). We plotted these slope values for all networks that were built with that specific hyperparameter value (Figure 5C). Slope values spanned both negative and positive values, indicating that the number of training trials were increasing with heterogeneity levels for some networks, while decreasing for others (Figure 5C). The distribution of slopes also spanned a large range of approximately $-10,000$ to $10,000$ trials per heterogeneity level for J3 and B2 networks, although the median across the population was 127 trials per heterogeneity level (Figure 5C). Strong dependencies of cross-heterogeneity average and slope values on hyperparameters governing recurrent and input weights were noted from these plots as well (Figure 5C).

To visualize the interdependencies of the effects of individual hyperparameters and cross-dependencies with heterogeneities, we plotted the slope and average of the number of training trials across heterogeneity levels (from Figures 5B and 5C) for all the five hyperparameter dimensions (Figures 5D and 5E). These visualizations emphasized our earlier observations on higher number of training trials being required for networks initialized with J3 or B2 (Figure 5D). Consistent with this, the slopes were also steeper for the same combination of seed values but spanned both negative and positive values (Figure 5E). The strong dependencies of the number of training trials on the seed values governing the two sets of weight values (recurrent and input) were discernable with this visualization. Importantly, we note that no hyperparameter individually governed the number of training trials, emphasizing the strong interdependencies and specific combinations required for either higher or lower values of the average or the slope (Figures 5D and 5E). For instance, even for the strong J3-B2 combination, the J3-B2- x_0 2- Δ 2- τ 1 combination yields high average value and highly positive slope value, the J3-B2- x_0 2- Δ 3- τ 1 combination yields high average value and highly negative slope value, and the J3-B2- x_0 2- Δ 1- τ 1 combina-

tion yields low average value and ~ 0 slope value (these are the three example plots provided in Figure 5B as Δ 1, Δ 2, and Δ 3 with the rest of the hyperparameters identical; Figures 5D and 5E).

We repeated similar analyses for task performance error values for all 1,458 networks, with trials spanning all input values and epoch durations (Figure S4). Hyperparameters governing recurrent and synaptic weights had strong impacts on the median error values (Figure S4A), the average error values computed across heterogeneities (Figure S4C), and the slope defining the dependence on heterogeneity levels (Figure S4C). There was substantial network-to-network variability in average error values (Figure S4D) as well as in the slope values (Figure S4E), which spanned both negative and positive values. Average error was higher for certain combinations such as J1-B2, J2-B2, and J2-B3 networks (Figure S4D) whereas the slopes were the steepest for J3-B2 set of networks (Figure S4E). These observations on pronounced network-to-network variability in the dependence of performance errors on heterogeneity levels were consistent with our earlier observations on the 53 distinct networks tested earlier (Figures 2D–2E).

To compute dependencies of network dynamics and activity trajectories on each hyperparameter, we individually computed PCA on the activity patterns of each of the 1,458 networks. We then transformed these to a common latent space using N1H0 (which is J1-B1- x_0 1- Δ 1- τ 1-H0 with the hyperparameter-based notation) as the reference network (as in Figure 3A). We then computed the Euclidean distances between the latent-space dynamics for each network with respect to its homogeneous counterpart (as in Figures 3 and 4). We plotted trial-averaged distance values for networks with the same specific hyperparameters (Figure 6A). The distribution of average distance in time for all the networks spanned 0 to 4 units with a median close to 0.5 units across different hyperparameters (Figure 6A). Consistent with the number of training trials (Figure 5A) and trial-averaged performance errors (Figure S4A), there was a strong dependence of median distance values on hyperparameters governing recurrent and input synaptic weights, with other hyperparameters influencing the median to a lesser extent (Figure 6A). Consistent with our prior observations with the 53 distinct networks (Figure 4E), the median value of the distance values progressively increased through H1–H5 (Figure 6A).

To evaluate the dependence of network dynamics on training heterogeneities, we computed the cross-heterogeneity average and slope values across all heterogeneous networks and plotted those as well for networks with the specific hyperparameter values (Figures 6B–6E). Homogeneous networks were not included for the estimation of average or slope values since distances were calculated with respect to the corresponding H0

(B) Number of trials taken to train an example network J3-B2- x_0 2- τ 1 with Δ 1 (gray), Δ 2 (purple), and Δ 3 (green). The actual number of trials (dots) and the best linear fits (lines) are plotted along the heterogeneity axis (H0–H5). The dashes represent the average number of training trials across heterogeneity levels.

(C) Distribution of the average number of training trials across heterogeneity level (top) and the slope of training trials as a function of the heterogeneity level (bottom). The distribution is plotted for all networks together (All) and for different values of the five hyperparameters (J1– τ 3). The black line depicts the median value of average (top) and the slope (bottom) along the heterogeneity axis for each network.

(D and E) Average (D) and slope (E) of the number of training trials across heterogeneity levels plotted as a heatmap for all combinations of the hyperparameter values. The three plots contain all the 81 networks with the initial input synaptic weights initialized as B1 (left), B2 (middle), and B3 (right). In each matrix, the x axis depicts the three different values of recurrent weights (J1–J3) as the major subdivisions and the three values of the initial activity values (x_0 1– x_0 3) as the minor subdivisions. Similarly, the y axis depicts the three different sets of time constant values (τ 1– τ 3) as the major subdivisions and the three sets of impulse patterns (Δ 1– Δ 3) introduced during training as the minor subdivisions.

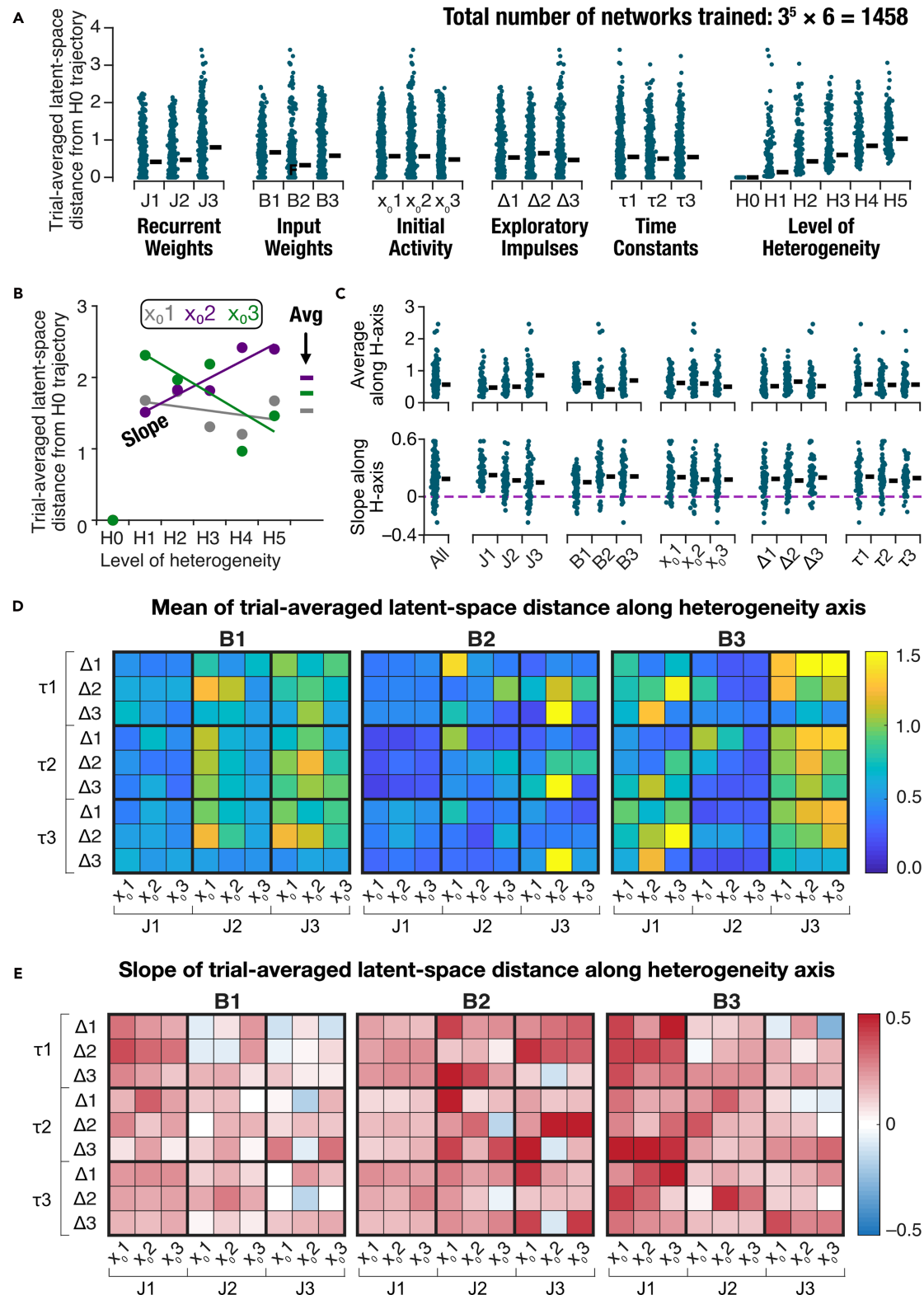


Figure 6. Variability in the impact of different network hyperparameters on network dynamics

(A) Variability in the mean distance calculated with respect to H0 network trajectory across each value of the five hyperparameters and heterogeneity levels. The black lines depict the median of mean distance across all the networks with the corresponding value of hyperparameter ($N = 486$ for each of the five hyperparameters and 243 for each heterogeneity level).

(legend continued on next page)

network, trivially yielding zero values for H0 networks in all cases. We observed a strong dependence of average distance values (computed across heterogeneity levels for every single network) on hyperparameters governing recurrent and input synaptic weights (Figures 6C–6D). Importantly, although the median slope values were positive across all hyperparameter values (Figure 6E), indicating that network dynamics progressively diverged from the H0 network with increasing level of heterogeneities, there was pronounced network-to-network variability in this dependence (Figures 6C and 6E). Specifically, there were networks and hyperparameter combinations that showed weak negative slopes, indicating a reduction in distance with increasing level of training heterogeneity (Figures 6C and 6E). These plots also emphasized the importance of combinations of hyperparameters rather than dominance of one over the other. For instance, the J3-B2- x_02 - $\Delta3$ combination (irrespective of the hyperparameter governing τ) yields high average value and weakly negative slope value, the J1-B3 combination yields variable average value and highly positive slope value, and the J3-B3- x_03 - $\Delta2$ - $\tau3$ combination yields high average value and ~ 0 slope value (Figures 6D–6E).

An extremely important observation that emerges from comparing the three metrics we used on training (Figure 5), performance errors (Figure S4) and latent-space distances of network trajectories (Figure 6) is that different hyperparameter combinations strongly affect each of these metrics. For instance, the J3-B2 combination was dominant in inducing variability in the average and slope values of training trials (Figures 5D–5E), the B2-J1 and B2-J2 combinations were dominant for average performance error (Figure S4D), the B2-J3 combination was dominant for the slope of performance error (Figure S4E), and the J3-B3 combination was dominant in inducing variability in average and slope values of the distance measure (Figures 6D–6E). These observations further emphasize the need for multiple metrics in the evaluation of the impact of neural heterogeneities on circuit functions.

Together, the systematic evaluation of the impact of individual hyperparameters on different metrics of network functions showed the dominant role of hyperparameters governing recurrent and input synaptic weights in regulating network training, performance, and dynamics. There was pronounced network-to-network variability in all measurements across all assessed hyperparameters, together emphasizing that no single hyperparameter individually governed network training, performance, or dynamics. These observations also unveiled strong interdependencies across hyperparameters and the importance of specific combinations for achieving specific network outcomes. A critical outcome of these results is that the metrics

governing network training, performance, and dynamics are differentially regulated by different hyperparametric combinations. Collectively, these results highlight each of the points emphasized in our framework involving training graded heterogeneities, the need for the population-of-network approach to study network-to-network variability, and the importance of using multiple performance metrics rather than focusing on one metric. They also suggest that the focus must be on specific non-random combinations of hyperparameters governing network function, rather than narrow focus on any single hyperparameter.

Variable robustness of trained networks to post-training intrinsic perturbations

Biological neurons and networks constantly undergo changes in the process of learning new tasks or homeostatic mechanisms or non-specific disturbances, thereby introducing perturbations in different circuit components. We tested the robustness of the trained networks to these different kinds of such perturbations, specifically toward assessing the robustness of networks trained with different levels of training heterogeneities (H0–H5) to these post-training perturbations. We introduced post-training perturbations (referred to as P0–P5 to distinguish from training heterogeneities H0–H5) of different forms to each of the 318 networks at six different levels, starting at P0 representing no perturbation and P5 denoting the highest level of perturbation. The graded levels in different post-training perturbations were essential because (1) the exact level of perturbation in biological systems might be variable and (2) this approach allows for assessing the extent of different kinds of perturbations that the learned system was robust to. For each of the several forms of post-training perturbations (Figures 7–10), we performed all test trials and compared error levels and network dynamics across different levels of post-training perturbations for all 318 networks. The difference in errors as well as the distances measured in latent-space trajectories were computed for each of the P1–P5 heterogeneity levels with reference to the P0 network. As a population, the distance distributions across different durations of stimulus and response epochs were broadly comparable (Figure 4). Therefore, for the perturbation analyses below, all distance-based calculations are shown for the longest duration trials.

We studied the impact of post-training perturbations on the intrinsic properties of recurrent units after the training was complete. Specifically, we tested two types of post-training intrinsic perturbations in time constant distributions. First, the extent of heterogeneities in terms of the distance of time constants from the mean was kept constant, while the mean time constant was gradually shifted (Figure 7A). Across networks, we observed an

(B) Mean distance across all 16 trials in an example network J3-B3- $\Delta1$ - $\tau1$ with x_01 (gray), x_02 (purple), and x_03 (green). The mean distance (dots) and the best linear fits (lines) are plotted along the heterogeneity axis (H0–H5). The dashes represent the average distance across all trials and heterogeneity levels.

(C) Distribution of the average distance across all trials and heterogeneity levels (top) and the slope of the mean distance as a function of the heterogeneity level (bottom). The distribution is plotted for all networks together (All) and for different values of the five hyperparameters (J1– $\tau3$). The black line depicts the median value of average (top) and slope (bottom) along the heterogeneity axis for each network.

(D and E) Average (D) and slope (E) of mean distance across heterogeneity levels plotted as a heatmap for all combinations of the hyperparameter values. The three plots contain all the 81 networks with the initial input synaptic weights initialized as B1 (left), B2 (middle), and B3 (right). In each matrix, the x axis depicts the three different values of recurrent weights (J1–J3) as the major subdivisions and the three values of the initial activity values (x_01 – x_03) as the minor subdivisions. Similarly, the y axis depicts the three different sets of time constant values ($\tau1$ – $\tau3$) as the major subdivisions and the three sets of impulse patterns ($\Delta1$ – $\Delta3$) introduced during training as the minor subdivisions.

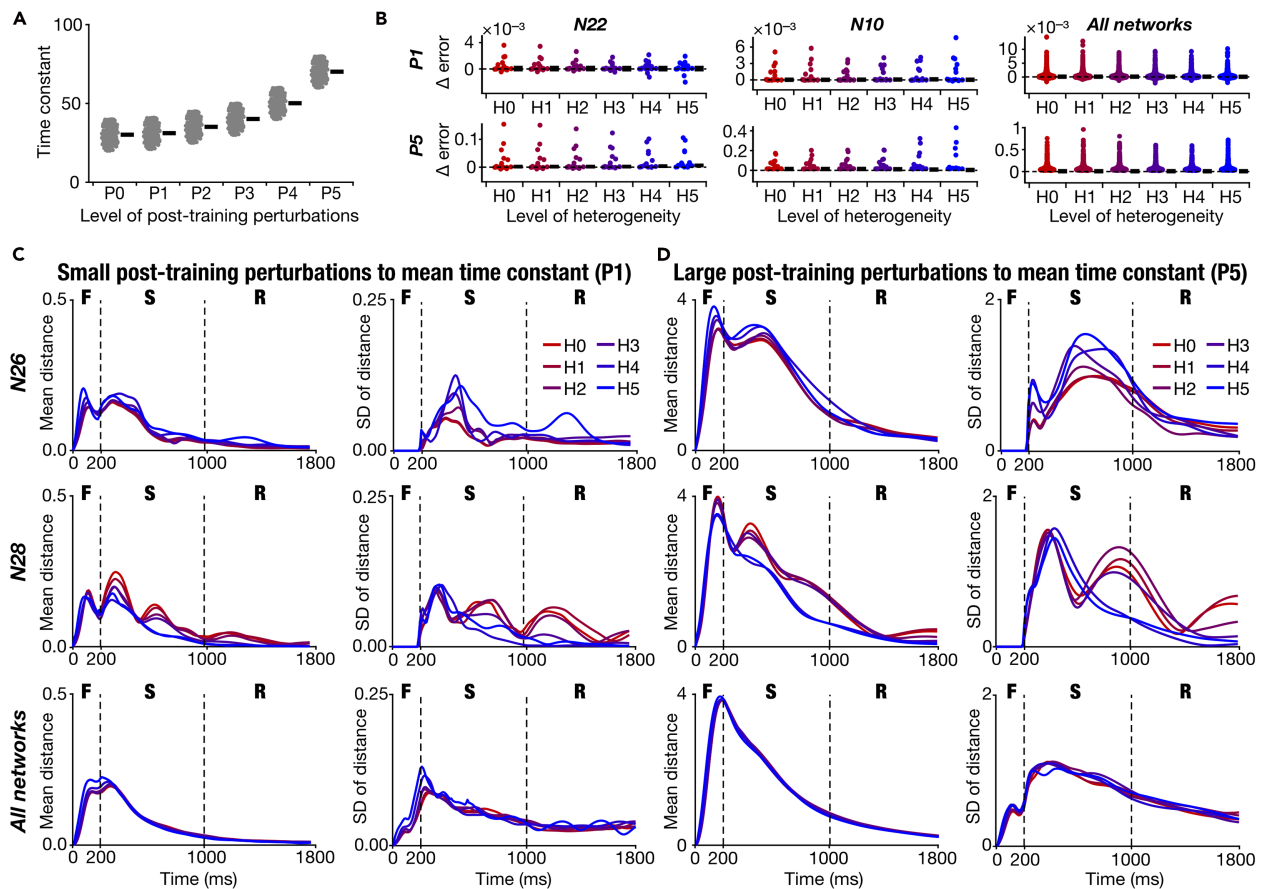


Figure 7. Pronounced network-to-network variability in the robustness of heterogeneous recurrent networks to post-training intrinsic perturbations

(A) Distribution of time constants for six different levels of post-training perturbations (P0–P5). Note that the mean (black bar) increases as the level of post-training perturbations is increased, which is distinct from heterogeneities introduced during training (H0–H5; Figure 1) where the mean was kept constant.

(B) Distribution of the difference in response errors (across different trials) with respect to P0 for low (P1; top row) and high (P5; bottom row) levels of post-training perturbations, introduced in networks trained with H0–H5 heterogeneities. The first column shows an example network (N22) that showed a reduction in error with increase in the level of training heterogeneities (H0–H5). The second column depicts another network (N10) where the error increased in networks trained with higher level of heterogeneities. The third column summarizes the difference in errors across all 53 networks, spanning different trials. Black bars represent the median values.

(C) Mean (left column) and standard deviation (SD, right column) of the distance distribution (computed from trajectories in the latent space) with respect to P0 for low level of post-training perturbation (P1). Each panel shows plots for networks trained with different levels of heterogeneity (H0–H5). The first row shows an example network (N26) where the mean distance and the associated variability increased with increasing level of training heterogeneities. The second row depicts another network (N28) where the mean distance and the associated variability decreased with increasing level of training heterogeneities. The third row depicts mean distance and the associated variability across all 53 networks.

(D) Same as (C) but for P5 instead of P1.

F, fixation epoch; S, stimulus epoch; R, response epoch.

increased difference in trial performance errors (Figure 7B), as well as an increase in distances of latent-space trajectories (Figure 7C vs. 7D), both computed with reference to the P0 trials, with increase in post-training perturbations from P1 to P5. With increase in the level of post-training perturbations from P1 (Figure 7C) to P5 (Figure 7D), we found an increase in both mean distance and the variability (SD) across different types of inputs.

We then asked if the robustness of the network in terms of error performance or network dynamics to post-training perturbations (P0–P5) was dependent on the level of intrinsic heterogeneities used during training (H0–H5). We found that in certain networks (e.g., N22 for P5 level of perturbation), the change in

maximum error (from P0) decreased with increase in the level of training heterogeneities (from H0 to H5), whereas in other networks (e.g., N10 for P5 level of perturbation), the change in maximum error increased from H0–H5. Similarly, there were certain networks (e.g., N28) where the mean and variance in trajectory distances decreased with increasing training heterogeneities, whereas in other networks (e.g., N26), the opposite was true. As a consequence of pronounced variability within networks (across trials) and across networks, the overall differences in errors (Figure 7B; all networks) as well as the mean and SD in distances (Figures 7C and 7D; all networks) were comparable across H0–H5 levels of training heterogeneities.

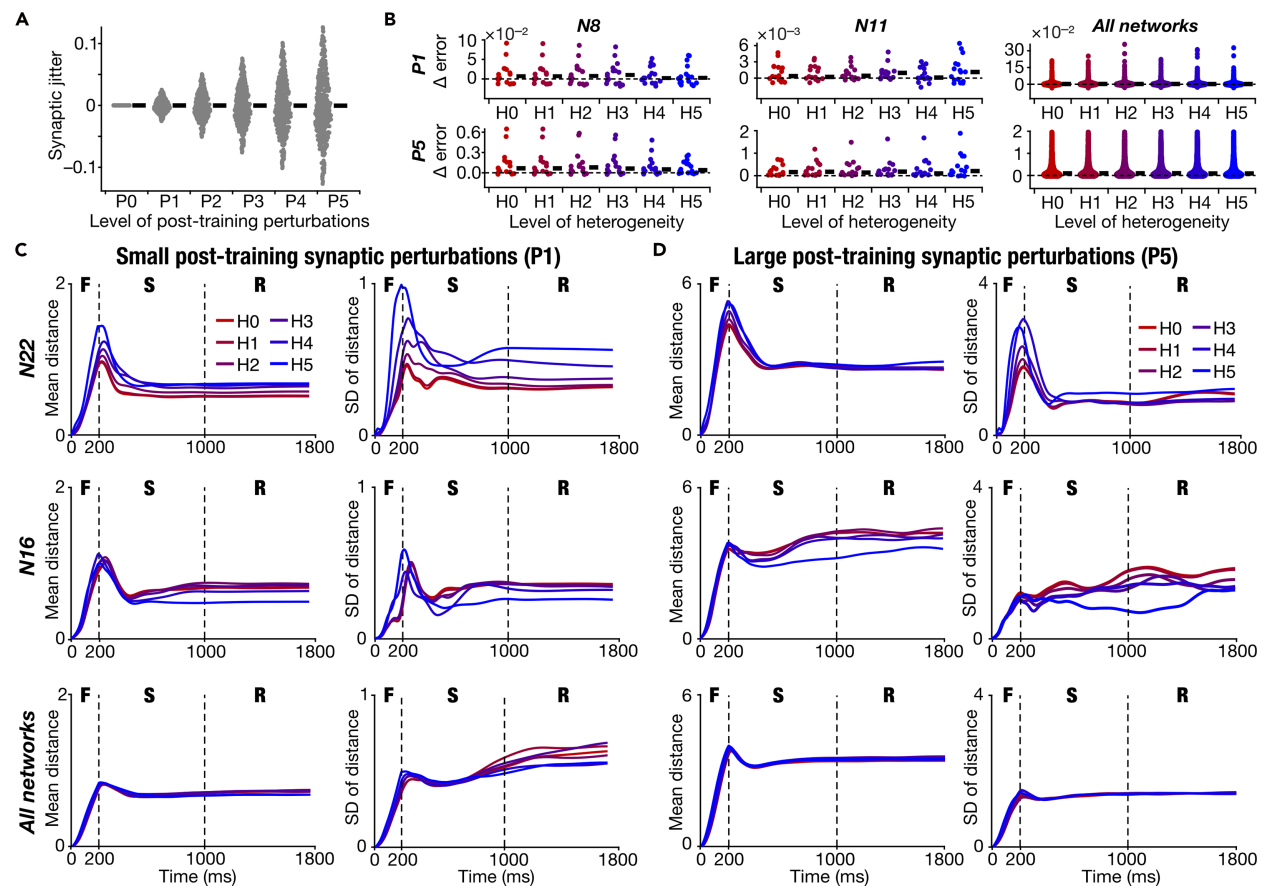


Figure 8. Pronounced network-to-network variability in the robustness of heterogeneous recurrent networks to post-training synaptic jitter (A) Distribution of synaptic jitter (deviation from trained values) for six different levels of post-training perturbations (P0–P5). (B) Distribution of the difference in response errors (across different trials) with respect to P0 for low (P1; top row) and high (P5; bottom row) levels of post-training perturbations, introduced in networks trained with H0–H5 heterogeneities. The first column shows an example network (N8) that showed a reduction in error with increase in the level of training heterogeneities (H0–H5). The second column depicts another network (N11) where the error was high in networks trained with higher level of heterogeneities. The third column summarizes the difference in errors across all 53 networks, spanning different trials. Black bars represent the median values. (C) Mean (left column) and standard deviation ([SD], right column) of the distance distribution (computed from trajectories in the latent space) with respect to P0 for low level of post-training perturbations (P1). Each panel shows plots for networks trained with different levels of training heterogeneity (H0–H5). The first row shows an example network (N22) where the mean distance and the associated variability increased with increasing level of training heterogeneities. The second row depicts another network (N16) where the mean distance and the associated variability decreased with higher level of heterogeneities introduced during training. The third row depicts mean distance and the associated variability across all 53 networks. (D) Same as (C) but for P5 instead of P1. Note that in (C and D), the mean distance was high even during the response window, indicating the deleterious impact of synaptic jitter on network performance. F, fixation epoch; S, stimulus epoch; R, response epoch.

In a second scenario (Figure S5A), we retained the mean to be the same (30 ms) with the variance of post-training perturbations changing across P0–P5 (with P0 representing zero variance). As this was identical to how H0–H5 were defined (Figure 1), a network trained with Hx showed minimal error for Px (Figure S5B). Therefore, in this case, errors and distances were computed with reference to the heterogeneity level that the specific network was trained with, rather than with the respective P0. The changes in error were minimal for the heterogeneity level that the network was trained with and increased with increasing distance from the level that it was trained with (Figure S5B). A similar observation held for the mean and SD of distances in the latent-space distance trajectories (Figures S5C and 5D). Importantly, there was network-to-

network variability in terms of error performance in H0–H5 networks (Figure S5B), as well as in the mean and variance of latent-space distance trajectories (Figures S5C and S5D).

Variable robustness of trained networks to post-training synaptic perturbations

Learned synaptic strengths in a biological network could change because of learning additional tasks or homeostatic mechanisms or non-specific perturbations. To assess the impact of such post-training synaptic changes, we added Gaussian noise to the trained synaptic weights (Figure 8A) and evaluated network performance and dynamics of all 318 networks. We refer to these post-training synaptic changes as synaptic jitter,

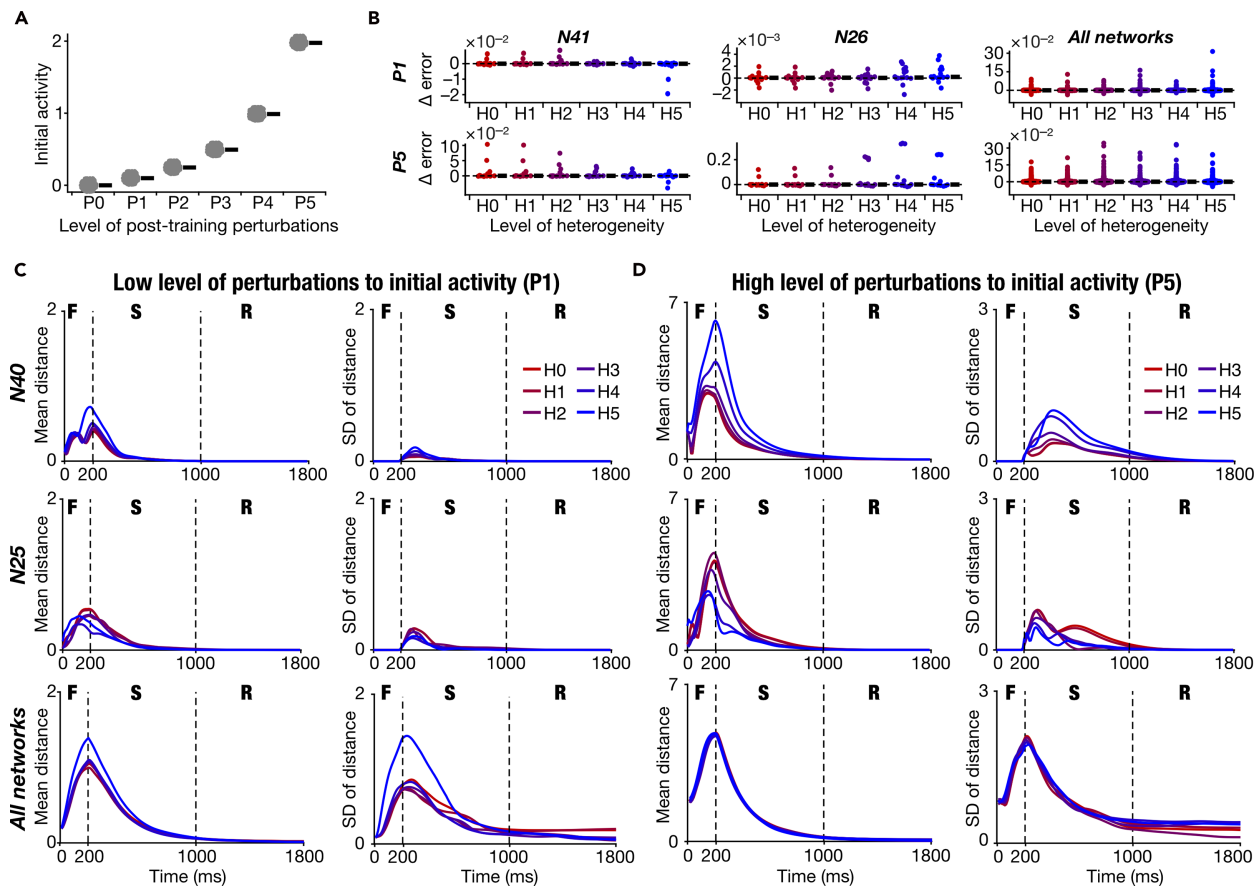


Figure 9. Pronounced network-to-network variability in the robustness of heterogeneous recurrent networks to post-training shift in network activity at task initiation

(A) Distribution of network activity range in all recurrent units at network initiation for six different levels of post-training perturbations (P0–P5). Note the shift in the mean value of the initial activity distribution with increase in the level of post-training perturbations.

(B) Distribution of the difference in response errors (across different trials) with respect to P0 for low (P1; top row) and high (P5; bottom row) levels of post-training perturbations, introduced in networks trained with H0–H5 heterogeneities. The first column shows an example network (N41) that showed a reduction in error with increase in the level of training heterogeneities (H0–H5). The second column depicts another network (N26) where the error was high in networks trained with higher level of heterogeneities. The third column summarizes the difference in errors across all 53 networks, spanning different trials. Black bars represent the median values.

(C) Mean (left column) and standard deviation ([SD], right column) of the distance distribution (computed from trajectories in the latent space) with respect to P0 for low level of post-training perturbations (P1). Each panel shows plots for networks trained with different levels of training heterogeneity (H0–H5). The first row shows an example network (N40) where the mean distance and the associated variability increased with increasing level of training heterogeneities. The second row depicts another network (N25) where the mean distance and the associated variability decreased with higher level of heterogeneities introduced during training. The third row depicts mean distance and the associated variability across all 53 networks.

(D) Same as (C) but for P5 instead of P1.

F, fixation epoch; S, stimulus epoch; R, response epoch.

introduced at different levels P0–P5, defined by linearly increasing the SD (from 0 to 0.05) of the Gaussian noise. Performance errors were high for most networks and got progressively worse with increasing levels of synaptic jitter (Figure 8B). There were few networks where the difference in errors with synaptic jitter showed an increase (e.g., N11) or decrease (e.g., N8) from H0 to H5 networks (Figure 8B). On the average, error distribution across H0–H5 was comparable for P1 and P5 (Figure 8B). Importantly, the mean and the variability in the distances between trajectories of P1–P5 networks, compared to their P0 counterparts, were high even during the response epoch (Figures 8C and 8D). Deviations from the P0 network dynamics

progressively increased from P1 to P5 (Figures 8C vs. 8D). These observations indicated that the networks were unable to converge to the required fixed points when post-training synaptic jitter was introduced (Figures 8C and 8D). Network-to-network variability in the robustness of network dynamics on pre-training heterogeneities was observed in a few networks, with some showing increased distances (N22), while others showing a reduction (N16) through H0–H5 (Figures 8C and 8D). Of all the post-training perturbations tested, we found synaptic jitter to be the most detrimental post-training perturbation, which severely affected all networks across all training heterogeneity levels even during the response epochs.

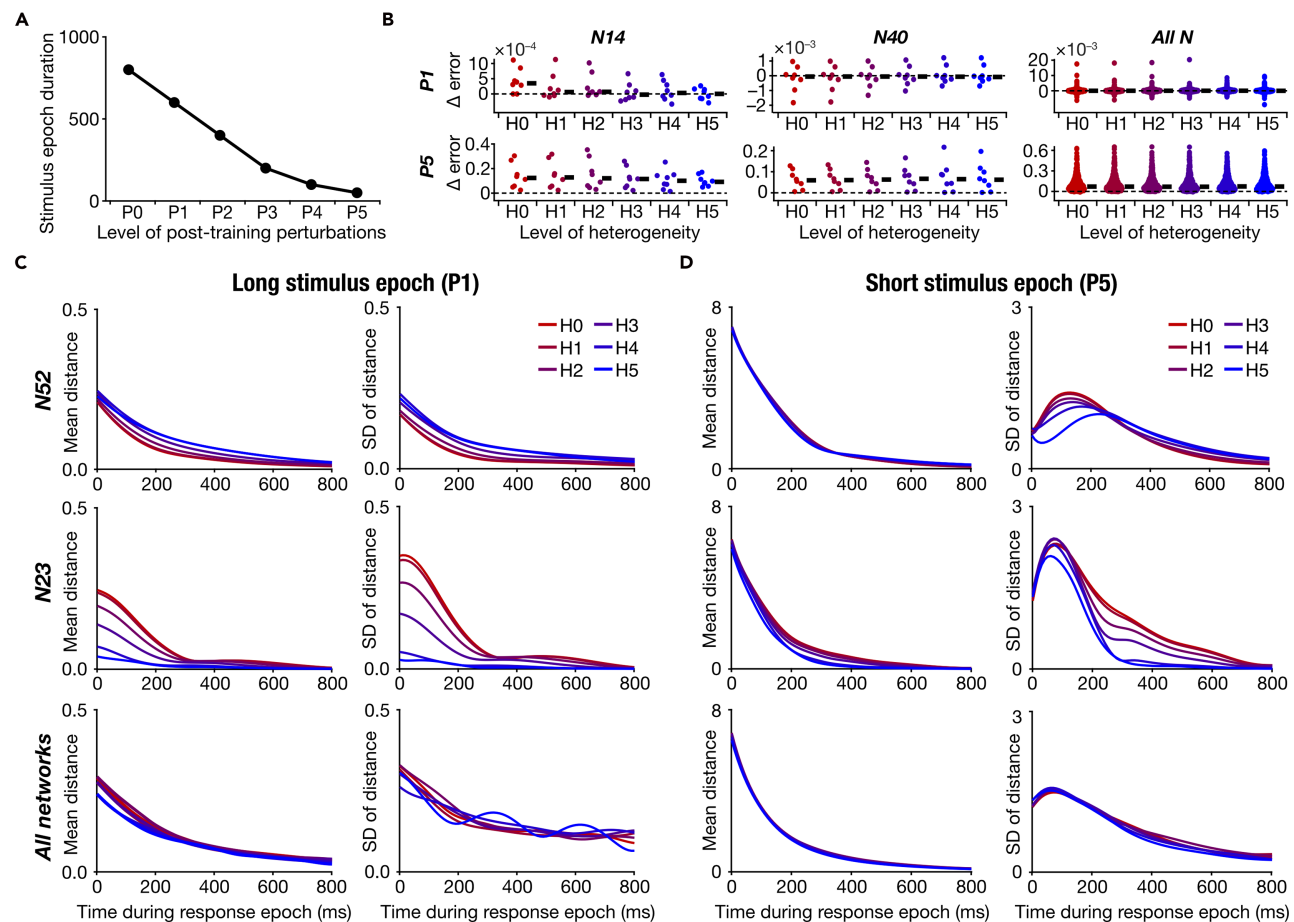


Figure 10. Pronounced network-to-network variability in the robustness of heterogeneous recurrent networks to post-training shift in the stimulus epoch duration

(A) Value of stimulus epoch duration for six different levels of post-training perturbations (P0–P5). (B) Distribution of the difference in response errors (across different trials) with respect to P0 for low (P1; top row) and high (P5; bottom row) levels of post-training perturbations, introduced in networks trained with H0–H5 heterogeneities. The first column shows an example network (N14) that showed a reduction in error with increase in the level of training heterogeneities (H0–H5). The second column depicts another network (N40) where the error was high in networks trained with higher level of heterogeneities. The third column summarizes the difference in errors across all 53 networks, spanning different trials. Black bars represent the median values. (C) Mean (left column) and standard deviation ([SD], right column) of the distance distribution (computed from trajectories in the latent space) with respect to P0 for low level of post-training perturbations (P1). Each panel shows plots for networks trained with different levels of training heterogeneity (H0–H5). The first row shows an example network (N52) where the mean distance and the associated variability increased with increasing level of training heterogeneities. The second row depicts another network (N23) where the mean distance and the associated variability decreased with higher level of heterogeneities introduced during training. The third row depicts mean distance and the associated variability across all 53 networks. (D) Same as (C) but for P5 instead of P1.

For (C and D), the mean distance and the associated variability are depicted for the response epoch, as distance computation during the stimulus epoch was impossible, given the differences in stimulus epoch duration across the different post-training perturbation levels.

Variable robustness of trained networks to post-training perturbations in the initial activity of the network and exploratory activity impulses

The brain is in continual flux, with neurons in the network continually responding to variable external stimuli and internal activity. Therefore, the initial conditions at the beginning of a task can be variable and could be a parameter that defines task performance. How do trained networks respond to changes in initial activity patterns? We tested the robustness of the network to two types of post-training perturbations in initial activity patterns at the beginning of the trials ($x(0)$). First, the extent of heterogeneities in terms of the initial activity patterns from the mean

was kept constant while gradually shifting the mean activity value from P0 to P5 (Figure 9A). Changes in the initial state affected few trials in certain networks, but mostly the changes in error values were minimal across networks across different levels of post-training perturbations (Figure 9B). Network-to-network variability in the dependence of performance error on training heterogeneities was prevalent (Figure 9B), with some networks showing less error for H0 compared to H5 (e.g., N26) and others showing the opposite trend (e.g., N41). The network dynamics manifested deviations in network trajectories, which progressively increased through P1–P5, with changes in the initial state of the network. However, these deviations were

predominantly limited to the first half of the trial period (Figures 9C and 9D). In most cases, the network converged to the same final fixed points, thereby minimizing the final errors during the response epochs (Figures 9C and 9D). We did observe some networks showing more deviation in network dynamics for high levels of training heterogeneities (e.g., N40), whereas others showed enhanced deviations for low levels of training heterogeneities (e.g., N25) for both P1 (Figure 9C) and P5 (Figure 9D) levels of post-training perturbations. In a second scenario (Figure S6), we kept the mean activity to be the same with the range of post-training perturbations changing across P0–P5. Our observations here were very similar to the scenario where we changed the mean activity, whereby the deviations in dynamics (which did show network-to-network variability) had negligible impact on the response epoch and final outcomes (Figure S6).

We tested the network by introducing exploratory activity impulses, Δ (which were otherwise used only during training trials, but not during test trials), at frequencies in between 0 and 10 Hz defining P0–P5 (Figure S7A). Since the networks were trained with low frequency Δ , we expected that the network would be robust to low-frequency activity exploratory activity impulses. We found the difference in errors to be low (<0.4) across all levels of training heterogeneities and post-training exploratory activity impulses (Figure S7B). Network-to-network variability in the dependence on the level of training heterogeneities was observed, with some networks showing increasing error through H0–H5 (e.g., N13), while others showing a reduction (e.g., N3). The distance between network trajectories was high due to random exploratory activity impulses (Figures S7C and S7D). On the average, across networks, the deviation in network dynamics peaked at the start of stimulus epoch and the maximal deviations were not markedly different across H0–H5 levels of training heterogeneity (Figures S7C and S7D). However, there were networks like N20 where the deviations and their variance were higher for H5 than for H0, or like N53 where the deviations and their variance were higher for the H0 network than others (Figures S7C and S7D).

Together, we found large deviations in network dynamics with changes in network initialization, coupled with pronounced network-to-network variability in the dependence on training heterogeneities. However, these deviations did not affect the final responses as the network converged to the same fixed points during the response epoch of most trials across P1–P5 levels of post-training perturbations in network initialization (Figure 9; Figure S6). With the introduction of exploratory activity impulses at different frequencies, we found pronounced network-to-network variability in the robustness of network performance and its dependence on training heterogeneities, although overall errors in performance were low across different frequencies of exploratory activity impulses (Figure S7).

Variable robustness of trained networks to changes in stimulus epoch duration

During the training process, we had used a stimulus epoch duration of either 200 or 800 ms (Figure 1). We tested the robustness of the learned networks to variations in the stimulus epoch duration from 800 to 50 ms spanning P0–P5 levels of post-training perturbations (Figure 10A). The changes in errors were not high

for H0–H5 networks when epoch durations remained higher than 100 ms (Figure 10B). However, for epoch durations less than 100 ms, the errors were high irrespective of the level of training heterogeneity (Figure 10B). Given that the design of these trials involved variable stimulus epoch durations (through P0–P5), computation of distances between the network trajectories (comparing P1–P5 with P0 dynamics) was feasible only for the response epoch (Figures 10C and 10D). The distance values at the beginning of the response epoch progressively increased with the reduction in stimulus epoch duration (Figures 10C vs. 10D). However, across P1–P5, the deviation in network trajectories sharply declined and showed convergence in many networks, although the variability was high with shorter stimulus epoch durations (Figures 10C and 10D). Network-to-network variability was prevalent in terms of how robustness to stimulus epoch duration depended on H0–H5 levels of training. Whereas some networks (e.g., N23) showed lower distances for H5 vs. H0, others (e.g., N52) showed the opposite dependence. Nevertheless, the average distance and the associated variance across all networks did not vary much across H0–H5 (Figures 10C and 10D). Together, decreasing the stimulus epoch reduced the accuracy of network performance and enhanced deviations in network dynamics and associated variability. There was pronounced network-to-network variability in how the robustness to changes in stimulus duration depended on training heterogeneities.

Collectively, we observed pronounced network-to-network variability in how the different networks reacted to all the tested post-training perturbations (Figures 7–10 and S5–S7). Although network robustness to perturbations was strongly dependent on the degree of training heterogeneity, there was no monotonic relationship between network robustness and degree of heterogeneity. Instead, there was strong network-to-network variability in how robustness depended on the degree of heterogeneity, with additional dependencies on the form and the level of specific post-training perturbation being tested (Figures 7–10 and S5–S7). Among all the different post-training perturbations studied, we found synaptic jitter (Figure 8; Table S2) to be the most detrimental. Among the others, perturbations to the mean of the time constant distribution (Figure 7) or to the stimulus epoch duration (Figure 10) yielded graded increases in performance error and in the mean/variance in distances of latent-space trajectories through P1–P5. On the other hand, perturbations to the range of time constant distribution (Figure S5), to the mean of initial activity distribution (Figure 9), to the range of initial activity distribution (Figure S6), or to the frequency of exploratory activity impulses (Figure S7) had minimal impact on performance error.

DISCUSSION

Our analyses revealed that heterogeneities impacted training performance, task-execution dynamics, and robustness to post-training perturbations in diverse ways even if training parameters and task definitions were identical. First, we show pronounced network-to-network variability in the dependence of training performance on the level of heterogeneity, as measured by the training trials required and task performance errors (Figure 2). Second, the impact of training heterogeneities on network dynamics during task execution manifested substantial

variability across networks (Figures 3 and 4). Finally, post-training perturbations had a marked impact on performance errors and network dynamics, with considerable network-to-network variability in how robustness depended on the level of training heterogeneities (Figures 7–10).

The fundamental conclusion of our analyses is that the impact of neural-circuit heterogeneities on network performance is diverse and depends heavily on several considerations (form of heterogeneities, degree of heterogeneities, metrics used for measurement of performance, type of performance that is measured, the time point at which heterogeneities are measured, etc.). These observations imply that strong conclusions based on studying a single type of heterogeneity at a single degree using a single performance metric should not be made.

The need for a population-of-networks approach in studying the impact of neural-circuit heterogeneities

The pronounced network-to-network variability in the impact of heterogeneities in training performance and in task-execution dynamics critically emphasize the need for using a population-of-networks approach to study neural-circuit heterogeneities. The use of a single network would have biased conclusions, showing that heterogeneities had either a beneficial or a deleterious impact on performance. The use of a population of networks allowed us to identify different networks that manifested diverse dependencies on heterogeneities, in training performance and task dynamics. The importance of using a population-of-networks approach has been strongly emphasized by several studies to account for variability as well as the role of distinct components in implementing network function.^{22,25,73} Our analyses strongly emphasize the need for the use of population-of-networks approach in studying the impact of heterogeneities on recurrent networks implementing cognitive tasks.

An important aspect of the population-of-networks approach is the impact of perturbations on network performance. It has been repeatedly observed that networks performing the same tasks, but with variable sets of components respond differently to the same perturbations.^{25,58–64} The variable impact of these perturbations has also been used to assess the dynamics of dysfunction in these neural circuits. In this context, our analyses demonstrate pronounced network-to-network variability in the impact of different forms of post-training perturbations on network performance. Whereas certain networks manifested strong dependence on training heterogeneities, others showed negligible impact on the level of training heterogeneities. The diversity in the dependence of robustness to different forms of post-training perturbations could be employed as a handle on how different networks accomplish the task and what contributions training heterogeneities make to this robustness. Together, our framework strongly recommends the training and assessment of a population of networks with different hyperparameters in studying the impact of different forms of neural-circuit heterogeneities on task performance, dynamics, and robustness to post-training perturbations.

Diversity in the impact of heterogeneities across networks and their functional outputs

The study of neural heterogeneities has revealed a diversity of roles played by heterogeneities in affecting circuit function. Het-

erogeneities have been shown to be beneficial in terms of performance and robustness of decorrelating networks,^{23,41,42,64} as well as in recurrent networks that have been trained to perform certain classification tasks.⁵⁷ In contrast, heterogeneities have been shown to hamper the performance of continuous attractor networks (CAN), although the addition of a negative feedback loop to individual units stabilizes heterogeneous CAN models.^{44,45} Our analyses here extend this diversity of impact by demonstrating pronounced network-to-network variability in the impact of heterogeneities on training, task-execution dynamics, and robustness of recurrent networks performing a cognitive task. Our framework involved the training of recurrent networks with different hyperparameters and with different levels of training heterogeneities. Our assessment of trained networks spanned training performance, task execution, and robustness to several levels of different post-training perturbations with a variety of performance metrics. In conjunction with other studies mentioned above, our analyses suggest that the diversity in the impact of heterogeneities is critically dependent on several factors. Factors that govern such diversity include the kind of network and its architecture, the type of model used for individual units, network hyperparameters, the specific task that is performed by the network and its complexity, the type of heterogeneity being introduced, the degree of heterogeneity introduced, the specific methodology used to introduce heterogeneity, and the metrics used to evaluate performance. In addition to these, our analyses also emphasize that the level and form of post-training perturbations also need to be explicitly accounted for in assessing the robustness of heterogeneous networks to perturbations. Overall, our analyses emphasize the critical need to account for each of these factors in assessing the impact of neural-circuit heterogeneities on circuit function. Without accounting for these factors, conclusions are bound to be biased by the choices made with reference to each of these different factors that govern the diversity in the impact of heterogeneities on circuit function.

A complex system interpretation of heterogeneities in neural circuits

How do we reconcile this diversity and variability in the dependence of network function on component heterogeneities? Our analyses suggest the complex systems framework as an ideal substrate for analyzing the impact of heterogeneities on neural circuits and their performance. Complex systems are systems where several functionally specialized subsystems interact to yield collective functional outcomes.^{79–82} With specific reference to biological neural circuits and the different forms of heterogeneities that manifest there, each neuron could be considered as a functionally specialized subsystem and the synaptic connectivity defining specific interactions.⁸³ Complex systems are defined by two key attributes. First, the interactions among subsystems of a complex system are neither fully determined nor completely random. This intermediate level of randomness is characterized by network motifs, defined as subnetworks that appear more frequently than expected in random networks. With reference to neural circuits and their heterogeneities, our analyses demonstrate that arbitrarily random synaptic connectivity with random neuronal heterogeneities would not accomplish a given task. This may be readily inferred from the high

pre-training errors in task execution (Figure 2) as well as with the deleterious impact of synaptic jitter on network performance (Figure 8). Thus, within the complex systems framework, learning could be considered as a mechanism that identifies these motifs (the non-random interactions among different forms of heterogeneities) that are essential for execution of the specific task in hand.^{2,84}

The second key feature of complex systems, which rules out the “completely determined” extreme in complex systems, is degeneracy, defined as the ability of multiple combinations of distinct subsystems to achieve the same collective function.^{20–22,28,79,85–87} In a heterogeneous neural circuit, each neuron is endowed with disparate intrinsic properties and neurons are connected by weights whose distributions are non-uniform. Thus, the ability of disparate sets of functionally specialized subsystems interacting with each other in heterogeneous ways to yield the same collective function (implement the task) constitutes a form of degeneracy. The presence of several such combinations, elucidated by different network initializations with disparate forms of intrinsic and synaptic heterogeneities, that are successfully able to learn the task (Figure 2) with disparate network dynamics (Figures 3 and 4) fulfills this. Thus, learning could be interpreted as the adaptive means to find one of the several combinations of *compatible* heterogeneities in the different neural-circuit components (subsystems) to achieve efficacious task performance. Network-to-network variability in performance, dynamics, and robustness are natural consequences within the complex systems framework^{20,24,25,58,63,67,74} because different networks *learn* disparate combinations of specific levels of heterogeneities in various subsystems to achieve the same outcome. Specifically, systems manifesting degeneracy exhibit pronounced system-to-system variability in how different functional metrics are affected by identical perturbations to the same component.^{20,25,43,58–64,88} As our analyses clearly demonstrate, different networks responded with pronounced variability to each of the tested perturbations (Figures 7–10 and S5–S7), together emphasizing the complex-systems perspective and the degenerate routes to achieving task performance.

Therefore, *individual* neural heterogeneities should not be seen through the prism of whether they are beneficial or detrimental to circuit function. Instead, they should be interpreted from the perspective of how specific *combinations* of disparate forms of heterogeneities come together to implement a specific function. When each of the different components of the network are assessed individually, they manifest widespread heterogeneities. But, when specific forms of heterogeneities interact with each other, they are able to accomplish the collective function of the complex system that they define. For example, in our analysis, disparate combinations of interactions among specific synaptic weights with specific intrinsic properties (each of which is heterogeneous by itself within that network) implements collective function. Even in scenarios where heterogeneities have been shown to be deleterious, the presence of other components with appropriate heterogeneities in them have negated the deleterious impact of heterogeneities. For instance, different forms of heterogeneities have been shown to negatively impact CANs.^{44,45} However, the incorporation of an additional heterogeneous component that implements a slow negative feedback loop, either

through inhibition or through ionic currents in individual neurons, could stabilize these networks toward accomplishing their collective function.^{44,45} Therefore, within the complex systems framework, we argue that the focus should not be on heterogeneities in individual components. Instead, the emphasis must be on the global structure of heterogeneities across all components, how they interact with each other, and how they adapt together toward achieving collective function.^{21,22,24,74,85,89}

Limitations and future directions

Our analyses here was to develop a framework for assessing the impact of neural heterogeneities on circuit function, with networks endowed with a simple form of heterogeneity trained with one of the simplest of cognitive tasks. Methodologically, these analyses should be extended to networks trained to perform multiple tasks, each with more complexity than the task that we had employed here. Such analyses also could bring in other biological attributes such as networks with separate excitatory and inhibitory neuronal populations, single units with more complexity reflecting biological neurons with active dendrites, concomitant plasticity of intrinsic and synaptic properties, adaptive regulation of network components in response to perturbations, and connectivity (local as well as afferent) that reflects specific biological networks.^{23,57,64,74,83,90–93} With specific reference to perturbations, in our analyses, we did not explicitly incorporate adaptive compensatory mechanisms that might restore function in the post-perturbation regime. Thus, direct generalizations of conclusions here to biological networks that are endowed with several compensatory regulatory mechanisms should be avoided, emphasizing the need for future studies explicitly incorporating such adaptive regulation to the network design. Such incorporations, apart from improving biological cross-interpretability, would also increase the number of places where heterogeneities could be introduced at different levels, providing avenues for assessment of training performance, task-execution dynamics, and robustness to perturbation when networks are trained with these heterogeneities. From a learning-theoretic standpoint, an important avenue for further exploration on the impact of heterogeneities is in continual learning systems, especially where the compositionality of dynamic motifs aids continual learning of different tasks in the same network.^{1,2}

From a pathophysiological standpoint, an important line of future research would be to assess the quantitative impact of different neurological disorders on neural-circuit heterogeneities and how such changes in the level or form of heterogeneities affect circuit function. Although most analyses focus on the average shifts in specific quantities, given the diverse and strong impact of neural-circuit heterogeneities on circuit function, it is essential to quantify this across different measurements and ask if altered heterogeneities contribute to circuit function.⁶⁵ Pathological conditions could be viewed as consequences of change in levels of heterogeneities. Specifically, circuit dysfunction could emerge from changes in the level of certain forms of heterogeneities *in the absence* of concomitant changes in other forms. If other changes had occurred concomitantly, they could have collectively compensated each other toward achieving comparable task performance. The absence of concomitant changes in heterogeneities could, therefore, yield dysfunction in the functional outcomes.

METHODS

Network architecture

A fully connected recurrent network model of N ($=200$) units (Figure 1) was used for all analyses. The dynamical equation that governed network activity was⁷⁸:

$$\tau \frac{dx(t)}{dt} = -x(t) + J.r(t) + B.u(t) + \Delta(t), \quad (\text{Equation 1})$$

$$r(t) = \tanh(x(t)), \quad (\text{Equation 2})$$

where $x(t)$ represented the N -dimensional activity vector spanning all units, $r(t)$ defined the response vector of the network units, and $u(t)$ represented the M -dimensional feedforward input vector at time t . J depicted an $N \times N$ matrix of recurrent connection weights, initialized from a normal distribution of mean 0 and variance g^2/N . B is an $N \times M$ matrix of input connection weights initialized from a uniform distribution in the range $[-0.01, 0.01]$. τ denoted an N -dimensional vector of time constants assigned to each unit in the network.

During training, random exploratory activity impulses Δ , were independently introduced to the activity of each unit at an average frequency of 3 Hz.⁷⁸ As our integration time constant Δt was set at 1 ms (below), this was implemented by picking a random number, one for each of the 200 units for every time step, between 0 and 1, and generating an activity impulse if this value is less than 0.003. This translated to 3 activity impulses (on the average) per unit in every 1,000 time steps, together translating to an average frequency of 3 Hz. The amplitude values of these exploratory activity impulses were sampled from a uniform distribution that spanned $[-0.5, 0.5]$ and occurred randomly.

The vector $u(t)$ was made of three different types of inputs: $u = (u_{fix}, u_{stim}, u_{task})$. The fixation input u_{fix} distinguishes the fixation period of the cognitive task from the response period. At the beginning of a trial, u_{fix} was assigned a value of 1, indicative of the fixation period. Switching the value of u_{fix} to 0 signals the beginning of the response epoch, during which the network response was evaluated. Stimulus input vector $u_{stim} = (u_{stim1}, u_{stim2})$ is 2-dimensional and carried task-specific stimulus information to all units of the network. Each input u_{stim1} and u_{stim2} switched between one of three values: -1 , 0 , or 1 , with 0 indicating the absence of any stimulus and $-1/1$ representing one of two distinct kinds of inputs that the network was presented with. $u_{task}(t)$ is a 6-dimensional one-hot vector with each dimension specifying a different task. For any trial, only the task input corresponding to the current task was set at 1, while the others remained at 0. For all experiments reported in this study, the dimension for the “Go task” was set to 1 and the rest were set to zero.

In our simulations, a first-order Euler approximation in time steps of Δt for each unit i was used for implementing the dynamical system represented in Equation 1:

$$x_i(t + \Delta t) = \left(1 - \frac{\Delta t}{\tau_i} \right) x_i(t) + \frac{\Delta t}{\tau_i} \left[\sum_{j=1}^N J_{ij} r_j(t) + \sum_{k=1}^M B_{ik} u_k(t) + \Delta_i \right], \quad (\text{Equation 3})$$

where $N = 200$, $M = 9$, $g = 1.5$, and $\Delta t = 1$ ms. The initial activity pattern $x(0)$ was set by sampling a uniform distribution spanning $[-0.1, 0.1]$. To ensure that the units received information about the stimulus inputs (u_{stim}), the column vectors in B corresponding to the stimulus inputs were scaled up by a factor of 100. Four arbitrarily chosen units were maintained at a constant activation of 1 to act as bias inputs to other units in the network. One of the N units was designated as the output unit of the network whose response is evaluated as the network’s response.⁷⁸

Incorporating training heterogeneity in the model

The time constant associated with individual units was initialized as a vector to incorporate intrinsic heterogeneity into the network. τ was initialized from a uniform distribution in range $[\tau_{min}, \tau_{max}]$. Consistent with our earlier analyses on heterogeneities,^{28,45} we introduced heterogeneities in a graded fashion rather than using a binary homogeneous vs. heterogeneous distinction in networks. In implementing this gradation, we introduced six levels of heterogeneities in unit time constants. For the six levels of heterogeneities H0–H5 in that order, the ranges of time constants $[\tau_{min}, \tau_{max}]$ were defined by $\tau_{min} = 30, 29, 25, 20, 10, 1$ and $\tau_{max} = 30, 31, 35, 40, 50, 59$ (Figure 1). The homogeneous level (H0) corresponded to a network where all units were assigned an identical time constant value of 30 ms. The level of heterogeneities progressively increases through H1–H5 as indicated by the respective values of τ_{min} and τ_{max} , translating to a progressive increase in the range of time constants always centered at 30 ms (Figure 1). The highest heterogeneity level (H5) was defined by units being randomly assigned to time constant values in the range 1–59 ms. The choice of the range 1–59 ms was to ensure that the time constant does not become zero for any unit and that the mean time constant across all heterogeneities was fixed at 30 ms.

To avoid biases introduced by the use of a single network for all analyses, we used a population-of-networks approach throughout our analyses to assess potential network-to-network variability in training performance and response dynamics. Specifically, the major hyperparameters of our network are: the seed value that determines the initial values of the connectivity matrices J and B , the seed value that determines the initial values of the activity vector $x(0)$, the seed value that determines the values of time constants from their respective distributions (in heterogeneous networks), and the seed value governing the pattern of exploratory activity impulses Δ . Instead of using one set of initial conditions associated with these hyperparameters, we created 53 distinct networks (N1–N53) by picking 53 different values for these hyperparameters. Each of these 53 distinct networks were initialized with the 6 different intrinsic heterogeneities (H0–H5), together providing a total of $53 \times 6 = 318$ total networks that were trained for the cognitive task. The presence of a population of networks, each trained with different levels of heterogeneities, allowed us to quantitatively assess network-to-network variability in the impact of heterogeneities on training performance and response dynamics.

Cognitive task description

Each of the 318 networks were trained on a memoryless Go task (Figure 1). The overall task structure and the input structure were modified from Yang et al., 2019.¹ The choice of the simple Go

task with simpler input structure was to reduce computational complexity of the task, to allow us to focus on the impact of multiple levels of heterogeneities and different hyperparameters. Each trial of the task consisted of three epochs. The *fixation epoch* was 200 ms long, during which $u_{fix} = 1$. At the beginning of the *stimulus epoch*, a single stimulus was presented through either u_{stim1} or u_{stim2} inputs. Specifically, one of u_{stim1} or u_{stim2} was randomly set to 1 or -1 (the active stimulus) and the other stimulus was set to zero (the inactive stimulus). These stimulus inputs remained at those specific values for the rest of the entire trial duration. The stimulus epoch duration switched randomly between 200 and 800 ms to ensure that the network learned the task and not the specifically timed responses. The *response epoch* follows the stimulus epoch, signaled by u_{fix} being reset to 0. The network response and associated error were evaluated during the response epoch period. For the memoryless Go task, the target response must be equal to the only active stimulus input (u_{stim1} or u_{stim2}) in that trial (which would be either +1 or -1). The duration of the response epoch randomly switched between 400 and 800 ms, again to ensure that the network learned the task and not the specific timings. Thus, the minimum trial duration was 800 ms (200 + 200 + 400 across the three epochs), while the maximum trial duration was 1,800 ms (200 + 800 + 800). During training, the choice between u_{stim1} and u_{stim2} , the specific input (-1 or +1) to the chosen stimulus, and the durations of the stimulus and the response epochs were randomly varied across trials.

Training and testing procedures

The input and recurrent synaptic weights were modified at the end of each trial based on a modified form of reward-modulated Hebbian learning.⁷⁸ During a trial, the potential weight change or eligibility trace for each recurrent synapse from the j^{th} unit to the i^{th} unit was accumulated over the course of the trial as

$$e_{ij}^J(t) = e_{ij}^J(t-1) + (r_j(t-1)(x_i(t) - \bar{x}_i))^3, \quad (\text{Equation 4})$$

where $\bar{x}_i$ maintained a short-term running weighted average of the x_i calculated as $\bar{x}_i(t) = \alpha_x \bar{x}_i(t-1) + (1 - \alpha_x)x_i(t)$. Thus, $(x_i(t) - \bar{x}_i)$ tracked sudden fluctuations in the unit's activity, primarily due to the exploratory activity impulse Δ_i . Similarly, for each synapse from the k^{th} external input to the i^{th} unit, the eligibility trace e_{ik}^B was as follows:

$$e_{ik}^B(t) = e_{ik}^B(t-1) + (u_k(t-1)(x_i(t) - \bar{x}_i))^3. \quad (\text{Equation 5})$$

At the end of each trial, error was calculated over the response epoch (RE) as the average absolute difference between the network output ($r_{obs}(t)$) and expected output ($r_{exp}(t)$) for the trial as follows:

$$E = \text{mean}(|r_{obs}(t) - r_{exp}(t)|_{t \in RE}). \quad (\text{Equation 6})$$

Following conventions from Miconi, 2017,⁷⁸ the reward for the trial was taken as the negative of the error, $R = -E$. It may be noted that R is always negative as E represents the average absolute difference. The reward prediction error ($R - \bar{R}$) signal was then computed with $\bar{R}$, the expected reward, calculated for trial n as follows:

$$\bar{R}(n) = \alpha_R \bar{R}(n-1) + (1 - \alpha_R)R(n). \quad (\text{Equation 7})$$

The reward prediction error was used to modulate the eligibility traces and compute weight changes,

$$\Delta J_{ij} = \eta_J e_{ij}^J \bar{R}(R - \bar{R}), \quad (\text{Equation 8})$$

$$\Delta B_{ik} = \eta_B e_{ik}^B \bar{R}(R - \bar{R}), \quad (\text{Equation 9})$$

where η_J and η_B defined the learning rates for the **J** and **B** matrices, respectively. In these simulations, the default values of the relative weightages for computing $\bar{x}_i(t)$ and $\bar{R}(n)$ were $\alpha_x = 0.05$ and $\alpha_R = 0.33$, respectively. The default learning rate parameters were set at $\eta_J = \eta_B = 0.15$. Synaptic weights were updated using these relationships, and the network then switched to the next trial of the cognitive task to repeat the whole updation process again. Under default conditions, similar to the convergence structure in Miconi, 2017,⁷⁸ convergence of the network (completion of training) was defined to be achieved if the error value E was less than 0.05 for a threshold number of consecutive trials (T_{th}). The default value of T_{th} was taken as 100.

By default, trained networks were tested without the exploratory activity impulses. Each network was tested on 16 different trials covering all 4 combinations of inputs (either u_{stim1} or u_{stim2} taking a value of either -1 or +1) and 4 trial durations (800, 1,200, 1,400, or 1,800 ms). The network performance was then quantified based on the absolute error, E , computed for each trial. A trial was counted as a correct trial if $E < 0.05$.

Analyzing latent network dynamics

The temporally aligned responses of all 200 units in the network across all 16 trials were concatenated into a matrix **R** of dimensions $T \times N$ where N defined the total number of recurrent units and T represented the total number of time points across all 16 trials. To obtain the latent-space dynamics of each network, we performed PCA on the associated **R** matrix. The projections to the first three principal components of the activity of the N recurrent units across time, during different kinds of trials, were used for visualizing the latent-space trajectories in 3D. The dimensionality of the system was computed using the eigen values (λ_i) associated with the **R** matrix as follows^{94–96}:

$$\text{Dimensionality} = \frac{\left(\sum_i \lambda_i\right)^2}{\sum_i \lambda_i^2}. \quad (\text{Equation 10})$$

Since latent-space dynamics of different networks trained on the same task form different trajectories in different latent spaces, the latent spaces themselves need to be aligned before the network trajectories can be compared across networks. CCA applies linear transformations to find a new set of dimensions for the latent-space dynamics of a pair of networks, L_1 and L_2 , such that their pairwise correlation is maximized.⁵ When M_1 and M_2 defined the coefficient matrices for projecting L_1 and L_2 to the new dimensions, respectively, the projections of the latent-space dynamics along the new dimensions were calculated as

$$\tilde{L}_1 = M_1^T L_1, \tilde{L}_2 = M_2^T L_2. \quad (\text{Equation 11})$$

Then, the latent-space dynamics of network 2 aligned with network 1 was obtained as

$$\hat{L}_2 = M_1^{-1} M_2^T L_2. \quad (\text{Equation 12})$$

For each network, the projections along the first 4 PCs were taken for CCA transformation and the latent-space trajectories of each pair of networks were aligned. The dimensionality value calculated using Equation 10, across all 318 networks was around 3. As dimensionality approximately represents the number of dimensions required for explaining around 80% of the total variance, we used 4 dimensions for the CCA transformation. Therefore, the M_i matrices were of dimension 4×4 . For all our analyses, network N1 was used as the reference network for mappings.

To observe the variability in latent-space dynamics due to change in network configurations, Euclidean distances between trajectories in the aligned latent space was computed for each pair of networks. Let $\hat{L}$ and $\hat{K}$ be $T \times m$ matrices of aligned latent-space trajectories of the dynamics of two networks, where T defined the number of time points and $m = 4$ denoted the number of dimensions of the aligned trajectories. The distance between the trajectories as a function of time was then calculated as

$$d(\hat{L}, \hat{K}, t) = \sqrt{\sum_{i=1}^m (l_i(t) - k_i(t))^2}, \quad (\text{Equation 13})$$

where $l_i(t)$ and $k_i(t)$ defined the columns of the matrix $\hat{L}$ and $\hat{K}$, respectively. To quantify the impact of heterogeneity level on network dynamics, the distance between the heterogeneous (H1–H5) network trajectories was calculated with respect to the homogenous (H0) network trajectory.

Assessing interdependencies of the impact of network hyperparameters on network training, task performance, and dynamics

To assess the specific impact of the different values of the five major hyperparameters, we selected the set of seed values used to generate networks N1–N3 (Table S1): (1) the initial value of the recurrent connectivity matrices **J**: J1, J2, and J3; (2) the initial value of the input connectivity matrix **B**: B1, B2, and B3; (3) the initial values of the activity vector **x**(0): x_{01} , x_{02} , and x_{03} ; (4) the pattern of exploratory impulses Δ during training: Δ_1 , Δ_2 , and Δ_3 ; and (5) and the values of time constants, τ : τ_1 , τ_2 , and τ_3 . With all combinations of the five hyperparameters (3^5), each with six levels of heterogeneity, we generated and trained a total of $3^5 \times 6 = 1,458$ networks on the Go task.

We quantified the performance of these 1,458 networks using (1) the number of training trials, (2) the task performance error, and (3) the latent-space distance of network trajectories with respect to the corresponding H0 network's trajectory. To assess how these different metrics varied across heterogeneity levels for different networks, we performed a linear fit to the metric as a function of heterogeneity level. The slope of the linear fit and the mean across heterogeneity levels were compared for all the combinations of hyperparameters. As distances were calculated with the H0 network as reference, H0 networks were removed from the slope and average calculation for the trial-averaged distances to avoid bias due to the zero values for H0 networks.

Assessing the robustness of networks to post-training perturbations

The different kinds of post-training perturbations for which robustness of all 318 networks was assessed are described below (Table S2). Unless otherwise stated, the performance error and dynamics of the networks with post-training perturbations (P1–P5) were compared to that of the respective P0 network (where there was no post-training perturbation). In each of the cases below, to explore the robustness of the network against the post-training perturbations, the difference in the absolute trial errors with and without heterogeneities was calculated. Similarly, the impact of post-training perturbations on the latent-space dynamics was quantified as the distance of trajectories with post-training perturbation (P1–P5) from the respective P0 trajectory.

Intrinsic heterogeneity in unit time constants

We tested two types of post-training intrinsic perturbations in time constant distributions. First, the extent of heterogeneities in terms of the distance of time constants from the mean was kept constant, while the mean time constant was gradually shifted. The shifts in the time constant of each unit (in ms) were (0, 1, 5, 10, 20, 40), in that sequence, for P0–P5 levels of post-training perturbations. In this case, we used the respective P0 network as the reference for computing errors and distances (the default scenario mentioned above). In a second scenario, we kept the mean to be the same with the variance of post-training perturbations changing across P0–P5 (with P0 representing zero variance). As this was identical to how H0–H5 were defined, a network trained with Hx showed minimal error for Px. Therefore, in this case, errors and distances were computed with reference to the heterogeneity level that the specific network was trained with.

Jitter in recurrent synaptic weights

To assess the robustness of networks to changes in the trained values of recurrent synaptic weights, we introduced Gaussian jitter with zero mean and increasing variance. The SD values of the Gaussian noise for P0–P5 were (0, 0.01, 0.02, 0.03, 0.04, 0.05).

Distribution pattern of initial activity

These post-training perturbations reflect differences in initial activity patterns at the beginning of the trials (**x**(0)). We tested two types of post-training perturbations in initial activity patterns here. First, the extent of heterogeneities in terms of the initial activity patterns from the mean was kept constant, while the mean activity value was gradually shifted. The shifts in the initial activity value of each unit were (0, 0.1, 0.25, 0.5, 1, 2), in that sequence, for P0–P5 levels of post-training perturbations. In a second scenario, we kept the mean activity to be the same with the range of post-training perturbations changing across P0–P5 (with P0 representing zero perturbations). The range of heterogeneities in the initial activity value of each unit was (0, 0.1, 0.25, 0.5, 1, 2), in that sequence, for P0–P5 levels of post-training perturbations. For instance, for the P5 value, the range would be -2 to $+2$ for how the initial activity heterogeneities will be introduced. At P0, **x**(0) = 0.

Stimulus epoch duration

During the training process, we had used a stimulus epoch duration of either 200 or 800 ms. We tested the robustness of the learned networks to changes in the stimulus epoch duration.

The stimulus epoch durations for P0–P5 post-training perturbations were (800, 600, 400, 200, 100, 50) in that sequence.

Frequency of exploratory activity impulses

While random exploratory activity impulses were introduced during training process, they were turned off during testing. We tested the robustness of the learned networks to introduction of increasing frequencies of exploratory activity impulses. The frequency of exploratory activity impulses (in Hz) for P0–P5 post-training perturbations, in that sequence, were (0, 2, 4, 6, 8, 10).

Computational details

All simulations and analyses were performed within the MATLAB (Mathworks Inc.) programming environment.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Rishikesh Narayanan (rishi@iisc.ac.in).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

All original code has been deposited at Zenodo at <https://doi.org/10.5281/zenodo.17750827>⁹⁷ and is publicly available as of the date of publication.

ACKNOWLEDGMENTS

The authors thank Dr. Alex Cayco-Gajic, Dr. Laura Driscoll, and Dr. Sara Solla for helpful discussions. The authors thank members of the cellular neurophysiology laboratory for helpful discussions and for comments on a draft of this manuscript. This work was supported by the Wellcome Trust DBT India Alliance (senior fellowship to R.N.; IA/S/16/2/502727), Pratiksha Trust (A.S.), and the Ministry of Education (A.S. and R.N.).

AUTHOR CONTRIBUTIONS

A.S. and R.N. designed the experiments, A.S. performed the experiments, A.S. analyzed the data, and A.S. and R.N. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.patter.2026.101685>.

Received: November 24, 2025

Revised: February 2, 2026

Accepted: August 31, 2026

REFERENCES

- Yang, G.R., Joglekar, M.R., Song, H.F., Newsome, W.T., and Wang, X.J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* 22, 297–306. <https://doi.org/10.1038/s41593-018-0310-2>.
- Driscoll, L.N., Shenoy, K., and Sussillo, D. (2024). Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nat. Neurosci.* 27, 1349–1363. <https://doi.org/10.1038/s41593-024-01668-6>.
- Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84. <https://doi.org/10.1038/nature12742>.
- Sussillo, D., and Abbott, L.F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63, 544–557. <https://doi.org/10.1016/j.neuron.2009.07.018>.
- Gallego, J.A., Perich, M.G., Chowdhury, R.H., Solla, S.A., and Miller, L.E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* 23, 260–270. <https://doi.org/10.1038/s41593-019-0555-4>.
- Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. *Curr. Opin. Neurobiol.* 46, 1–6. <https://doi.org/10.1016/j.conb.2017.06.003>.
- Safaie, M., Chang, J.C., Park, J., Miller, L.E., Dudman, J.T., Perich, M.G., and Gallego, J.A. (2023). Preserved neural dynamics across animals performing similar behaviour. *Nature* 623, 765–771. <https://doi.org/10.1038/s41586-023-06714-0>.
- Cembrowski, M.S., and Spruston, N. (2019). Heterogeneity within classical cell types is the rule: lessons from hippocampal pyramidal neurons. *Nat. Rev. Neurosci.* 20, 193–204. <https://doi.org/10.1038/s41583-019-0125-5>.
- Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K.F., et al. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* 34, 199–203. <https://doi.org/10.1038/nbt.3445>.
- Mishra, P., and Narayanan, R. (2020). Heterogeneities in intrinsic excitability and frequency-dependent response properties of granule cells across the blades of the rat dentate gyrus. *J. Neurophysiol.* 123, 755–772. <https://doi.org/10.1152/jn.00443.2019>.
- Marder, E. (2011). Variability, compensation, and modulation in neurons and circuits. *Proc. Natl. Acad. Sci. USA* 108, 15542–15548. <https://doi.org/10.1073/pnas.1010674108>.
- Han, X., Guo, S., Ji, N., Li, T., Liu, J., Ye, X., Wang, Y., Yun, Z., Xiong, F., Rong, J., et al. (2023). Whole human-brain mapping of single cortical neurons for profiling morphological diversity and stereotypy. *Sci. Adv.* 9, eadf3771. <https://doi.org/10.1126/sciadv.adf3771>.
- Langlieb, J., Sachdev, N.S., Balderrama, K.S., Nadaf, N.M., Raj, M., Murray, E., Webber, J.T., Vanderburg, C., Gazestani, V., Tward, D., et al. (2023). The molecular cytoarchitecture of the adult mouse brain. *Nature* 624, 333–342. <https://doi.org/10.1038/s41586-023-06818-7>.
- Nandi, A., Chartrand, T., Van Geit, W., Buchin, A., Yao, Z., Lee, S.Y., Wei, Y., Kalmbach, B., Lee, B., Lein, E., et al. (2022). Single-neuron models linking electrophysiology, morphology, and transcriptomics across cortical cell types. *Cell Rep.* 40, 111176. <https://doi.org/10.1016/j.celrep.2022.111176>.
- Piwecka, M., Rajewsky, N., and Rybak-Wolf, A. (2023). Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease. *Nat. Rev. Neurol.* 19, 346–362. <https://doi.org/10.1038/s41582-023-00809-y>.
- Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K.W., Ding, S.L., Hu, L., Lönnerberg, P., Bakken, T., Casper, T., Clark, M., et al. (2023). Transcriptomic diversity of cell types across the adult human brain. *Science* 382, eadd7046. <https://doi.org/10.1126/science.add7046>.
- Dembrow, N.C., Sawchuk, S., Dalley, R., Opitz-Araya, X., Hudson, M., Radaelli, C., Alfiler, L., Walling-Bell, S., Bertagnoli, D., Goldy, J., et al. (2024). Areal specializations in the morpho-electric and transcriptomic properties of primate layer 5 extratelencephalic projection neurons. *Cell Rep.* 43, 114718. <https://doi.org/10.1016/j.celrep.2024.114718>.
- Huckleberry, K.A., and Shansky, R.M. (2021). The unique plasticity of hippocampal adult-born neurons: Contributing to a heterogeneous dentate. *Hippocampus* 31, 543–556. <https://doi.org/10.1002/hipo.23318>.
- Dahmen, D., Hutt, A., Indiveri, G., Kennedy, A., Lefebvre, J., Mazzucato, L., Motter, A.E., Narayanan, R., Payvand, M., Planert, H., and Gast, R. (2026). How heterogeneity shapes dynamics and computation in the brain. *Neuron* 114, 804–819. <https://doi.org/10.1016/j.neuron.2025.11.023>.

20. Albantakis, L., Bernard, C., Brenner, N., Marder, E., and Narayanan, R. (2024). The brain's best kept secret is its degenerate structure. *J. Neurosci.* 44, e1339242024. <https://doi.org/10.1523/JNEUROSCI.1339-24.2024>.
21. Rathour, R.K., and Narayanan, R. (2019). Degeneracy in hippocampal physiology and plasticity. *Hippocampus* 29, 980–1022. <https://doi.org/10.1101/203943>.
22. Prinz, A.A., Bucher, D., and Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nat. Neurosci.* 7, 1345–1352. <https://doi.org/10.1038/nn1352>.
23. Mishra, P., and Narayanan, R. (2019). Disparate forms of heterogeneities and interactions among them drive channel decorrelation in the dentate gyrus: Degeneracy and dominance. *Hippocampus* 29, 378–403. <https://doi.org/10.1002/hipo.23035>.
24. Mittal, D., and Narayanan, R. (2024). Network motifs in cellular neurophysiology. *Trends Neurosci.* 47, 506–521. <https://doi.org/10.1016/j.tins.2024.04.008>.
25. Calabrese, R.L., and Marder, E. (2025). Degenerate neuronal and circuit mechanisms important for generating rhythmic motor patterns. *Physiol. Rev.* 105, 95–135. <https://doi.org/10.1152/physrev.00003.2024>.
26. Cropper, E.C., Dacks, A.M., and Weiss, K.R. (2016). Consequences of degeneracy in network function. *Curr. Opin. Neurobiol.* 41, 62–67. <https://doi.org/10.1016/j.conb.2016.07.008>.
27. Wang, Y., Weiss, K.R., and Cropper, E.C. (2019). Network Degeneracy and the Dynamics of Task Switching in the Feeding Circuit in *Aplysia*. *J. Neurosci.* 39, 8705–8716. <https://doi.org/10.1523/JNEUROSCI.1454-19.2019>.
28. Mittal, D., and Narayanan, R. (2022). Heterogeneous stochastic bifurcations explain intrinsic oscillatory patterns in entorhinal cortical stellate cells. *Proc. Natl. Acad. Sci. USA* 119, e2202962119. <https://doi.org/10.1073/pnas.2202962119>.
29. Seenivasan, P., and Narayanan, R. (2020). Efficient phase coding in hippocampal place cells. *Phys. Rev. Res.* 2, 033393. <https://doi.org/10.1103/PhysRevResearch.2.033393>.
30. Chelaru, M.I., and Dragoi, V. (2008). Efficient coding in heterogeneous neuronal populations. *Proc. Natl. Acad. Sci. USA* 105, 16344–16349. <https://doi.org/10.1073/pnas.0807744105>.
31. Balasubramanian, V. (2015). Heterogeneity and Efficiency in the Brain. *Proc. IEEE* 103, 1346–1358. <https://doi.org/10.1109/Jproc.2015.2447016>.
32. Shamir, M., and Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. *Neural Comput.* 18, 1951–1986. <https://doi.org/10.1162/neco.2006.18.8.1951>.
33. Teşileanu, T., Cocco, S., Monasson, R., and Balasubramanian, V. (2019). Adaptation of olfactory receptor abundances for efficient coding. *eLife* 8, e39279. <https://doi.org/10.7554/eLife.39279>.
34. Marsat, G., and Maler, L. (2010). Neural heterogeneity and efficient population codes for communication signals. *J. Neurophysiol.* 104, 2543–2555. <https://doi.org/10.1152/jn.00256.2010>.
35. Wada, T., Hironaka, K.-i., and Kuroda, S. (2021). Cell-to-cell variability serves as information not noise. *Curr. Opin. Syst. Biol.* 27, 100339. <https://doi.org/10.1016/j.coisb.2021.04.010>.
36. Ahn, J., Kreeger, L.J., Lubejko, S.T., Butts, D.A., and MacLeod, K.M. (2014). Heterogeneity of intrinsic biophysical properties among cochlear nucleus neurons improves the population coding of temporal information. *J. Neurophysiol.* 111, 2320–2331. <https://doi.org/10.1152/jn.00836.2013>.
37. Roy, A., and Narayanan, R. (2021). Spatial information transfer in hippocampal place cells depends on trial-to-trial variability, symmetry of place-field firing, and biophysical heterogeneities. *Neural Netw.* 142, 636–660. <https://doi.org/10.1016/j.neunet.2021.07.026>.
38. Renart, A., Song, P., and Wang, X.J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* 38, 473–485. [https://doi.org/10.1016/s0896-6273\(03\)00255-1](https://doi.org/10.1016/s0896-6273(03)00255-1).
39. Basak, R., and Narayanan, R. (2018). Spatially dispersed synapses yield sharply-tuned place cell responses through dendritic spike initiation. *J. Physiol.* 596, 4173–4205. <https://doi.org/10.1113/JP275310>.
40. Basak, R., and Narayanan, R. (2020). Robust emergence of sharply tuned place-cell responses in hippocampal neurons with structural and biophysical heterogeneities. *Brain Struct. Funct.* 225, 567–590. <https://doi.org/10.1007/s00429-019-02018-0>.
41. Padmanabhan, K., and Urban, N.N. (2010). Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nat. Neurosci.* 13, 1276–1282. <https://doi.org/10.1038/nn.2630>.
42. Tripathy, S.J., Padmanabhan, K., Gerkin, R.C., and Urban, N.N. (2013). Intermediate intrinsic diversity enhances neural population coding. *Proc. Natl. Acad. Sci. USA* 110, 8248–8253. <https://doi.org/10.1073/pnas.1221214110>.
43. Saini, S., and Narayanan, R. (2025). Degeneracy Explains Diversity in Interneuron Regulation of Pattern Separation in Heterogeneous Dentate Gyrus Networks. *Function* 6, zqaf035. <https://doi.org/10.1093/function/zqaf035>.
44. Gast, R., Solla, S.A., and Kennedy, A. (2024). Neural heterogeneity controls computations in spiking neural networks. *Proc. Natl. Acad. Sci. USA* 121, e2311885121. <https://doi.org/10.1073/pnas.2311885121>.
45. Mittal, D., and Narayanan, R. (2021). Resonating neurons stabilize heterogeneous grid-cell networks. *eLife* 10, e66804. <https://doi.org/10.7554/eLife.66804>.
46. Shridhar, S., Mishra, P., and Narayanan, R. (2022). Dominant role of adult neurogenesis-induced structural heterogeneities in driving plasticity heterogeneity in dentate gyrus granule cells. *Hippocampus* 32, 488–516. <https://doi.org/10.1002/hipo.23422>.
47. Josselyn, S.A., and Frankland, P.W. (2018). Memory Allocation: Mechanisms and Function. *Annu. Rev. Neurosci.* 41, 389–413. <https://doi.org/10.1146/annurev-neuro-080317-061956>.
48. Lau, J.M.H., Rashid, A.J., Jacob, A.D., Frankland, P.W., Schacter, D.L., and Josselyn, S.A. (2020). The role of neuronal excitability, allocation to an engram and memory linking in the behavioral generation of a false memory in mice. *Neurobiol. Learn. Mem.* 174, 107284. <https://doi.org/10.1016/j.nlm.2020.107284>.
49. Park, S., Kramer, E.E., Mercaldo, V., Rashid, A.J., Insel, N., Frankland, P.W., and Josselyn, S.A. (2016). Neuronal Allocation to a Hippocampal Engram. *Neuropsychopharmacology* 41, 2987–2993. <https://doi.org/10.1038/npp.2016.73>.
50. Rao-Ruiz, P., Yu, J., Kushner, S.A., and Josselyn, S.A. (2019). Neuronal competition: microcircuit mechanisms define the sparsity of the engram. *Curr. Opin. Neurobiol.* 54, 163–170. <https://doi.org/10.1016/j.conb.2018.10.013>.
51. Yiu, A.P., Mercaldo, V., Yan, C., Richards, B., Rashid, A.J., Hsiang, H.L.L., Pressey, J., Mahadevan, V., Tran, M.M., Kushner, S.A., et al. (2014). Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron* 83, 722–735. <https://doi.org/10.1016/j.neuron.2014.07.017>.
52. Wang, X.J., and Buzsáki, G. (1996). Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *J. Neurosci.* 16, 6402–6413. <https://doi.org/10.1523/JNEUROSCI.16-20-06402.1996>.
53. Tikidji-Hamburyan, R.A., Martínez, J.J., White, J.A., and Canavier, C.C. (2015). Resonant Interneurons Can Increase Robustness of Gamma Oscillations. *J. Neurosci.* 35, 15682–15695. <https://doi.org/10.1523/JNEUROSCI.2601-15.2015>.
54. Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* 11, 1621–1671. <https://doi.org/10.1162/089976699300016179>.
55. Maex, R., and De Schutter, E. (2003). Resonant synchronization in heterogeneous networks of inhibitory neurons. *J. Neurosci.* 23, 10503–10514. <https://doi.org/10.1523/JNEUROSCI.23-33-10503.2003>.
56. White, J.A., Chow, C.C., Ritt, J., Soto-Treviño, C., and Kopell, N. (1998). Synchronization and oscillatory dynamics in heterogeneous, mutually

- inhibited neurons. *J. Comput. Neurosci.* 5, 5–16. <https://doi.org/10.1023/a:1008841325921>.
57. Perez-Nieves, N., Leung, V.C.H., Dragotti, P.L., and Goodman, D.F.M. (2021). Neural heterogeneity promotes robust learning. *Nat. Commun.* 12, 5791. <https://doi.org/10.1038/s41467-021-26022-3>.
58. Marom, S., and Marder, E. (2023). A biophysical perspective on the resilience of neuronal excitability across timescales. *Nat. Rev. Neurosci.* 24, 640–652. <https://doi.org/10.1038/s41583-023-00730-9>.
59. Alonso, L.M., Rue, M.C.P., and Marder, E. (2023). Gating of homeostatic regulation of intrinsic excitability produces cryptic long-term storage of prior perturbations. *Proc. Natl. Acad. Sci. USA* 120, e2222016120. <https://doi.org/10.1073/pnas.2222016120>.
60. Gorur-Shandilya, S., Cronin, E.M., Schneider, A.C., Haddad, S.A., Rosenbaum, P., Bucher, D., Nadim, F., and Marder, E. (2022). Mapping circuit dynamics during function and dysfunction. *eLife* 11, e76579. <https://doi.org/10.7554/eLife.76579>.
61. Marder, E., Kedia, S., and Morozova, E.O. (2022). New insights from small rhythmic circuits. *Curr. Opin. Neurobiol.* 76, 102610. <https://doi.org/10.1016/j.conb.2022.102610>.
62. Ratliff, J., Franci, A., Marder, E., and O'Leary, T. (2021). Neuronal oscillator robustness to multiple global perturbations. *Biophys. J.* 120, 1454–1468. <https://doi.org/10.1016/j.bpj.2021.01.038>.
63. Schapiro, K., and Marder, E. (2024). Resilience of circuits to environmental challenge. *Curr. Opin. Neurobiol.* 87, 102885. <https://doi.org/10.1016/j.conb.2024.102885>.
64. Mishra, P., and Narayanan, R. (2021). Ion-channel regulation of response decorrelation in a heterogeneous multi-scale model of the dentate gyrus. *Curr. Res. Neurobiol.* 2, 100007. <https://doi.org/10.1016/j.crneur.2021.100007>.
65. Rich, S., Moradi Chameh, H., Lefebvre, J., and Valiante, T.A. (2022). Loss of neuronal heterogeneity in epileptogenic human tissue impairs network resilience to sudden changes in synchrony. *Cell Rep.* 39, 110863. <https://doi.org/10.1016/j.celrep.2022.110863>.
66. Planert, H., Mittermaier, F.X., Grosser, S., Fidzinski, P., Schneider, U.C., Radbruch, H., Onken, J., Holtkamp, M., Schmitz, D., Alle, H., et al. (2026). Electrophysiological classification of human layer 2-3 pyramidal neurons reveals subtype-specific synaptic interactions. *Nat. Neurosci.* 29, 455–466. <https://doi.org/10.1038/s41593-025-02134-7>.
67. Goaillard, J.M., and Marder, E. (2021). Ion Channel Degeneracy, Variability, and Covariation in Neuron and Circuit Resilience. *Annu. Rev. Neurosci.* 44, 335–357. <https://doi.org/10.1146/annurev-neuro-092920-121538>.
68. Marder, E., and Goaillard, J.M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nat. Rev. Neurosci.* 7, 563–574. <https://doi.org/10.1038/nrn1949>.
69. Proix, T., Bartolomei, F., Guye, M., and Jirsa, V.K. (2017). Individual brain structure and modelling predict seizure propagation. *Brain* 140, 641–654. <https://doi.org/10.1093/brain/awx004>.
70. Hashemi, M., Depannemaeker, D., Saggio, M., Triebkorn, P., Rabuffo, G., Fousek, J., Ziaemehr, A., Sip, V., Athanasiadis, A., Breyton, M., et al. (2026). Principles and Operation of Virtual Brain Twins. *IEEE Rev. Biomed. Eng.* 19, 111–139. <https://doi.org/10.1109/RBME.2025.3562951>.
71. Wang, H.E., Dollomaja, B., Triebkorn, P., Duma, G.M., Williamson, A., Makhalova, J., Lemarechal, J.D., Bartolomei, F., and Jirsa, V. (2025). Virtual brain twins for stimulation in epilepsy. *Nat. Comput. Sci.* 5, 754–768. <https://doi.org/10.1038/s43588-025-00841-6>.
72. Wang, H.E., Triebkorn, P., Breyton, M., Dollomaja, B., Lemarechal, J.D., Petkoski, S., Sorrentino, P., Depannemaeker, D., Hashemi, M., and Jirsa, V.K. (2024). Virtual brain twins: from basic neuroscience to clinical use. *Natl. Sci. Rev.* 11, nwae079. <https://doi.org/10.1093/nsr/nwae079>.
73. Schaeffer, R., Khona, M., and Fiete, I.R. (2024). No free lunch from deep learning in Neuroscience: a case study through models of the entorhinal-hippocampal circuit. In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (Curran Associates Inc.).
74. Mishra, P., and Narayanan, R. (2021). Stable continual learning through structured multiscale plasticity manifolds. *Curr. Opin. Neurobiol.* 70, 51–63. <https://doi.org/10.1016/j.conb.2021.07.009>.
75. Jedlicka, P., Bird, A.D., and Cuntz, H. (2022). Pareto optimality, economy-effectiveness trade-offs and ion channel degeneracy: improving population modelling for single neurons. *Open Biol.* 12, 220073. <https://doi.org/10.1098/rsob.220073>.
76. Pallasdies, F., Norton, P., Schleimer, J.H., and Schreiber, S. (2021). Neural optimization: Understanding trade-offs with Pareto theory. *Curr. Opin. Neurobiol.* 71, 84–91. <https://doi.org/10.1016/j.conb.2021.08.008>.
77. Alon, U. (2019). *An Introduction to Systems Biology: Design Principles of Biological Circuits, Second Edition* (Chapman and Hall/CRC Press).
78. Miconi, T. (2017). Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks. *eLife* 6, e20899. <https://doi.org/10.7554/eLife.20899>.
79. Edelman, G.M., and Gally, J.A. (2001). Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. USA* 98, 13763–13768. <https://doi.org/10.1073/pnas.231499798>.
80. Kim, J., and Wilhelm, T. (2008). What is a complex graph? *Phys. Stat. Mech. Appl.* 387, 2637–2652. <https://doi.org/10.1016/j.physa.2008.01.015>.
81. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827. <https://doi.org/10.1126/science.298.5594.824>.
82. Watts, D.J., and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442. <https://doi.org/10.1038/30918>.
83. Luo, L. (2021). Architectures of neuronal circuits. *Science* 373, eabg7285. <https://doi.org/10.1126/science.abg7285>.
84. Kuan, A.T., Bondanelli, G., Driscoll, L.N., Han, J., Kim, M., Hildebrand, D.G.C., Graham, B.J., Wilson, D.E., Thomas, L.A., Panzeri, S., et al. (2024). Synaptic wiring motifs in posterior parietal cortex support decision-making. *Nature* 627, 367–373. <https://doi.org/10.1038/s41586-024-07088-7>.
85. Seenivasan, P., and Narayanan, R. (2022). Efficient information coding and degeneracy in the nervous system. *Curr. Opin. Neurobiol.* 76, 102620. <https://doi.org/10.1016/j.conb.2022.102620>.
86. Huang, A., Singh, S.H., Martinelli, F., and Rajan, K. (2025). Measuring and Controlling Solution Degeneracy across Task-Trained Recurrent Neural Networks. Conference paper at. *Advances in Neural Information Processing Systems* 38. <https://doi.org/10.52202/085713-3517>.
87. Pezon, L., Schmutz, V., and Gerstner, W. (2026). Linking neural manifolds to circuit structure in recurrent networks. *Neuron* 114, 1682–1694.e21. <https://doi.org/10.1016/j.neuron.2025.12.047>.
88. Mishra, P., and Narayanan, R. (2021). Ion-channel degeneracy: Multiple ion channels heterogeneously regulate intrinsic physiology of rat hippocampal granule cells. *Physiol. Rep.* 9, e14963. <https://doi.org/10.14814/phy2.14963>.
89. Goldman, M.S., Golowasch, J., Marder, E., and Abbott, L.F. (2001). Global structure, robustness, and modulation of neuronal models. *J. Neurosci.* 21, 5229–5238. <https://doi.org/10.1523/JNEUROSCI.21-14-05229.2001>.
90. Chavlis, S., and Poirazi, P. (2025). Dendrites endow artificial neural networks with accurate, robust and parameter-efficient learning. *Nat. Commun.* 16, 943. <https://doi.org/10.1038/s41467-025-56297-9>.
91. Pagkalos, M., Makarov, R., and Poirazi, P. (2024). Leveraging dendritic properties to advance machine learning and neuro-inspired computing. *Curr. Opin. Neurobiol.* 85, 102853. <https://doi.org/10.1016/j.conb.2024.102853>.
92. Song, H.F., Yang, G.R., and Wang, X.J. (2016). Training Excitatory-Inhibitory Recurrent Neural Networks for Cognitive Tasks: A Simple and Flexible Framework. *PLoS Comput. Biol.* 12, e1004792. <https://doi.org/10.1371/journal.pcbi.1004792>.
93. Khona, M., and Fiete, I.R. (2022). Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* 23, 744–766. <https://doi.org/10.1038/s41583-022-00642-0>.
94. Abbott, L.F., Rajan, K., and Sompolinsky, H. (2011). Interactions between Intrinsic and Stimulus-Evoked Activity in Recurrent Neural Networks. In

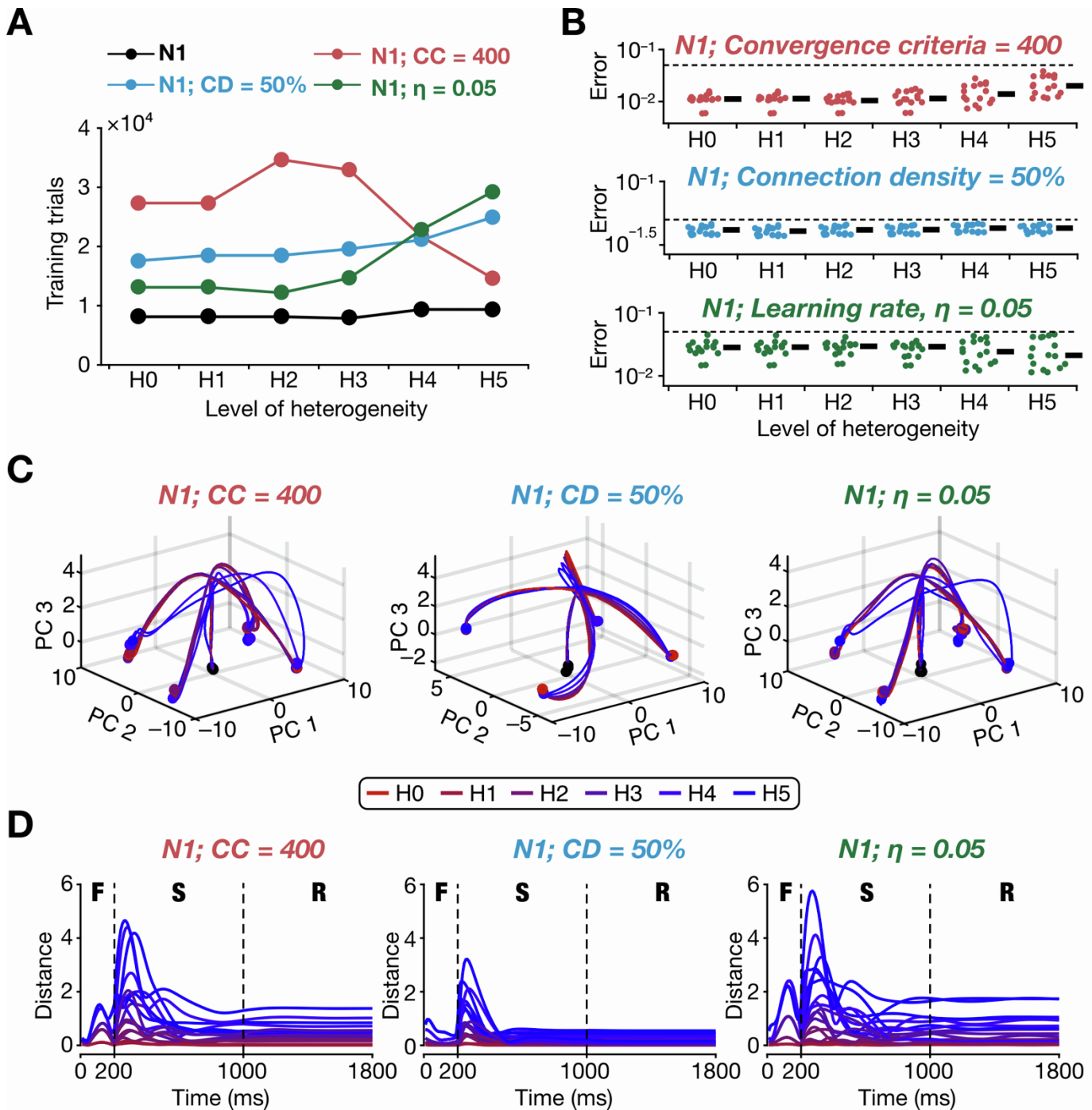
- The Dynamic Brain: An Exploration of Neuronal Variability and Its Functional Significance, P.M. Ding and P.D. Glanzman, eds. (Oxford University Press). <https://doi.org/10.1093/acprof:oso/9780195393798.003.0004>.
95. Litwin-Kumar, A., Harris, K.D., Axel, R., Sompolinsky, H., and Abbott, L.F. (2017). Optimal Degrees of Synaptic Connectivity. *Neuron* 93, 1153–1164.e7. <https://doi.org/10.1016/j.neuron.2017.01.030>.
96. Mazzucato, L., Fontanini, A., and La Camera, G. (2016). Stimuli Reduce the Dimensionality of Cortical Activity. *Front. Syst. Neurosci.* 10, 11. <https://doi.org/10.3389/fnsys.2016.00011>.
97. Santhosh, A., and Narayanan, R. (2026). Source code for assessing the impact of heterogeneities on recurrent networks performing different cognitive tasks. Zenodo. <https://doi.org/10.5281/zenodo.17750827>.

Patterns, Volume 7

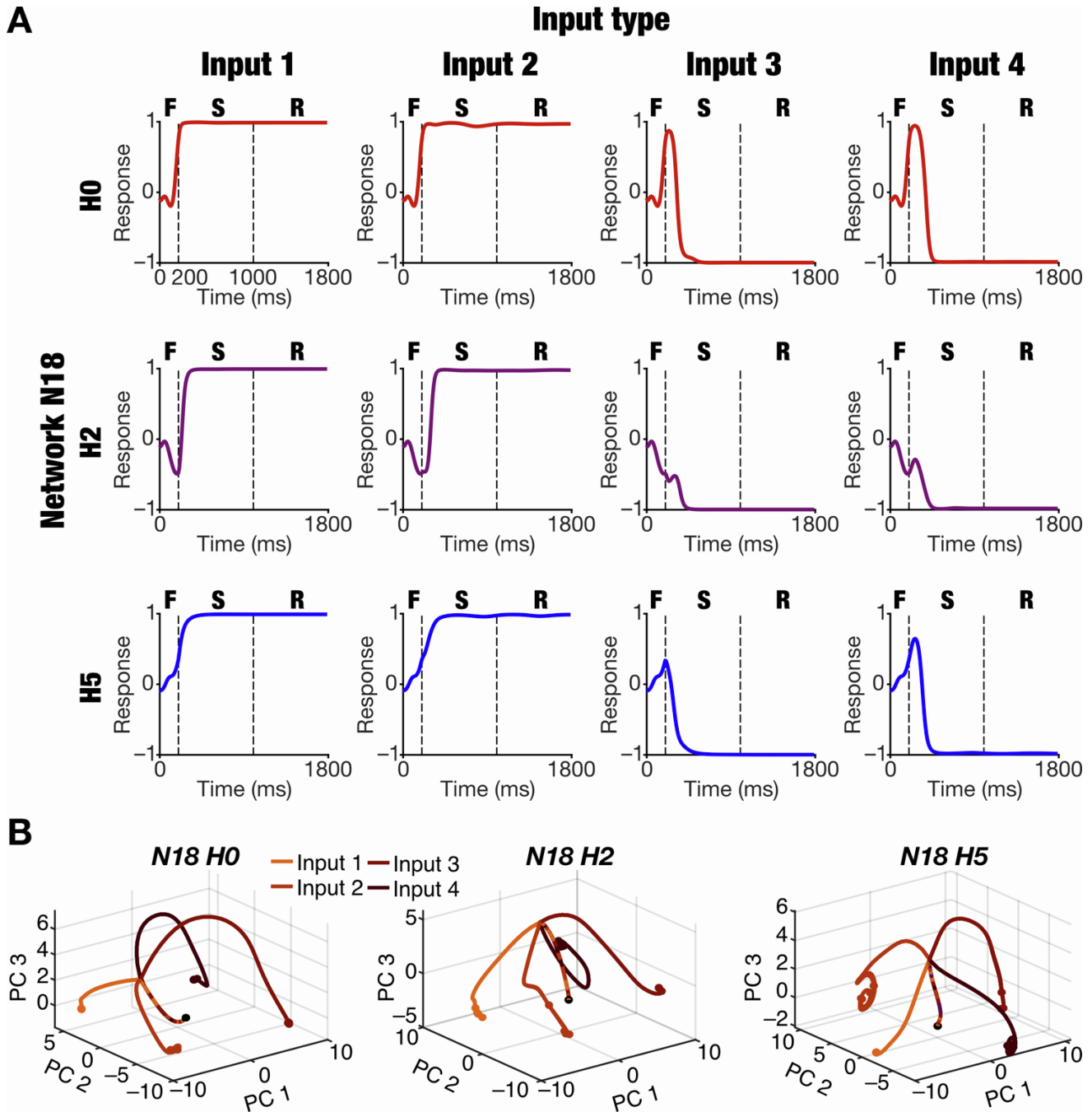
Supplemental information

Diversity in the impact of heterogeneities on recurrent networks performing a cognitive task

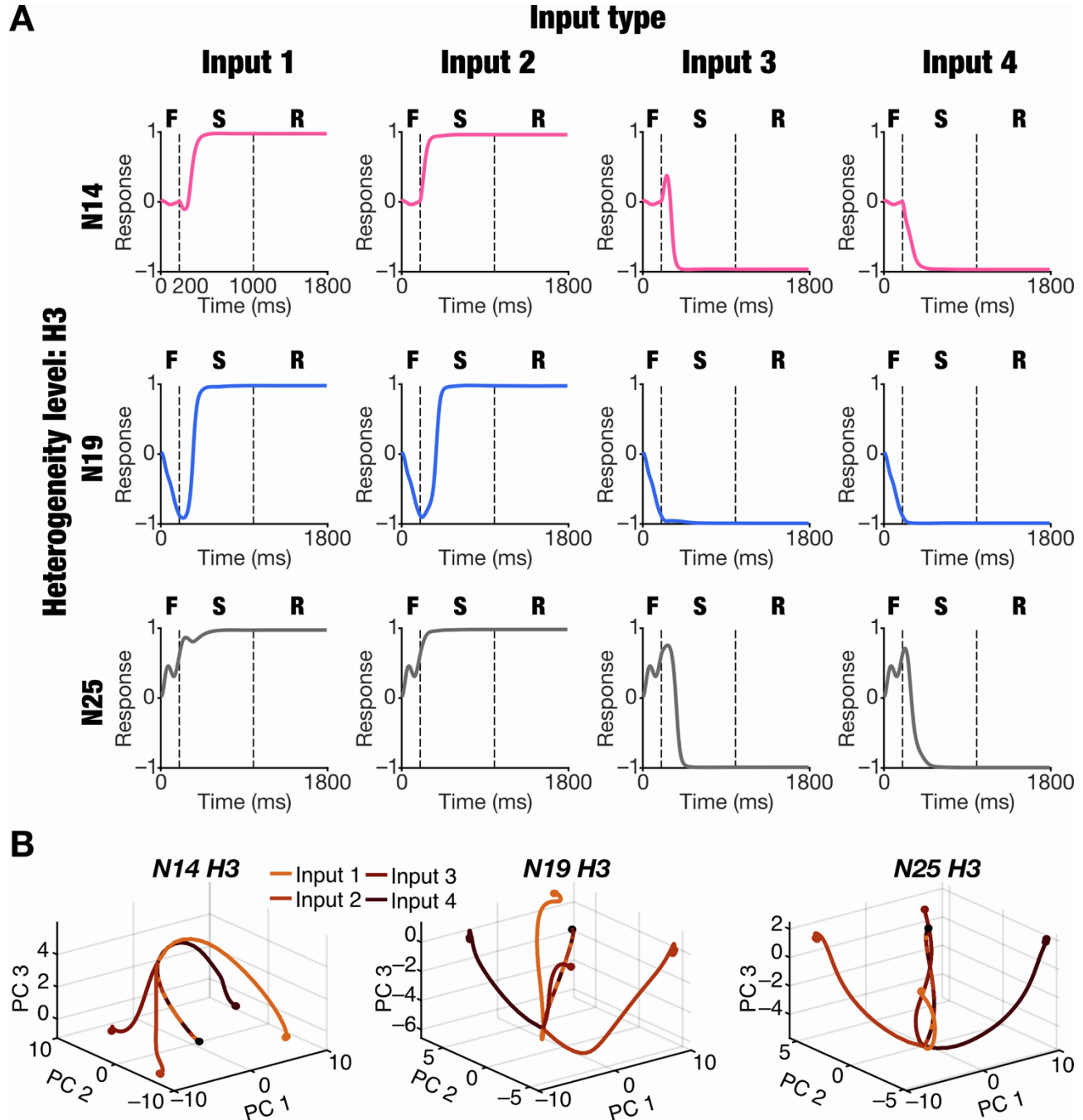
Anjana Santhosh and Rishikesh Narayanan



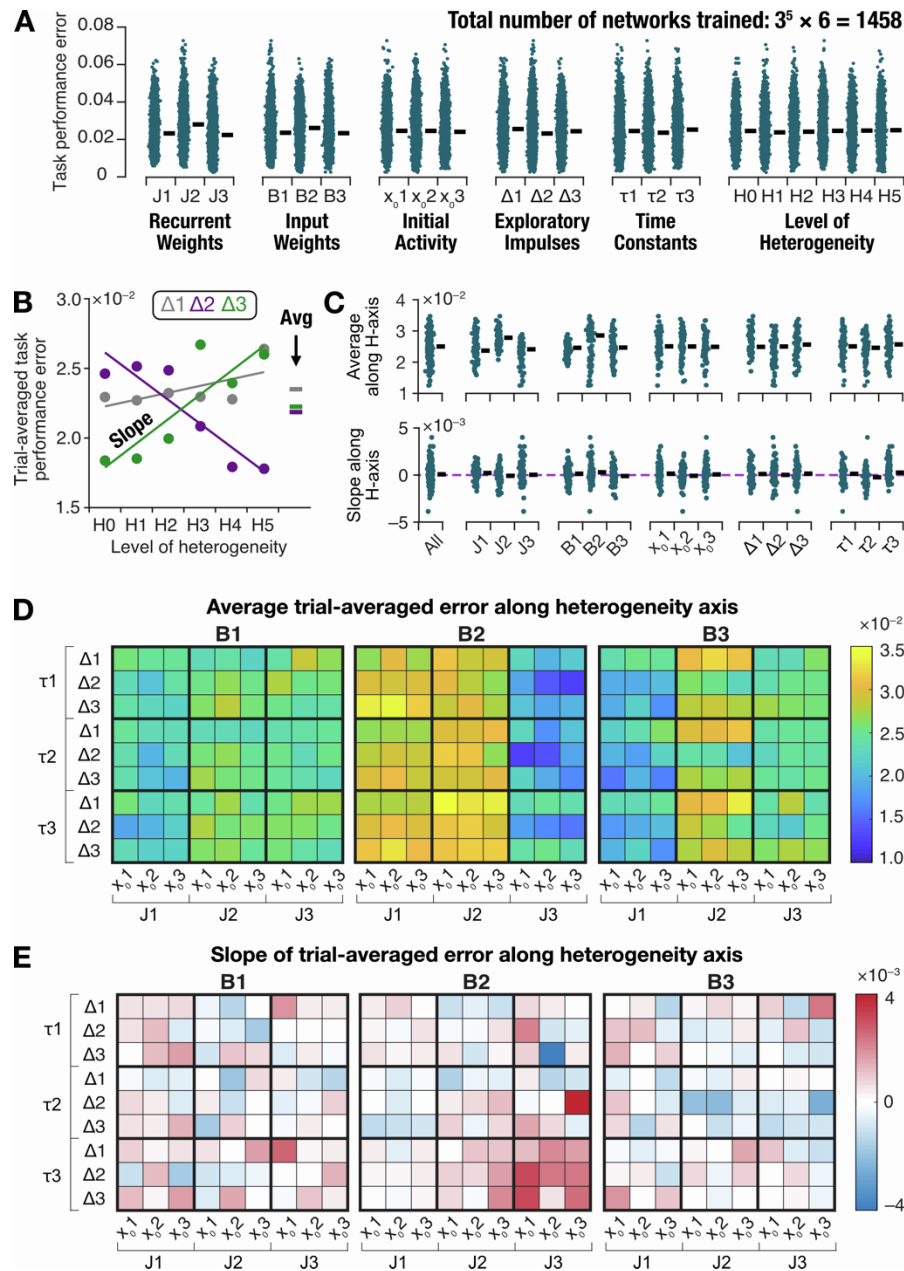
Supplemental Figure S1. Training performance and dynamics of heterogeneous recurrent networks with more stringent hyperparameters. (A) Number of trials taken to train network N1 to reach convergence criteria with default hyperparameters (N1), with more stringent convergence criteria (N1; CC=400), with reduced connection probability between neurons (N1; CD=50%), and reduced learning rate (N1; $\eta=0.05$). For the default network (N1), the convergence criterion required error to be below threshold for 100 trials, the connection probability was 100%, and learning rate $\eta=0.15$. (B) Bee-swarm plot of distribution of errors for different levels of training heterogeneity with each of the different groups. The black dashed line represents the threshold error (0.05) below which the trial is considered a correct trial. Solid lines represent the median values for each heterogeneity level. (C) CCA-aligned trajectories of networks with the different stringent training criteria, with all 6 levels of heterogeneities (H0–H5). (D) The distance between latent-space trajectories for all levels of heterogeneity with respect to their respective H0 network shown for all training heterogeneity levels. F: Fixation epoch, S: Stimulus epoch, R: Response epoch.



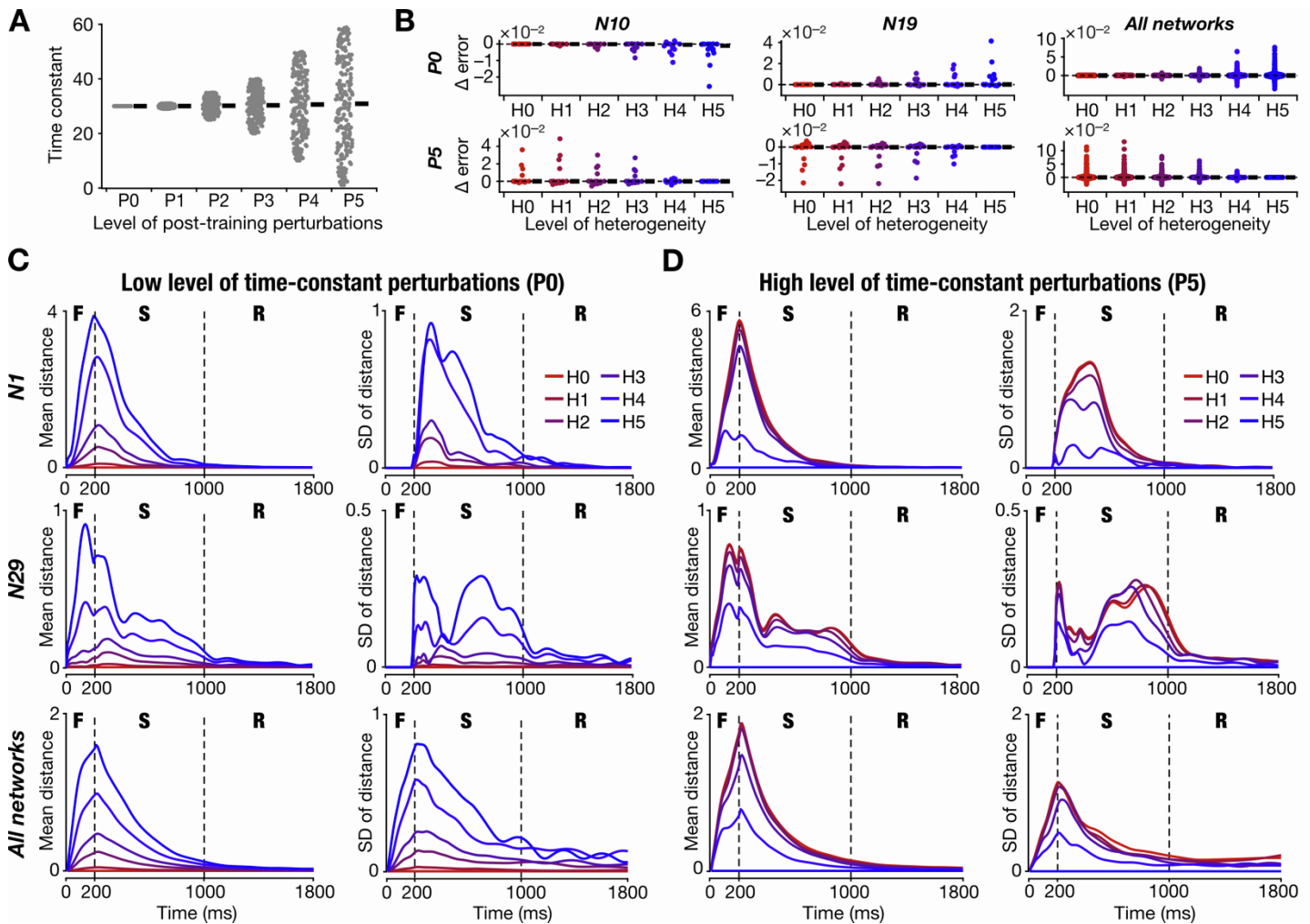
Supplemental Figure S2. Dynamics of recurrent networks with the same hyperparameters and different levels of training heterogeneities. (A) Temporal evolution of the response of the output unit in network N18 trained with three different levels of heterogeneities (H0, H2, and H5). Shown are responses of the output unit to the four distinct input types (**Input 1:** $u_{stim1} = 1$, $u_{stim2} = 0$; **Input 2:** $u_{stim1} = 0$, $u_{stim2} = 1$; **Input 3:** $u_{stim1} = -1$, $u_{stim2} = 0$; **Input 4:** $u_{stim1} = 0$, $u_{stim2} = -1$). It may be noted the output unit has converged to its respective expected output values in each of the four input types across all levels of heterogeneities. The variability in the temporal dynamics across different levels of heterogeneities, especially during the fixation epoch and the initial part of the stimulus epoch, may also be noted. F: Fixation epoch, S: Stimulus epoch, R: Response epoch. (B) The dynamics of all units in the network are represented by input-dependent trajectories in their respective reduced dimensional spaces, computed for each network using principal component analysis (PCA). These are shown for the same networks that are depicted in panel A. The input-dependence of the trajectories and the fixed points may be noted. Variability in the locations of fixed points and in the trajectories across the different levels of heterogeneities may also be noted.



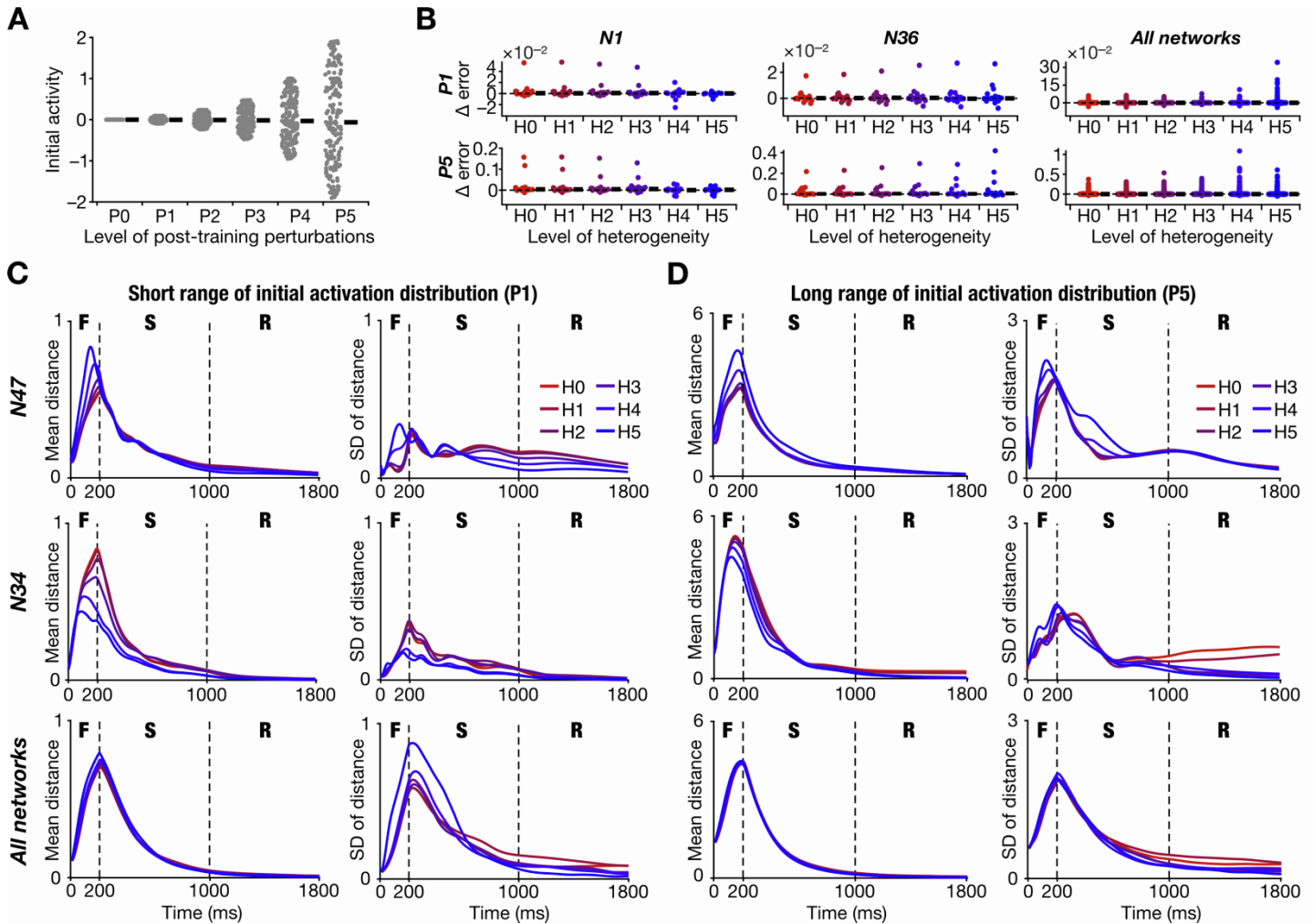
Supplemental Figure S3. Dynamics of recurrent networks with distinct hyperparameters and same level of training heterogeneities. (A) Temporal evolution of the response of the output unit in three different networks (N14, N19, and N25) trained with the same levels of heterogeneity (H3). Shown are responses of the output unit to the four distinct input types (**Input 1:** $u_{stim1} = 1$, $u_{stim2} = 0$; **Input 2:** $u_{stim1} = 0$, $u_{stim2} = 1$; **Input 3:** $u_{stim1} = -1$, $u_{stim2} = 0$; **Input 4:** $u_{stim1} = 0$, $u_{stim2} = -1$). It may be noted the output unit has converged to its respective expected output values in each of the four input types across all networks. The variability in the temporal dynamics across different networks, especially during the fixation epoch and the initial part of the stimulus epoch, may also be noted. F: Fixation epoch, S: Stimulus epoch, R: Response epoch. (B) The dynamics of all units in the network are represented by input-dependent trajectories in their respective reduced dimensional spaces, computed for each network using principal component analysis (PCA). These are shown for the same networks that are depicted in panel A. The input-dependence of the trajectories and the fixed points may be noted. Variability in the locations of fixed points and in the trajectories across the different networks may also be noted.



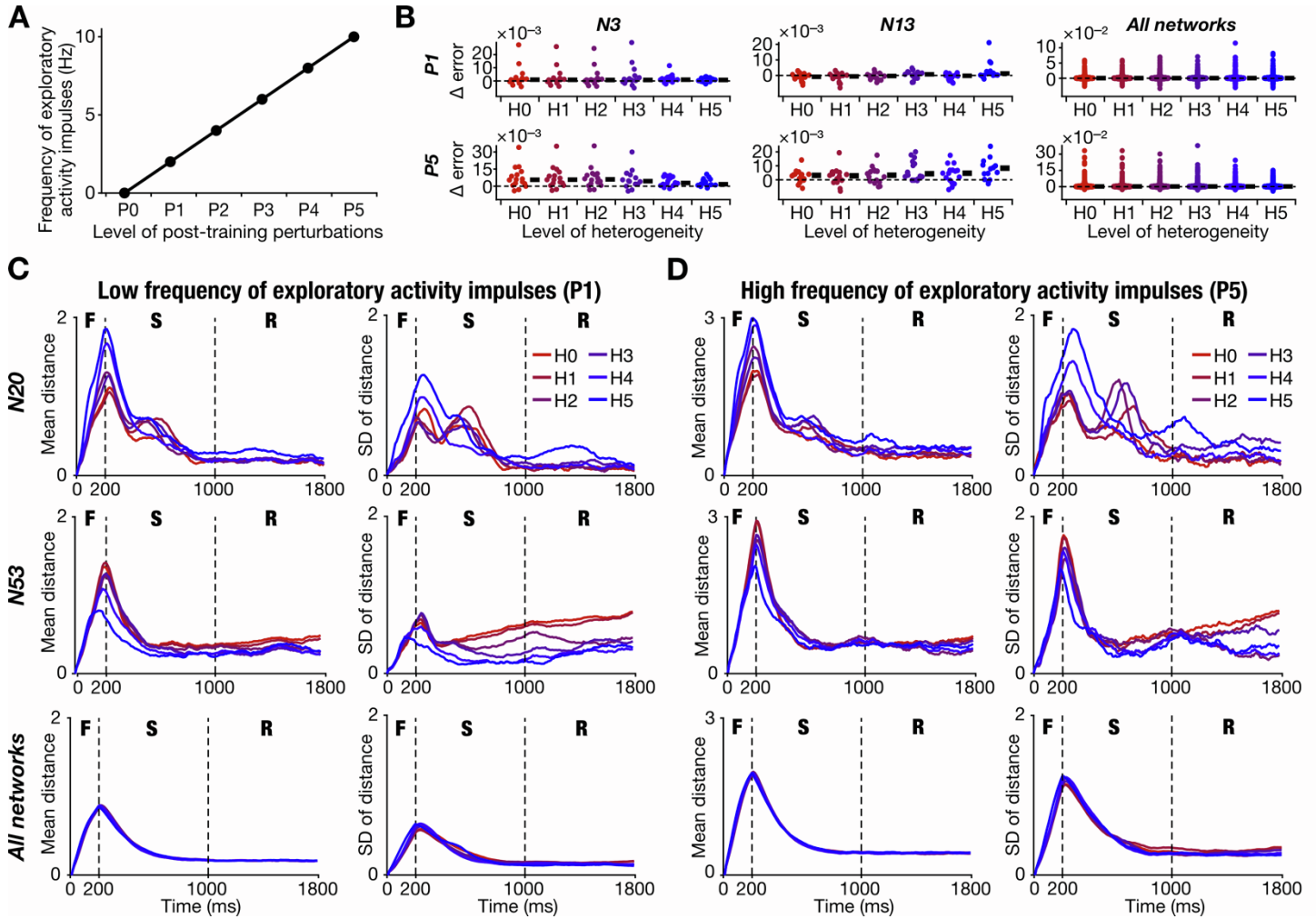
Supplemental Figure S4. Variability in the impact of different network hyperparameters on performance. (A) Variability in the errors in all trials plotted across each value of the five hyperparameters and heterogeneity levels. The black lines depict the median errors across all the networks with the corresponding value of hyperparameter ($N=7776$ for each of the five hyperparameters and 3888 for each heterogeneity level). (B) Mean error across all 16 trials in an example network J1–B1– x_0 3– τ 3 with Δ_1 (gray), Δ_2 (purple), and Δ_3 (green). The mean errors (dots) and the best linear fits (lines) are plotted along the heterogeneity axis (H0–H5). The dashes represent the average errors across all trials and heterogeneity levels. (C) Distribution of average errors across all trials and heterogeneity level (top) and slope of mean errors as a function of heterogeneity level (bottom). The distribution is plotted for all networks together (All) and for different values of the five hyperparameters (J1 to τ 3). The black line depicts the median value of average (top) and slope (bottom) along the heterogeneity axis for each network. (D–E) Average (D) and slope (E) of mean errors across heterogeneity levels plotted as a heatmap for all combinations of the hyperparameter values. The three plots contain all the 81 networks with the initial input synaptic weights initialized as B1 (left), B2 (middle), and B3 (right). In each matrix, the x-axis depicts the three different values of recurrent weights (J1–J3) as the major subdivisions and three values of the initial activity values (x_0 1– x_0 3) as the minor subdivisions. Similarly, the y-axis depicts the three different sets of time constant values (τ 1– τ 3) as the major subdivisions and three sets of impulse patterns (Δ 1– Δ 3) introduced during training as the minor subdivisions.



Supplemental Figure S5. Pronounced network-to-network variability in the robustness of heterogeneous recurrent networks to post-training intrinsic perturbations. (A) Distribution of time constants for six different levels of post-training heterogeneity (P0–P5). Note that the mean (black bar) remained the same, while the range of time constants increased as the level of post-training perturbations are increased. This is akin to heterogeneities introduced during training (H0–H5; Fig. 1) and constitutes a shift in level of post-training perturbations. Therefore, errors and distances in panels B–D were calculated with respect to the heterogeneity level the networks were trained with. For instance, when networks trained with H0–H5 are subjected to perturbation P5, the outcomes are compared with H5 network. (B) Distribution of the difference in response errors (across different trials) for low (P1; top row) and high (P5; bottom row) levels of post-training perturbations, introduced in networks trained with H0–H5 heterogeneities. The first column shows an example network (N10) that showed a reduction in error with increase in level of training heterogeneities (H0–H5). The second column depicts another network (N19) where the error increased in networks trained with higher level of heterogeneities. The third column summarizes the difference in errors across all 53 networks, spanning different trials. Black bars represent the median values. (C) Mean (left column) and standard deviation (SD; right column) of the distance distribution (computed from trajectories in the latent space) with respect to H0 for low level of post-training perturbations (P0). Each panel shows plots for networks trained with different levels of heterogeneity (H0–H5). The first row shows an example network (N1) where the mean distance and the associated variability were high. The second row depicts another network (N29) where the mean distance and the associated variability were low. The third row depicts mean distance and the associated variability across all 53 networks. (D) Same as C but for P5 instead of P0. For panels C and D, the error for the trained level of heterogeneity is low and increases progressively with increasing distance from the trained level of heterogeneity. F: Fixation epoch, S: Stimulus epoch, R: Response epoch.



Supplemental Figure S6. Pronounced network-to-network variability in the robustness of heterogeneous recurrent networks to post-training shift in network activity at task initiation. (A) Distribution of network activity range in all recurrent units at network initiation for six different levels of post-training perturbations (P0–P5). Note the shift in the range of the initial activity distribution (without shift in mean) with increasing in level of post-training perturbations. (B) Distribution of the difference in response errors (across different trials) with respect to P0 for low (P1; top row) and high (P5; bottom row) levels of post-training perturbations, introduced in networks trained with H0–H5 heterogeneities. The first column shows an example network (N1) that showed a reduction in error with increase in level of training heterogeneities (H0–H5). The second column depicts another network (N36) where the error was high in networks trained with higher level of heterogeneities. The third column summarizes the difference in errors across all 53 networks, spanning different trials. Black bars represent the median values. (C) Mean (left column) and standard deviation (SD; right column) of the distance distribution (computed from trajectories in the latent space) with respect to P0 for low level of post-training perturbations (P1). Each panel shows plots for networks trained with different levels of training heterogeneity (H0–H5). The first row shows an example network (N47) where the mean distance and the associated variability increased with increasing level of training heterogeneities. The second row depicts another network (N34) where the mean distance and the associated variability decreased with higher level of heterogeneities introduced during training. The third row depicts mean distance and the associated variability across all 53 networks. (D) Same as C but for P5 instead of P1. F: Fixation epoch, S: Stimulus epoch, R: Response epoch.



Supplemental Figure S7. Pronounced network-to-network variability in the robustness of heterogeneous recurrent networks to post-training perturbation in the frequency of exploratory activity impulses. (A) Increasing frequency of exploratory activity impulses introduced for six different levels of post-training perturbations (P0–P5). (B) Distribution of the difference in response errors (across different trials) with respect to P0 for low (P1; top row) and high (P5; bottom row) levels of post-training perturbations, introduced in networks trained with H0–H5 heterogeneities. The first column shows an example network (N3) that showed a reduction in error with increase in level of training heterogeneities (H0–H5). The second column depicts another network (N13) where the error was high in networks trained with higher level of heterogeneities. The third column summarizes the difference in errors across all 53 networks, spanning different trials. Black bars represent the median values. (C) Mean (left column) and standard deviation (SD; right column) of the distance distribution (computed from trajectories in the latent space) with respect to P0 for low level of post-training perturbations (P1). Each panel shows plots for networks trained with different levels of training heterogeneity (H0–H5). The first row shows an example network (N20) where the mean distance and the associated variability increased with increasing level of training heterogeneities. The second row depicts another network (N53) where the mean distance and the associated variability decreased with higher level of heterogeneities introduced during training. It may be noted that the variability in distances were high for the N53 network even during the response epoch, especially when trained with low level of heterogeneities (H0–H1). The third row depicts mean distance and the associated variability across all 53 networks. (D) Same as C but for P5 instead of P1. F: Fixation epoch, S: Stimulus epoch, R: Response epoch.

Supplemental Table S1. Seed values used to initialize the three sets of hyperparameters for Figs. 5–6 and Supplemental Fig. S4.

Hyperparameter	Seed values
Initial recurrent connectivity matrix (J)	$J1 = 1$
	$J2 = 104$
	$J3 = 1509$
Initial input connectivity matrix (B)	$B1 = 2$
	$B2 = 200$
	$B3 = 1507$
Initial value of activity vector ($x(0)$)	$x_{01} = 3$
	$x_{02} = 300$
	$x_{03} = 1505$
Pattern of exploratory impulses during training (Δ)	$\Delta1 = 4$
	$\Delta2 = 400$
	$\Delta3 = 1503$
Value of time constants (τ)	$\tau1 = 5$
	$\tau2 = 500$
	$\tau3 = 1501$

Supplemental Table S2. Different levels of the post-training perturbations and their impact on task performance.

Type of post-training perturbations	Description of perturbations	Level of perturbations	Values	Impact on performance
Mean of time constant distribution (Fig. 7)	Time constants of each unit, τ , are shifted by τ_{shift} such that the new time constants are $\tau + \tau_{shift}$	P0	$\tau_{shift} = 0$	Graded increases in performance error and mean/variance in distances of latent-space trajectories through P1–P5.
		P1	$\tau_{shift} = 1$	
		P2	$\tau_{shift} = 5$	
		P3	$\tau_{shift} = 10$	
		P4	$\tau_{shift} = 20$	
		P5	$\tau_{shift} = 40$	
Range of time constant distribution (Supplemental Fig. S5)	τ is sampled from a uniform distribution spanning τ_{range} with the same seed value used for training	P0	$\tau_{range} = 30$	Minimal impact on performance error. Mean/variance in distances of latent-space trajectories varied, but did not have a strong impact on response epochs.
		P1	$\tau_{range} = [29,31]$	
		P2	$\tau_{range} = [25,35]$	
		P3	$\tau_{range} = [20,40]$	
		P4	$\tau_{range} = [10,50]$	
		P5	$\tau_{range} = [1,59]$	
Noise in recurrent synaptic weight (Fig. 8)	Noise J^{noise} sampled from a normal distribution $\mathcal{N}(0, \sigma^2)$ were added to the recurrent synaptic weights, J	P0	$\sigma = 0$	Most detrimental among all tested perturbations. Graded increases in performance error and mean/variance in distances of latent-space trajectories through P1–P5.
		P1	$\sigma = 0.01$	
		P2	$\sigma = 0.02$	
		P3	$\sigma = 0.03$	
		P4	$\sigma = 0.04$	
		P5	$\sigma = 0.05$	
Initial activity distribution mean (Fig. 9)	$x(0)$ is sampled from a uniform distribution of range $x_{shift} + [-0.1, 0.1]$	P0	$x_{shift} = 0$	Minimal impact on performance error. Mean/variance in distances of latent-space trajectories varied, but did not have a strong impact on response epochs.
		P1	$x_{shift} = 0.1$	
		P2	$x_{shift} = 0.25$	
		P3	$x_{shift} = 0.5$	
		P4	$x_{shift} = 1$	
		P5	$x_{shift} = 2$	
Initial activity distribution range (Supplemental Fig. S8)	$x(0)$ is sampled from a uniform distribution of range $[-x_{range}, x_{range}]$	P0	$x_{range} = 0$	Minimal impact on performance error. Mean/variance in distances of latent-space trajectories varied, but did not have a strong impact on response epochs.
		P1	$x_{range} = 0.1$	
		P2	$x_{range} = 0.25$	
		P3	$x_{range} = 0.5$	
		P4	$x_{range} = 1$	
		P5	$x_{range} = 2$	
Stimulus epoch duration (Fig. 10)	The duration of the stimulus epoch was set to T_{SE}	P0	$T_{SE} = 800$	Graded increases in performance error and mean/variance in distances of latent-space trajectories through P1–P5.
		P1	$T_{SE} = 600$	
		P2	$T_{SE} = 400$	
		P3	$T_{SE} = 200$	
		P4	$T_{SE} = 100$	
		P5	$T_{SE} = 50$	
Frequency of exploratory activity impulses (Supplemental Fig. S7)	Random exploratory activity impulses, Δ , are introduced to each unit with the frequency f_{Δ}	P0	$f_{\Delta} = 0$	Minimal impact on performance error. Mean/variance in distances of latent-space trajectories varied, but did not have a strong impact on response epochs.
		P1	$f_{\Delta} = 2$	
		P2	$f_{\Delta} = 4$	
		P3	$f_{\Delta} = 6$	
		P4	$f_{\Delta} = 8$	
		P5	$f_{\Delta} = 10$	