Residual dynamics resolves recurrent contributions to neural computation

Aniruddh R. Galgali[1,2], Maneesh Sahani[3] and Valerio Mante[1,2]

[1] Institute of Neuroinformatics, University of Zurich & ETH Zurich, Zurich, Switzerland
[2] Neuroscience Center Zurich, University of Zurich & ETH Zurich, Zurich, Switzerland
[3] Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom
Corresponding Authors: galgalia@ini.uzh.ch, valerio@ini.uzh.ch

Abstract

Relating neural activity to behavior requires an understanding of how neural computations arise from the coordinated dynamics of distributed, recurrently connected neural populations. However, inferring the nature of recurrent dynamics from partial recordings of a neural circuit presents significant challenges. Here, we show that some of these challenges can be overcome by a fine-grained analysis of the dynamics of neural residuals, i.e. trial-by-trial variability around the mean neural population trajectory for a given task condition. Residual dynamics in macaque pre-frontal cortex (PFC) in a saccade-based perceptual decision-making task reveals recurrent dynamics that is time-dependent, but consistently stable, and implies that pronounced rotational structure in PFC trajectories during saccades are driven by inputs from upstream areas. The properties of residual dynamics restrict the possible contributions of PFC to decision-making and saccade generation, and suggest a path towards fully characterizing distributed neural computations with large-scale neural recordings and targeted causal perturbations.

Introduction

Perception, decisions, and the resulting actions reflect neural computations implemented by large, interacting neuronal populations acting in concert. Inferring the nature of these interactions from recordings of neural activity is a key step towards uncovering the neural computations underlying behavior[1–4]. One promising approach is based on the premise that neural computations reflect the action of a dynamical system[5–7], whereby the computations implemented by a neural population emerge from the interplay between external inputs into a distributed neural population and the internal dynamics resulting from the recurrent connections between neurons. The utility of such a "computation-through-dynamics" framework hinges critically on our ability to characterize the nature of this interplay, and disentangle the individual contributions of inputs and recurrent dynamics. In practice, disentangling these two factors based on recordings of neural responses alone is challenging, as typically neither the exact properties of the inputs into a brain area, nor the nature of recurrent connectivity within and across areas, are known a priori[8–11].

Here, we show that some of the challenges inherent to inferring the contribution of recurrent dynamics to neural responses can be overcome by analyzing the dynamical structure of neural population residuals, i.e. the trial-to-trial variability in neural population responses[12–22]. Our approach involves solving a statistical inference problem, but is ultimately based on the intuitive idea that the effect of recurrent computations can be revealed by observing how a local perturbation of the state of the neural population evolves over time[11,23–27]. Unlike in causal perturbation experiments, where the perturbations are generated externally[28–30], we rely entirely on an analysis of recorded response residuals, which we interpret as naturally occurring perturbations within the repertoire of neural patterns produced by a recurrent neural network[31,32]. We term the time-varying dynamics of response residuals as "residual dynamics", and show that in many settings it can resolve key properties of the recurrent dynamics underlying recorded neural responses. Obtaining a complete and quantitative description of residual dynamics is difficult, because neural population residuals are typically dominated by unstructured noise. To obtain reliable and unbiased estimates of residual dynamics, we thus developed novel statistical methods based on subspace identification[33,34] and instrumental variable regression[35].

Our findings are organized in three sections. First, we illustrate the challenges in disentangling inputs and recurrent dynamics based on the simulations of a few, simple dynamical systems (Fig. 1-2). These dynamical systems are analogous to single-area, artificial recurrent neural networks (RNN) previously proposed for explaining the network-level mechanisms underlying sensory evidence integration[36–42] and movement generation in cortical areas[43–47]. We demonstrate that our estimates of residual dynamics can reveal the essential features of the computations implemented by these models, even when the time-course of the inputs are unknown. Second, we study neural population recordings from pre-frontal cortex (PFC) of macaque monkeys during decision-making and saccadic choices (Fig. 3-5). While neural population trajectories in PFC are consistent with a number of previously proposed models of evidence integration and movement generation, we are able to rule out several candidate models based on the properties of the inferred residual dynamics. Third, we analyze simulated responses of a previously proposed multi-area RNN model of decision-making[48], to illustrate how inferred residual dynamics can be used to deduce circuit-level implementations of distributed recurrent computations. (Fig. 6-8).

Results

A prevalent approach for studying population-level neural computations relies on extracting low-dimensional neural trajectories from the population response[49–55]. The time-course of such trajectories and their dependency on task-variables can be compared to those generated by hand-designed[41,56–60] and task-optimized RNN models[42,46,47,61–66], or statistical models of neural dynamics[67–75]. Such an approach has been very successful in generating hypotheses about the nature of neural computations, but typically cannot unambiguously resolve the properties of recurrent dynamics based on the measured population responses[10,11]—estimating these properties is generally an ill-posed problem whenever other factors contributing to the responses, like external inputs, are unknown or unobserved.

Neural population trajectories poorly constrain recurrent computations

To illustrate the nature of this problem, we consider simulated responses of a number of distinct models of neural population dynamics during perceptual decision-making[40] and movement generation[43] (Fig. 1). While these hand-designed models are not meant to precisely reproduce neural recordings, they do capture the distinctive features of rather complex, non-linear RNNs trained to integrate sensory evidence towards a choice[41,42,76] (Fig. 1a) or generate complex motor sequences[46,47] (Fig. 1b).

In the models, the temporal evolution of the neural population response ($\mathbf{z}_t$) at time t is governed by a non-linear differential equation, which describes the momentary change in the response ($\dot{\mathbf{z}}_\mathrm{t}$) as resulting from the combined action of the *recurrent dynamics* ($\mathbf{F}$), the *input* ($\mathbf{u}_t$), and the *noise* ($\boldsymbol{\epsilon}_t$):

$$\dot{\mathbf{z}}_\mathrm{t} = \mathbf{F}(\mathbf{z}_t) + \mathbf{u}_t + \boldsymbol{\epsilon}_t \qquad (1)$$

Any solution to the above equation is also determined by the *initial condition* $\mathbf{z}_0$ (the neural state at the start of the trial). Differences in responses across task-conditions (e.g., different choices or movements) are explained by allowing $\mathbf{u}_t$ or $\mathbf{z}_0$ to vary across conditions (Fig. 1c, red vs. blue; Fig. 1b, initial conditions IC_1 vs. IC_2).

We simulated single-trial responses for two task-conditions and represented them as trajectories in a 2-dimensional neural state-space (Fig. 1a,b, choice 1 & 2; dark-gray curves). The recurrent dynamics ($\mathbf{F}$) can be represented as a flow field (Fig. 1a,b, black arrows and light-gray curves), which describes how the instantaneous neural state ($\mathbf{z}_t$) evolves from a given location in state-space in the absence of inputs and noise. The action of the external input ($\mathbf{u}_t$) corresponds to injecting a pattern of activity into the neural population, and therefore pushing the trajectory along a direction in state space that can vary both across

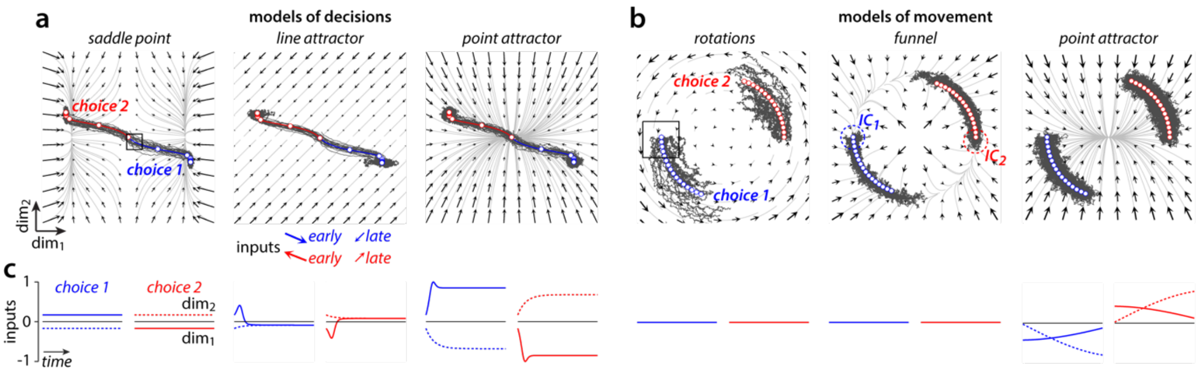


Fig. 1. Dynamical models of population-level computations underlying decisions and movements.

Each panel shows simulated single trials (dark-gray trajectories) and condition-averaged trajectories (blue and red trajectories) for two task conditions (choice 1 and 2). Black arrows show the effect of the recurrent dynamics on the response at any location in state-space. The effect of the inputs is constant across state-space, but can change over time and across task conditions (middle, example inputs at bottom). ***a****, Models of decision-making. Left: a model implementing a saddle point close to the initial conditions for both choice 1 and 2. Middle: a line attractor model. Right: a point-attractor model. The three models implement unstable (left), perfect (middle), and leaky integration (right) of an appropriately chosen input.* ***b****, Models of movement-generation. Left: purely rotational dynamics. Perturbations of the condition-averaged trajectory along both state-space dimensions are persistent; Middle: funnel model. Perturbations along the radial dimension decay, perturbations along the circular "channel" are persistent. Right: point attractor model. Responses are pushed away from the point attractor by strong inputs. IC: approximate extent of the initial conditions, shown as an example for the funnel model.* ***c****, Deterministic component of the inputs, for the models in* ***a*** *and* ***b****. Curves indicate the components of the input along the two state-space dimensions (solid vs dashed) as a function of time (horizontal axis) and condition (red vs blue). The inputs are chosen such that the different models of decision-making in* ***a****, and of movement-generation in* ***b****, cannot be distinguished based on the condition-averaged trajectories. Boxes in* ***a*** *and* ***b*** *(left) show the regions of state-space analyzed in Fig. 2.*

time and task conditions (Fig. 1a; red and blue arrows; Fig. 1c). For simplicity, $\mathbf{u}_t$ only captures the component of the inputs that is deterministic, i.e. repeatable across trials of the same condition. Any trial-to-trial variability in the inputs, together with moment-to-moment variability generated intrinsically within the recurrent population, are explained by the noise $\boldsymbol{\epsilon}_t$ and the initial-condition $\mathbf{z}_0$.

Critically, the simulations show that very different combinations of these factors can result in very similar trajectories. For example, the three models of decision-making differ in the nature of their inputs and recurrent dynamics, each mimicking a specific behavioral "strategy" for perceptual decision-making[36–38,77–80], from unstable, impulsive decisions (Fig. 1a, saddle point), to optimal accumulation of evidence (Fig1a, line attractor), and leaky, forgetful accumulation (Fig. 1a, point attractor). Yet, for the chosen inputs, which depending on the model are either constant (Fig. 1c, saddle point) or transient (Fig. 1c, line and point attractor), all three models result in similar single-trial trajectories (Fig. 1a, gray curves) and essentially indistinguishable condition-averaged trajectories (Fig. 1a, blue and red curves).

Analogous observations hold for the models of movement generation (Fig. 1b). Two of the models have no inputs, and are driven entirely by recurrent dynamics starting from condition-dependent initial conditions—one model implements rotational dynamics[45,46], implying that any variability in the initial condition on a given trial is reflected throughout the entire trajectory (Fig. 1b rotations; gray curves); the other implements what we refer to as "funnel" dynamics[47,61], whereby neural activity is pushed through a narrow channel in state space, and any variability along directions orthogonal to the channel is suppressed (Fig. 1b, funnel). In the third model, the recurrent dynamics implements a point attractor, and responses are mostly input driven[11] (Fig. 1b, point attractor). The simulated condition-averages can neither distinguish between the models with or without inputs, nor between the different recurrent dynamics associated with models that lack an input.

Residual dynamics as a window onto recurrent dynamics

More insights into the underlying computations can be obtained by considering the dynamics of response residuals, the component of single-trial responses that is not explained by the condition-averaged responses[12,17,20,81]. Residuals are defined as the difference between a given single-trial trajectory and the corresponding condition-averaged trajectory (Extended Data Fig. 1). We interpret residuals as perturbations away from the condition-averaged trajectory, and then describe how these perturbations evolve over time (Extended Data Fig. 1).

In the simulated models, the dynamics of residuals can be derived analytically (Fig. 2a, Extended Data Fig. 1). First, we define the *effective dynamics*, which describes how the population response would evolve from any given location in state-space and time in the absence of noise. The effective dynamics is obtained by summing the contributions of the recurrent dynamics and the input. The *residual dynamics* is then obtained by subtracting, from the effective dynamics, a component corresponding to the instantaneous direction of change along the condition-averaged trajectory (Fig. 2a, see labels over each panel).

The residual dynamics describes how a perturbation away from the condition-averaged neural state would evolve relative to the trajectory over the course of one time-step. In Fig. 2c,d, the blue dot indicates the unperturbed, "reference" neural state, which lies along the average trajectory. The tail of each arrow indicates the residual (the perturbed state), and the arrow-head shows how this residual evolves over one time-step. For the saddle point model (Fig. 2c, saddle point), perturbations along the horizontal direction, away from the trajectory, expand over time (arrows point away from the reference state), whereas perturbations along the vertical direction decay back to the trajectory (arrows point towards the reference state). These dynamics correctly reflect the influence of a saddle point in the vicinity of the examined region of state space (Fig. 1a, box). Likewise, the residual dynamics

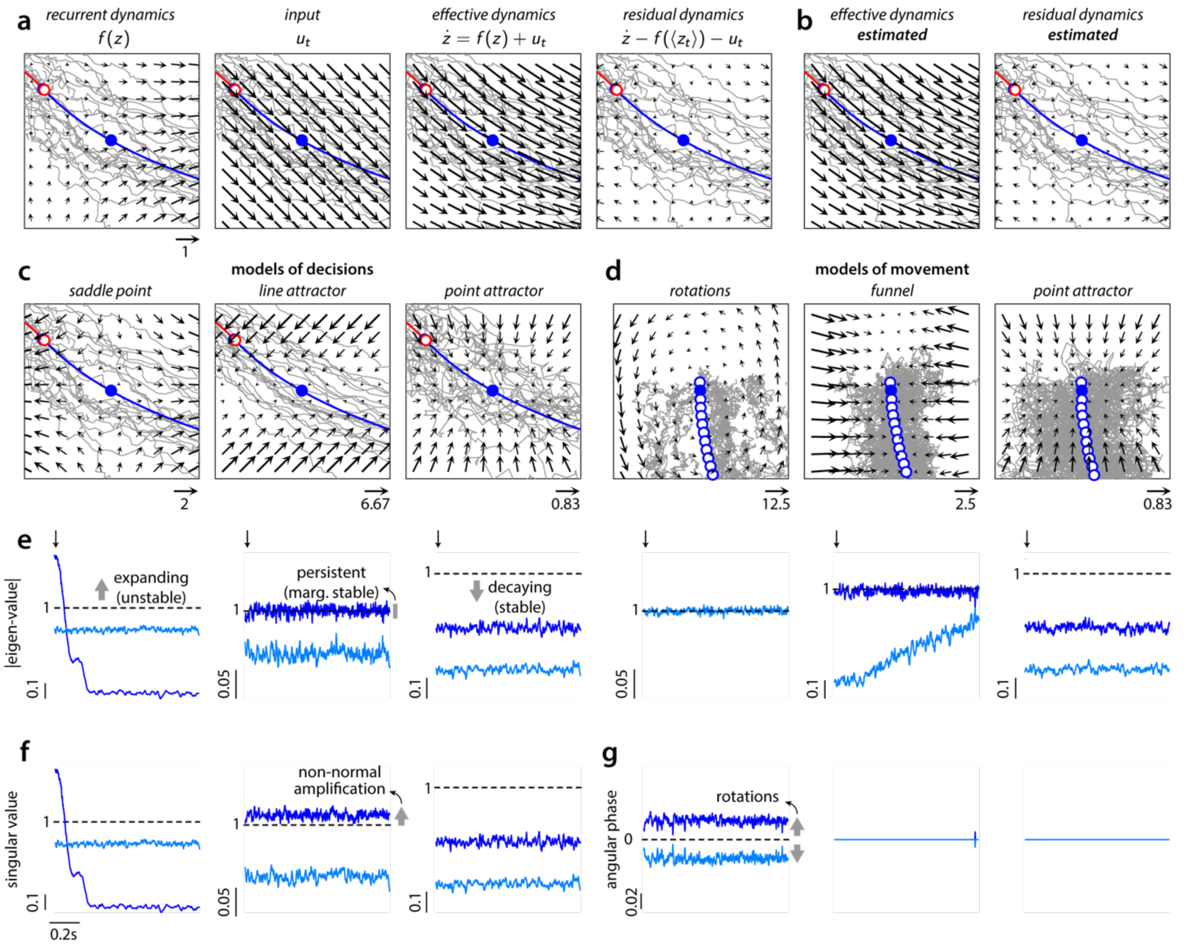


Fig. 2. Residual dynamics reveals population-level computations.

***a**, Different factors contributing to the local dynamics of the saddle point model, shown in the state-space region marked in Fig. 1a for an early time in choice 1 trials (box). Same conventions as in Fig. 1a. Recurrent dynamics and input sum to generate the effective dynamics, determining the evolution of the response in the absence of noise. The residual dynamics is the component of the effective dynamics that explains the evolution of perturbations away from the condition-averaged trajectory (blue line; blue dot: reference time). **b**, Effective and residual dynamics estimated directly from simulated single-trial responses match the ground-truth in **a**. **c**, Ground-truth residual dynamics for the models of decisions, same state-space region and reference time as in **a**. The residual dynamics reflects the key properties of the recurrent dynamics at the corresponding state-space region in Fig. 1a. The arrows in each flow field were scaled by a fixed factor that differed across models and with **a** (numbers close to arrows at the bottom). **d**, Analogous to **c**, but for the models of movement at an early time in choice 1 trials (box in Fig. 1b). **e-g**, Properties of the estimated residual dynamics for the models in Fig. 1. Only residual dynamics for choice 1 is shown. The residual dynamics is described by a time and condition-dependent, autonomous, linear dynamical system. The corresponding dynamics matrices describe the residual dynamics at particular locations along one of the condition-averaged trajectories (Extended Data Fig. 1). **e**, Magnitude of the eigen-values (EV, y-axis) of the 2-dimensional dynamics matrix as a function of time (x-axis). **f**, Singular values (SV) of the dynamics matrix as a function of time for the models of decisions. The difference between EV and SV in the line-attractor model is a consequence of non-normal dynamics. **g**, Angular phase associated with complex-valued EV for models of movement. Larger angular phase implies faster rotational dynamics. EV, SV, and angular phase together distinguish between the different models.*

correctly reveals line attractor and point attractor dynamics in the other two models of decisions (Fig. 2c), as well as the main properties of the recurrent dynamics in the models of movement, i.e. rotational dynamics, decay towards the funnel, and point attractor dynamics (Fig. 2d). More generally, the residual dynamics only reflects the recurrent dynamics, rather than any external inputs, when two constraints are met. First, inputs and recurrent dynamics must combine additively. Second, the noise in the inputs (captured by $\boldsymbol{\epsilon}_t$ in Eq. 1) must be temporally uncorrelated. Both constraints hold exactly for the models in Fig. 1, and at least approximately for many previously proposed RNN models. The second constraint, however, is likely to be violated at the level of many single areas in biological networks, as input variability may be temporally correlated when the input originates in upstream areas that themselves implement recurrent dynamics. Nonetheless, we show below that even in such scenarios residual dynamics can provide insights into the nature of recurrent dynamics in the recorded area. Unlike residual dynamics, the effective dynamics and the condition-averaged trajectories always reflect the properties of both, the recurrent dynamics and the inputs, even when the two constraints above are met.

A further, key property of residual dynamics simplifies the task of estimating it directly from neural responses, even when the underlying computations are non-linear and vary both in time and across state-space location. Residual dynamics is always expressed relative to a "reference" neural state, corresponding to a particular time and location along a condition-averaged trajectory (Fig. 2c,d, blue dot). By this definition, residual dynamics always has a fixed point at the location of the reference state (Fig. 2c,d, blue dot; see methods) making it amenable to be estimated using easily interpretable, statistical models characterized by dynamics that is linear and autonomous (i.e. without inputs). Specifically, the residual dynamics can be approximated by a condition and time-dependent, locally linear system, whereby time parameterizes location in state-space along the condition-averaged trajectory (Extended Data Fig. 1). We estimate these linear systems from neural response residuals by combining methods from subspace identification[33,34] and instrumental variable regression[35] (Extended Data Fig. 2). These methods, unlike simpler linear regression approaches, can produce robust and unbiased estimates of residual dynamics in biologically realistic settings (Extended Data Fig. 3).

We summarize the residual dynamics through the main properties of the estimated local linear dynamical systems, specifically the magnitude of the eigen-values (EV), the singular values (SV), and the rotation frequency associated with the EV (Fig. 2e-g). For locations close to the saddle point in the model of decision-making, one of the EV is larger than 1, implying that perturbations along the associated eigen-vector (the horizontal direction in Fig. 1a, left) *expand* over time; the other EV is smaller than one, corresponding to *decay* along the vertical direction (Fig. 1a, left; center of flow field; Fig. 2e, left-most panel; early times). A line attractor results in a single EV of 1 (Fig. 2e, second from left) as horizontal perturbations are *persistent*, i.e. neither expand nor decay, and a point attractor in all EV smaller than 1 (Fig. 2e, third from left; all directions decay). Rotational dynamics results in EV that are complex-valued and thus associated with a non-zero rotation frequency (Fig. 2g). Finally, differences between the magnitude of SV and EV reflect non-normal dynamics, a critical feature of a number of previous models of neural computation[82–87]. The SV larger than 1 in the line attractor model implies that small perturbations along the corresponding right singular vector transiently expand, even though they are persistent (EV=1) or decay (EV<1) over longer time-scales (Fig. 2e,f).

Neural population responses of decisions and movements in PFC

We compared these model dynamics to neural population responses recorded in the pre-frontal cortex (PFC; area 8Ar) of two macaque monkeys performing a saccade-based perceptual decision-making task[39,81,88,89] (Fig 3a,b; Extended Data Fig. 4). To increase the statistical power of our analyses, we employed a dimensionality reduction technique to “align” the task-related subspaces of neural activity from different experiments with a similar task-configuration (Extended Data Fig. 4; 14-61 experiments per configuration; 150-200 units per experiment). This alignment yielded a shared, 20-dimensional neural state-space explaining >90% of task-related variance in the average neural responses measured across different experiments[90] (Extended Data Fig. 5). All the analyses below are performed within this aligned subspace, although the main results can be reproduced from sufficiently long single experiments (Extended Data Fig. 6).

The condition-averaged population trajectories in PFC shared important features with the average trajectories of the models in Fig. 1. We visualized the population trajectories through projections onto four distinct, two-dimensional activity subspaces: a “choice” plane, emphasizing choice-related activity; a “time” plane, emphasizing time-varying activity common to both choices; and two “jPC” planes[45], emphasizing rotational dynamics (Fig. 3c,d; left to right). We estimated these planes separately during a decision-epoch, which coincided with the presentation of a random-dots stimulus (Fig 3c), and during a movement-epoch aligned to the execution of the saccade (Fig. 3d). As in the decision-models (Fig. 1a), PFC responses started in an undifferentiated state prior to stimulus onset (Fig 3c; choice plane; filled dots mark stimulus onset) and gradually diverged based on the upcoming choice of the animal (Fig. 3c, red vs. blue). PFC responses during the movement period showed pronounced rotational components (Fig. 3d, jPC_{12} plane; filled dots mark movement onset) similar to those in the movement models (Fig. 1b). Prior to saccade-onset, PFC responses fell into largely stationary, choice-dependent states and then transitioned into rotational dynamics following the presentation of the go cue (Fig. 3d, jPC planes).

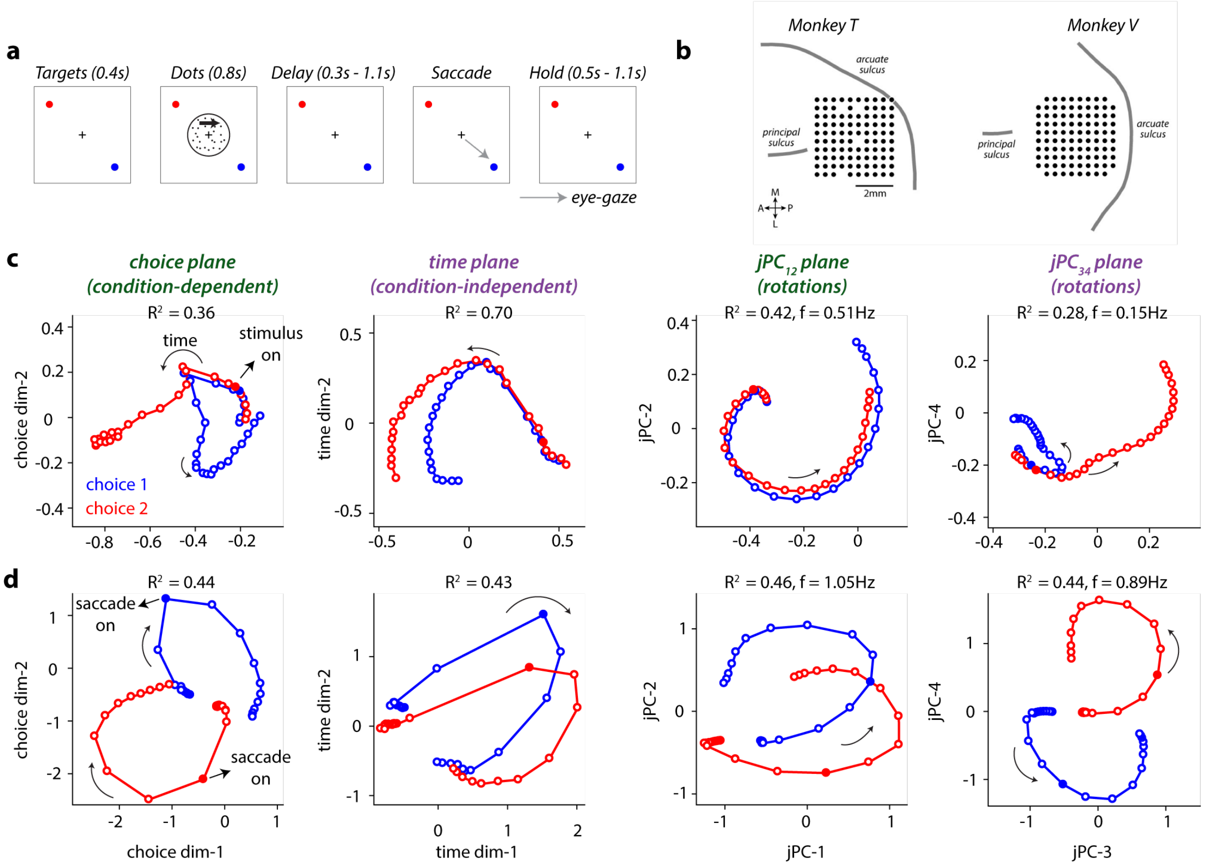


Fig. 3. Average neural dynamics in prefrontal cortex during perceptual decisions and saccades.

***a**, Behavioral task. Monkeys fixating at the center of a screen (fixation point, black cross) viewed a random dot stimulus for 800ms. After a delay period of random duration, they reported the perceived direction of motion with a saccade to one of two targets (red and blue circles; blue: choice 1; red: choice 2). Following the saccade, the monkeys had to fixate on the chosen target during a hold period of random duration. Saccade targets were located at different locations in the visual field for different recording sessions (see Extended Data Fig. 4). **b**, Position of the 10 x 10 electrode array in pre-arcuate cortex of the two monkeys. Black circles indicate the cortical locations of the 96 electrodes used for recordings. **c-d**, Neural trajectories in monkey T, averaged over trials of the same choice. Trajectories are obtained after aligning neural responses (see Extended Data Fig. 5) from experimental sessions with a similar configuration of saccade targets (config-3 in Extended Data Fig. 4). Aligned responses are projected into four activity-subspaces: the choice, time, jPC$_{12}$, and jPC$_{34}$ planes, capturing variance due to choice, time, and rotations, respectively (R^2: fraction of variance explained; f: rotation frequency associated with the jPC plane). **c**, Trajectories in the decision-epoch (-0.2 to 1s relative to stimulus onset, filled circle). **d**, Trajectories in the movement-epoch (-0.7 to 0.5s relative to saccade onset, filled circle).*

The measured PFC responses also differed from the model responses in several ways. Consistent with past reports of population dynamics during decisions, working memory and movements, PFC responses reflected strong condition-independent components during both task-epochs (e.g. Fig. 3c,d, time-plane) [23,25,51,91–95]. Such condition-independent components were not implemented in the models in Fig 1. Unlike in the models, pronounced choice-related activity occurred along more than one state-space direction (Fig. 3c, choice plane) and rotational dynamics within more than one plane. Moreover, rotational dynamics was observed also during the decision-epoch (Fig. 3c, jPC planes). As for the models in Fig. 1, it is not clear which of these features of the condition-averaged trajectories reflect the influence of inputs, recurrent dynamics, or both.

Residual dynamics in PFC

To better resolve the contributions of recurrent dynamics to the recorded responses, we characterized residual dynamics in PFC, by proceeding in two steps. First, we estimated a "dynamics subspace", contained within the previously defined aligned subspace (Fig. 4a, Extended Data Figs. 2,5,7). The dynamics subspace was defined such that within it, but not outside of it, residuals at any given time are significantly correlated with residuals at previous or future times. Second, we exploited these correlations to estimate residual dynamics within the dynamics subspace, following the same approach as for responses simulated from the models above (Fig. 2e-g, Extended Data Fig. 2,8).

We found that residual dynamics in PFC was stable and decaying across the decision and movement epochs (Fig. 4b), as the largest estimated EV magnitudes were consistently smaller than 1 in both monkeys (Fig. 4e; $p < 0.001$, single tailed t-test, n= 144 data points across times, choices and configurations). The dynamics subspace was close to 8-dimensional in all configurations (Fig 4a, Extended Data Fig. 7,8) and was best aligned with directions that explained most task-related variance within the aligned subspace (Fig. 4a, largest dot products at small values along y-axis; Extended Data Fig. 5). Any directions lying outside the dynamics subspace can be thought of as being associated with an EV equal zero, meaning that perturbations along these directions completely decay within a single time step.

The EV magnitudes were strongly time-dependent. For all task configurations, the largest EV were attained during the decision epoch or the delay period preceding the saccade. These EV were associated with decay time-constants in the range 187-745ms during the decision period (0s to +0.8s following stimulus onset) and 110-913ms during the delay period (-0.5s to +0.3s relative to saccade onset) for monkey T (95% CI, medians = 352ms and 293ms; Fig. 4e, top), and 309-1064ms and 192-3586ms for monkey V (95% CI, medians = 489ms and 491ms; Fig. 4e, bottom). Concurrently with the saccade onset, the EV consistently underwent a strong contraction—the largest measured time constants at saccade onset fell to median values of 159ms in monkey T and 310ms in monkey V (Fig. 4e), implying that perturbations away from the average trajectory during movement quickly fall back to the trajectory.

These findings alone rule out several models of recurrent dynamics in PFC. Even the largest EV in PFC during the decision epoch are inconsistent with unstable dynamics (EV>1, Figs. 1a,2e; saddle point) and for the most part substantially smaller than what would be expected

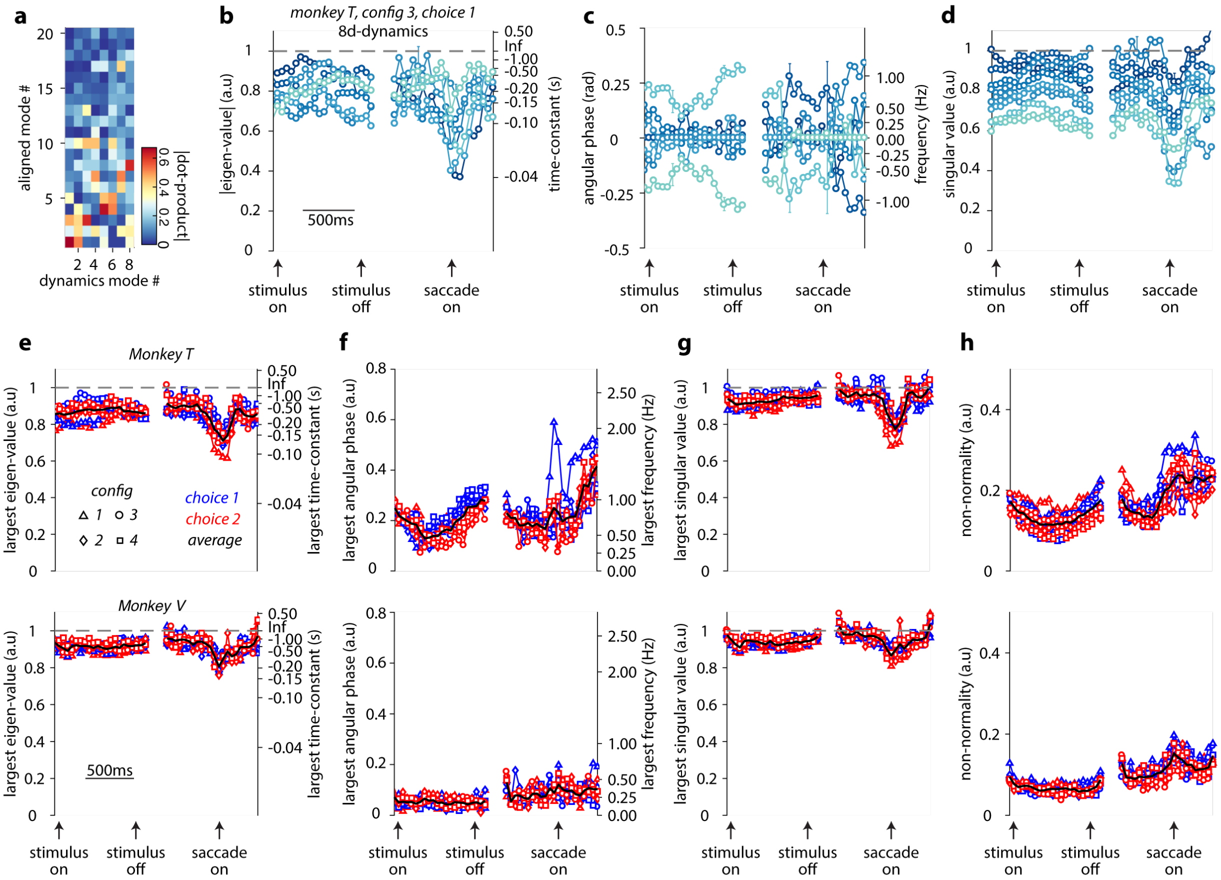


Fig. 4. Residual dynamics in prefrontal cortex during perceptual decisions and saccades.

***a-d**, Estimated residual dynamics in prefrontal cortex in monkey T, same task configuration as in Fig 3c-d. The residual dynamics was 8-dimensional for this example dataset. **a**, Relative alignment between the modes spanning the 8d-dynamics subspace and the modes spanning the 20d-aligned subspace (see Extended Data Fig. 5), measured as the absolute value of the corresponding dot-product. The dynamics modes project strongly onto the first few aligned modes, which capture most of the task-relevant variance in the responses. **b-d**, Properties of the residual dynamics for a single choice condition (choice 1). Error bars: 95% bootstrap confidence intervals (shown at selected times). **b**, Eigen-values (EV) of the dynamics (left axis), and associated time-constants of decay (right axis) as a function of time (x-axis). **c**, Angular phase of the EV (left axis; angular phase = 0: real-valued EV) and associated rotation frequencies (right axis). **d**, Singular values (SV) of the dynamics. The eigenvectors and singular vectors associated with the shown EV and SV can vary over time. Line colors reflect the magnitude of the EV or SV at the first time of the decision epoch. At later times, colors match those associated with the closest eigen-vector or right singular vector at the previous time. **e-h**, Properties of the residual dynamics across all animals (Monkey T, top; Monkey V, bottom), choices (blue: choice 1; red: choice 2), and task configurations (markers; see Extended Data Fig. 4). Black curves: averages across all choices and configurations. **e**, Magnitude of the largest EV (left axis) and the associated decay time-constants (right axis). **f**, Largest angular phase of the EV and the corresponding frequency of rotation. **g**, Largest singular value. **h**, Time course of the index of non-normality.*

from persistent dynamics (EV≈1, Figs. 1a,2e; line attractor). Likewise, the small EV around the time of the saccade are inconsistent with purely rotational or funnel dynamics, which would both result in directions with very slow decay (EV≈1, Figs. 1b,2e; rotations and funnel). Rather, the inferred EV are consistent with quickly decaying recurrent dynamics (Figs. 1b,2e; point attractor).

The absence of strong rotational dynamics is bolstered by the finding that the largest estimated rotation frequencies are either close to zero or very small for most EV in both monkeys (Fig. 4f). We did observe a few EV with rotation frequency considerably larger than zero (≈0.5-1Hz) in monkey T (Fig. 4c). However, around the time of movement the associated EV magnitudes were small (e.g. time constants between 70-110ms, Fig. 4b; dark blue) implying that perturbations decay within 1/15th of a rotational cycle. Overall, these findings are inconsistent with the large rotation frequencies and slow decay expected for purely rotational recurrent dynamics (Fig. 2e,g; rotations).

Finally, the largest SV had a somewhat larger magnitude than the largest EV throughout both task epochs, particularly in monkey T (compare Fig 4e to 4g). This finding indicates that dynamics in PFC is non-normal, albeit only weakly. Even the largest SV are smaller than 1, implying that the non-normal recurrent dynamics does not amplify perturbations, it only transiently slows down their decay. The degree of non-normality, quantified as the discrepancy between the EV and the SV, followed a consistent time-course across animals and configurations, and was most pronounced around the time of the saccade (Fig. 4h).

Condition-averaged trajectories reflect time-dependent input contributions

Additional insights into the relative strengths of recurrent dynamics and inputs can be gained by comparing the properties of residual dynamics and condition-averaged trajectories. When inputs are weak, the trajectories mostly reflect the properties of the recurrent dynamics, which in turn results in distinct relations between trajectories and residual dynamics. For example, in the saddle-point and line-attractor models, the condition-averaged trajectories for the two choices diverge along a direction that is closely aligned with the eigenvector associated with the largest EV in the residual dynamics (Fig. 1a, left-most panels; horizontal direction; Fig. 2e). Similarly, in the funnel and rotation models, the condition-averaged trajectories rotate in the plane containing residual dynamics with EV close to 1 (Fig. 1b, left-most panels; Fig. 2e), or EV with large angular phase (Fig. 1b, left panels; Fig. 2g). When such relations are absent, two scenarios are possible. First, the neural trajectories may mostly be driven by a strong input (Fig. 1b,2e, point-attractor model: trajectories rotate, whereas residual dynamics is decaying and non-rotational). Second, the recurrent dynamics may implement strong non-normal amplification, where population trajectories can display pronounced excursions along directions that are largely orthogonal to the eigenvectors associated with the largest EV[82,85,96,97]. While the latter scenario is ruled out by the properties of the residual dynamics (Fig. 4g,h, SV≤1; no non-normal amplification), the former is not.

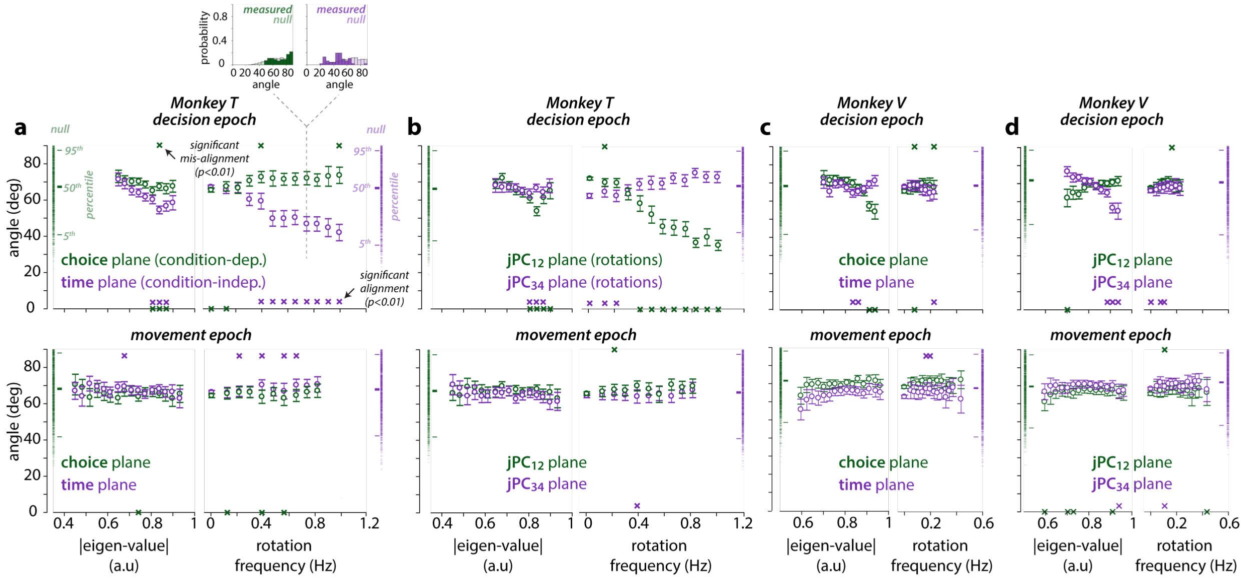


Fig. 5. Relationship between residual dynamics and average dynamics.

*Overlap between the eigenvectors of the residual dynamics and the four activity subspaces defined as in Fig 3c-d (choice, time, jPC$_{12}$, and jPC$_{34}$ planes; see legends), in monkeys T (**a-b**) and V (**c-d**) during the decision and movement epochs (top and bottom). Overlap is defined as the subspace angle between a particular eigenvector (real-valued EV) or pair of eigenvectors (complex-valued EV) and a given plane. Subspace angles for both choices and task-configurations are averaged within bins defined based on EV magnitude or rotation frequency (left and right halves of the x-axis; errorbars: normal 95% confidence intervals). To determine whether eigenvectors within a bin are significantly aligned or misaligned with a given plane (crosses; close to subspace angles of 0 and 90) we compared the corresponding subspace angles to null distributions obtained from randomly sampled directions in the dynamics subspace (vertically arranged purple and green ticks, on the left and right of each plot, see Methods). Measured and null distributions for two example bins are shown in **a** (inset, top). **a-b**, subspace angles with the choice and time planes (**a**) and the jPC$_{12}$ and jPC$_{34}$ planes (**b**), in monkey T. **c-d**, same as **a-b**, for monkey V.*

We quantified the relationship between residual dynamics and the condition-averaged trajectories in PFC (Fig. 5) as the subspace angle between the eigenvectors of the residual dynamics and the four activity subspaces that we defined based on condition-averaged trajectories (Fig. 3). To reveal relationships of the kind predicted by some of the decision and movement models, we sorted subspace angles, based either on the magnitude (Fig. 5a,b left half of x-axis) or rotation frequency of the associated EV (Fig. 5a,b, right half of x-axis). For each magnitude and angular phase, we also determined whether the measured subspace angles were significantly smaller (i.e. “aligned”) or larger (“mis-aligned”) than expected based on randomly chosen directions within the dynamics subspace (Fig. 5, crosses).

During the saccade epoch, subspace angles with the jPC planes showed no dependency on either the rotation frequency or the magnitude of the associated EV in both monkeys (Fig. 5a-d, bottom). In fact, the mean subspace angle obtained for any given EV magnitude or rotation frequency closely matched that expected from the null distribution (Fig. 5a-d, bottom; vertically aligned green and purple points on the left and right). The prominent rotations in the condition-averages during the saccade epoch (Fig. 3d) are thus not preferentially aligned with eigenvectors associated with EV of large magnitude or large rotation frequency. This finding is inconsistent with the predictions of the rotation and funnel models, and instead suggests a prominent role of inputs in driving saccade-related activity (Fig. 1b, point attractor).

On the other hand, the subspace angles were related to the properties of the residual dynamics during the decision-epoch, although the observed relations differed across animals. In monkey V, the choice plane is best aligned with the eigenvectors of the largest magnitude EV (Fig. 5c top, left half of x-axis; green points), consistent with a role of the recurrent dynamics in generating choice responses. This relation is less pronounced in monkey T (Fig. 5a top, left half of x-axis; green points), for which the residual dynamics was better aligned with the time-plane, capturing choice-independent variance, than with the choice plane (Fig. 5a top, left half of x-axis; purple points). Nonetheless, a pronounced relation between residual dynamics and condition-averaged responses was apparent in monkey T, although of an unexpected kind. The subspace angles with the time-plane (Fig. 5a top, right half of x-axis; purple points) and the jPC_{12} plane (Fig. 5b top, right half of x-axis; green points) showed a strong dependence on EV rotation frequency, suggesting that the rotational structure of the trajectories in those planes during the decision epoch reflects the influence of rotational recurrent dynamics.

In both monkeys, the properties of residual dynamics (Fig. 4) and its relation to condition-averaged trajectories (Fig. 5) thus suggest that recurrent dynamics substantially contributes to shaping the condition-averaged trajectories measured in PFC only during the decision-epoch (Fig. 3c). The large excursions in the trajectories observed during the movement epoch (Fig. 3d) instead seem more consistent with the influence of strong external inputs[11].

Interpreting local residual dynamics in distributed cortical circuits

These above conclusions, however, are based on a comparison to simplified models of neural dynamics, for which inputs and recurrent contributions are well defined (Fig. 1). Biological circuits tend to be modular, i.e. are subdivided into areas, with both local recurrence within areas, as well as long-range, feedforward or feedback connections between areas[98,99]. At the level of any single area, a clear distinction between inputs and recurrent dynamics may then be challenging, raising the question of how residual dynamics should be interpreted when computations are distributed across many areas.

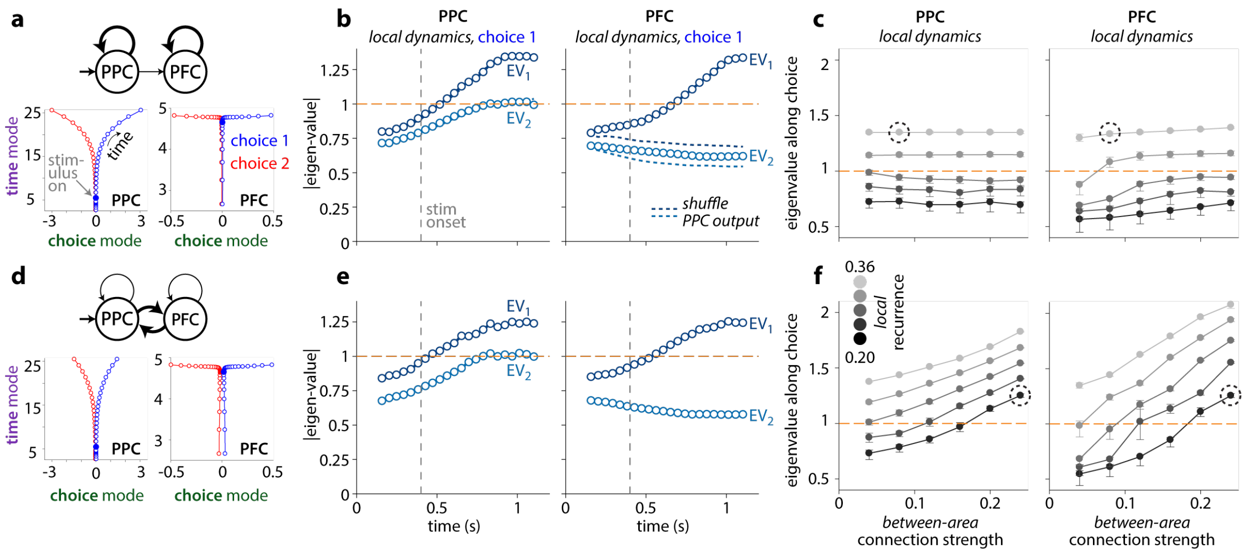


Fig. 6. Local residual dynamics in multi-area networks of perceptual decision making.

*Each network consists of two interconnected modules (PPC and PFC), whereby a module mimics an RNN with a given level of local recurrence. PPC is driven by an external input, and feedback connections from PFC to PPC are either absent (**a-c**) or present (**d-f**). **a**, Connectivity (top) and average trajectories (bottom) for an example network with weak feedforward connectivity between areas (top, thin arrow) and strong local recurrence (thin arrows). Condition-averaged trajectories are shown separately for each area for two choices (blue: choice 1, red: choice 2). Trajectories are visualized in a subspace spanned by the choice mode, explaining variance due to choice, and a time mode, explaining condition-independent variance. **b**, Time-varying EV magnitude of the local residual dynamics estimated from residuals in PPC (left) or PFC (right) for choice 1, in the example network in **a**. The external input is turned on 400ms after the start of the trial (gray dashed line). EV magnitudes in PFC are strongly reduced upon shuffling the feedforward output of PPC across trials (blue dashed curves). **c**, Maximum EV magnitude (measured across time) for residuals projected onto the choice modes in PPC (left) or PFC (right), as a function of the strengths of local recurrence (black to gray: small to large recurrence) and between-area connections (x-axis). Errorbars indicate 95 percentile bootstrap confidence intervals. The dashed circle marks the example network shown in **a-b**. **d-e**, Same conventions as in **a-c**, but for networks with between-area feedback.*

To address this question, we consider simulations of a two-area, non-linear, recurrent neural network previously proposed to explain the interplay of posterior parietal cortex (PPC) and PFC during decision-making and working-memory[48]. The network implements both local recurrence within each area (PPC and PFC), as well as long-range connectivity between the two areas. PPC is assumed to be upstream of PFC, as it alone receives an input encoding external stimuli. Here we consider only a limited set among all possible network configurations. First, the strength of local recurrence is set to be equal in both areas. Second, when feedback connections from PFC to PPC are present, their strength equals those of the feedforward connections from PPC to PFC.

Simulated responses of a random-dots task show choice-dependent and condition-independent components, both in PPC and PFC (Fig. 6a,d; choice and time modes). The EV of the residual dynamics, estimated *locally* in PPC or PFC, are typically time-dependent (Fig. 6b,e, Extended Data Fig. 9). In particular, the dynamics can change from stable (EV<1) to unstable (EV>1) after the input is turned on, reflecting the non-linear nature of these networks.

To assess the interaction of local recurrence and long-range connections, we focus on residuals dynamics estimated along the choice mode in each area (Fig. 6c,f, Extended Data Fig. 9). By design, the choice modes define the "communication subspace" between PPC and PFC in these networks[20,48]—the feedforward and feedback connections between areas are constructed such that activity along the choice mode in one area drives activity along the choice mode in the other area (Extended Data Fig. 9). We summarize the residual dynamics in each network with the peak magnitude of the EV along the choice mode achieved within a trial (Fig. 6c,f, Extended Data Fig. 9).

In networks lacking feedback between areas, the residual dynamics in PPC naturally only reflects the local recurrence, whereby the largest EV gradually increases with stronger local recurrence. (Fig. 6c, PPC). The residual dynamics in PFC closely resembles that in PPC (Fig. 6c, PFC), but this resemblance conceals a critical difference between the two areas. In PPC, the residual dynamics reflects the properties of the local recurrent dynamics. The same is not true in PFC, where any EV>1 mostly reflects recurrent dynamics implemented upstream, in PPC. Indeed, if the output of PPC is "shuffled" to remove any temporal correlations, while retaining its time-varying mean, the EV estimated in PFC fall below 1, indicating that recurrent dynamics in PFC is actually decaying in these networks (Fig. 6b, dashed). We refer to this effect as an "inflation" of the EV in PFC, due to the correlated input from PPC.

Such an inflation of local residual dynamics can occur whenever trial-by-trial variability in the inputs into an area displays correlations across time, as can be the case when the upstream areas themselves implement recurrent dynamics (Extended Data Figs. 10,11). This effect implies that the EV magnitudes we estimated in PFC (Fig. 4) set an upper limit to the "true" values one would observe based on local PFC recurrence alone. Notably, not just the magnitude of the estimated EV can be inflated, but also their rotation frequency (Extended Data Fig. 10b,d). Estimated EV with large rotation frequency could thus reflect rotational dynamics occurring locally, or that are implemented in areas upstream to the recorded area (Extended Data Fig. 10d,e).

In networks with long-range feedback, the residual dynamics in PPC and in PFC reflects both the strength of local recurrence and of long-range connections, whereby reduced local recurrence can be entirely compensated by increased global feedback (Fig. 6f). Unlike in the feedforward networks, where the choice results entirely from dynamics unfolding locally in PPC, here the choice dynamics reflects a process distributed across both areas.

Overall, these simulations show that local residual dynamics in an area cannot be assumed to only reflect local recurrence in that area, as very different combinations of local and long-range connectivity can result in virtually indistinguishable residual dynamics at the level of single areas (Fig. 6a,d vs. b,e). At the same time, these analyses also demonstrate that local residual dynamics can resolve recurrent computations implemented outside of the recorded area, as long as they are unfolding within the output subspace of an upstream area.

Global residual dynamics resolves local and global recurrent computations

The simulations in Fig. 6 imply that the properties of recurrent dynamics in PFC can be constrained, but are not unambiguously revealed, by *local* estimates of residual dynamics (see Discussion). However, we find that detailed insights into the interaction of local and long-range recurrence are possible when considering the *global* residual dynamics, which is estimated from recordings across all areas in a network.

We estimated global residual dynamics from the concurrent, pooled responses simulated in PPC and PFC, for the two example networks with long-range feedforward and feedback connections (Fig. 7). EV magnitudes are qualitatively similar in the two networks, with one EV unstable (EV>1), one persistent (EV≈1), and the others decaying (EV<1; Fig. 7a). The number of global EV does not robustly distinguish between networks, as it reflects a somewhat arbitrary cutoff in the dimensions to include in the dynamics subspace (excluded dimensions effectively have EV=0).

Critically, the alignment between the eigenvectors of the global EV and the local task-activity subspaces can distinguish between the two networks. As above (Fig. 5), we quantified the alignment as the angle between the estimated global eigenvectors and the local choice and time modes in PPC and PFC (Fig. 7b; gray: feedforward, black: feedback). Eigenvectors can be either "shared" across areas, or "private" to an area, depending on whether they have substantial projections (i.e. angle<90) onto modes in both areas or only a single area. For example, EV_1 is shared in both networks, albeit to different degrees, whereas EV_2 is consistently private to PPC. While several global eigenvectors are aligned differently with the PPC and PFC modes in the two networks (Fig. 7b, EV_1 and EV_3), such differences are not evident at the level of local residual dynamics (Fig. 7c).

Global residual dynamics can distinguish between the two networks because variability evolves differently within and across areas depending on the connectivity between areas. To explore the possible nature of these differences, we first consider the effect of perturbations in two simple models implementing time-independent, linear dynamics (Fig. 7d), which mimics key properties of the inferred global dynamics (Fig. 7b). We considered activity that is only two-dimensional, whereby the two cardinal dimensions represent the choice modes

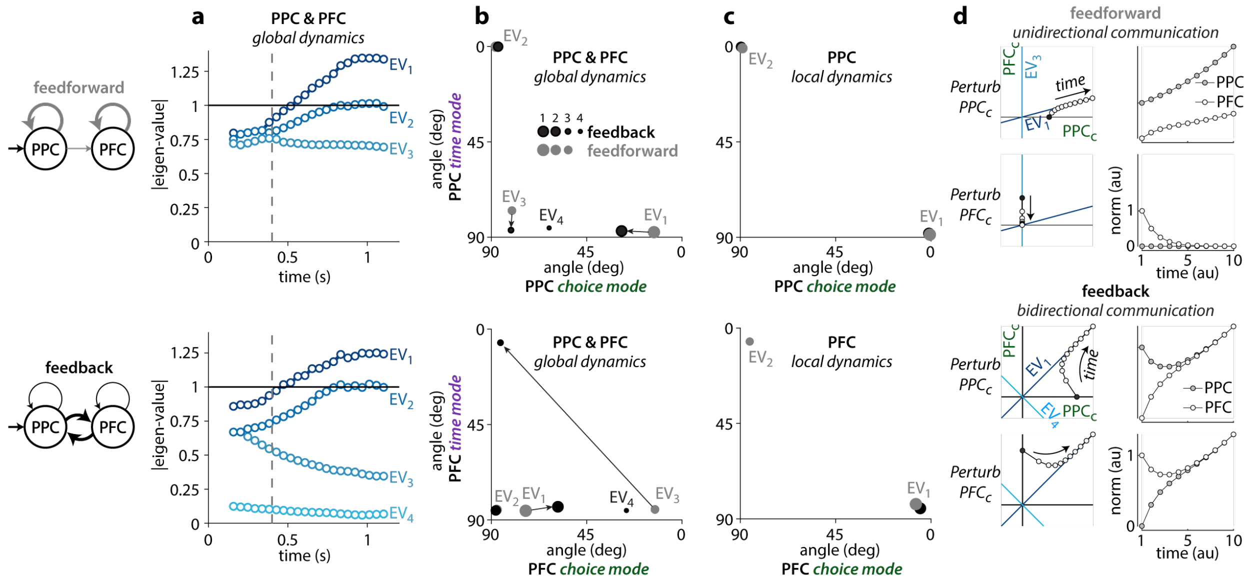


Fig. 7. Global residual dynamics resolves local and between-area recurrent contributions.

***a**, Time-varying EV magnitudes of the global residual dynamics for the example networks in Fig. 6a (top) and Fig. 6d (bottom). Global residuals are obtained by pooling observations from both areas for a single choice condition (here choice 1). **b**, Alignment between the eigenvectors of the global residual dynamics and the choice mode (horizontal axis) or the time mode (vertical axis). Alignment is defined as the angle between the corresponding directions and is shown separately for the example feedforward (gray) and feedback (black) networks. Angles of 0 and 90 deg indicate perfect alignment and complete lack of overlap, respectively. As in Fig. 6a,d, the choice and time modes are defined locally in PPC (top) or PFC (bottom). Angles are shown for a single time late in the trial. Arrows mark large differences in alignment between the two example models. Marker size is proportional to EV magnitude. Bootstrap confidence intervals (95th percentile) for the angles are smaller than the marker sizes. **c**, Analogous to **b**, but for the eigenvectors of the local residual dynamics (see Fig 6b,e) estimated separately based on PPC responses (top) or PFC responses (bottom). **d**, Effect of local perturbations in two simple models implementing linear dynamics that mimic key features of the estimated global dynamics in **a**,**b**. Two-dimensional dynamics evolve in a subspace spanned by the PPC and PFC choice modes (PPC_c and PFC_c). An unstable, global eigenvector is shared across areas; the corresponding eigenvector projects equally onto PPC_c and PFC_c in the feedback model (bottom), but predominantly onto PPC_c in the feedforward model (top; analogous to EV_1 in **b**). A quickly decaying eigenvector is shared in the feedback model (EV_4, bottom) but is largely private to PFC in the feedforward model (EV_3, top; analogous to the corresponding EV in **b**). Activity is perturbed along the PPC_c and PFC_c axes (black circles, left; see labels) and then evolves based on the dynamics determined by the respective EV (white circles, left). The right column shows the norm of activity within each area (i.e. projected onto PPC_c or PFC_c) for the different perturbation types (perturb PPC_c or PFC_c) and models.*

in PPC and PFC, respectively (Fig. 7d). The two models differ in the arrangement of the two eigenvectors of the dynamics, but not in the magnitudes of the associated EV. In the “feedforward” model, an unstable eigenvector projects mostly onto the PPC choice mode, while a stable eigenvector is aligned with the PFC choice mode (Fig. 7d, top; EV_1 and EV_3; similar to the corresponding gray points in Fig. 7b). In the feedback model, both the unstable and stable eigenvectors have large projections onto the PPC and PFC choice modes (Fig. 7d, bottom; EV_1 and EV_4; similar to the corresponding black points in Fig. 7b).

We mimicked a local perturbation either in PPC or PFC by initializing activity along the corresponding choice mode (Fig. 7d, left; black points), and then letting activity evolve based on the linear dynamics determined by the respective EV (Fig 7d, left; white points).

These simple models exemplify how the arrangement of global eigenvectors determines the directionality of the communication between areas. In the feedforward model, a PPC perturbation causes expanding activity in PPC that propagates to PFC, whereas a PFC perturbation decays in PFC, and does not propagate to PPC (Fig. 7d, top, right column). This unidirectional communication results from non-normal dynamics, as EV_1 is shared, while EV_3 is private to PFC (Fig. 7d, top; EV_1 not orthogonal to EV_3). In the feedback model, perturbations in either PPC and PFC propagate to the other area (Fig. 7d, bottom, right column). Such bidirectional communication results from normal dynamics, and the fact that both EV_1 and EV_4 are shared equally between PPC and PFC.

Notably, the existence of bidirectional communication is also reflected in the activity of the perturbed area. Somewhat counter-intuitively, activity in the area that was perturbed initially decays, and expands only later; activity in the unperturbed area does not show this dip (Fig. 7d, feedback; PPC and PFC activity in right panels). This dip in activity occurs because any local perturbation is only partially aligned with the shared, unstable direction (EV_1). Initially, activity in the perturbed area then mostly reflects the rapidly decaying component of activity along the second, global eigenvector (EV_4).

Inferring global dynamics with local causal perturbations

We directly verified the insights from these simple linear models by simulating the effect of causal perturbations in the example two-area networks (Fig. 8). We applied *local* perturbations, either in PPC or PFC, by “injecting” an activity pattern corresponding either to the choice mode or the time mode in each area. For each trial, we applied a brief perturbation at one of six different times after stimulus onset, and then let the activity evolve under the influence of the recurrent dynamics and the input. We visualize the effect of a given perturbation as the time-varying norm of the population activity in PPC and PFC for a brief time-window following the onset of the perturbation, averaged over many trials (Fig. 8b-c,e-f; a group of three connected points; analogous to Fig. 7d). The effects of a perturbation depend on the time at which it was applied (Fig. 8b-c,e-f, compare time-courses within each panel), reflecting the time-varying dynamics in these networks (Fig. 7a).

For perturbations applied late in the trial, when dynamics is unstable (Fig. 7a, EV>1), perturbations of the choice modes result in activity that largely matches the dynamics of the simple models above (Fig. 7d). In the feedforward network, PPC perturbations lead to expanding activity in PPC and PFC (Fig. 8b,c; top-left, green), whereas PFC perturbations lead

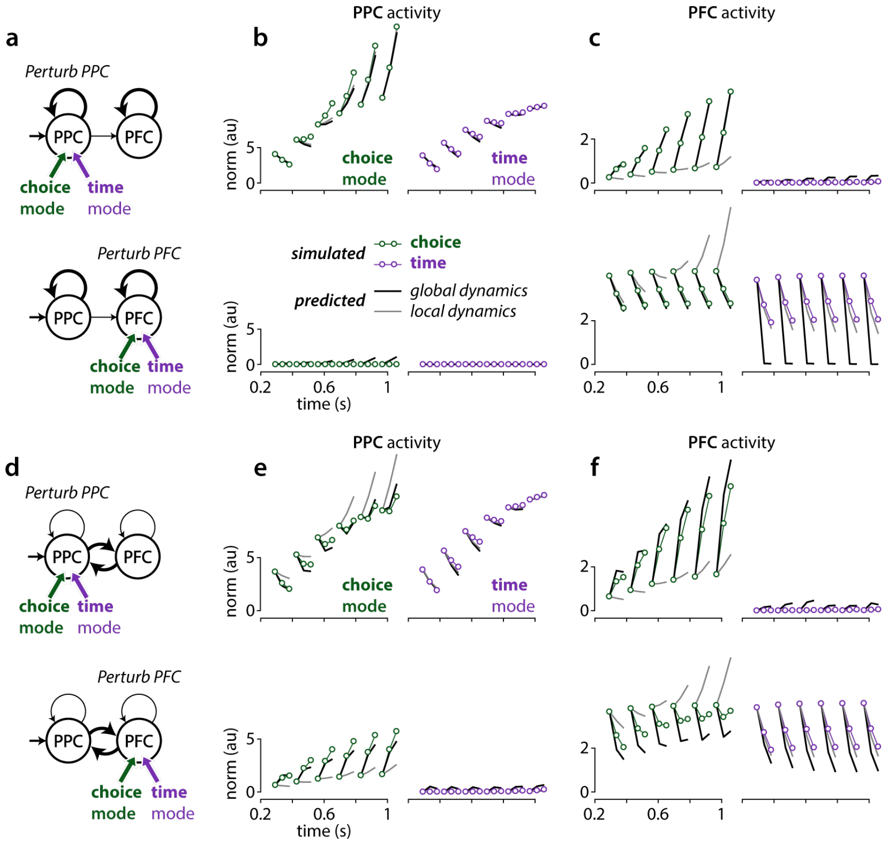


Fig. 8. Residual dynamics explains the effects of targeted causal perturbations.

*Simulated responses to brief perturbations for the two example networks in Fig. 6,7 (small circles) are compared to predictions based on residual dynamics (**a-c** and **d-f**: network without and with feedback between areas). Perturbations are applied locally in each area, along the choice or time mode (green and purple circles) at one of six times in the trial (the first point of each curve in **b-c** and **e-f**). Predictions are based either on the local residual dynamics in the simulated area (gray curves; b,e: PPC; c,f: PFC) or on the global residual dynamics (black curves). **a**, Schematic of the location and type of perturbations shown in **b** and **c** for the network without feedback. **b**, Simulated impulse responses in PPC for perturbations in PPC (top) or PFC (bottom) along the respective choice (left) and time modes (right) compared to the corresponding predictions based on local PPC residual dynamics (gray) or global residual dynamics (black). The norm of the impulse response (y-axis) is shown against time in the trial (x-axis). The last two points on each curve correspond to responses for the two time-steps following the offset of each perturbation. **c**, Analogous to **b**, but for responses in PFC. **d-f**, Analogous to **a-c**, but for the network with feedback. Predictions based on the global, but not the local, residual dynamics capture the qualitative features of the simulated impulse responses, i.e. decay, expansion, or decay followed by expansion (e.g. **c**, top-left; **c**, bottom-left; **e**, top-left).*

to decaying activity in PFC (Fig. 8c, bottom-left) and no activity in PPC (Fig. 8b, bottom-left). In the feedback network, PPC and PFC perturbations lead to a dip in activity in the perturbed area (Fig. 8e, top-left and Fig. 8f, bottom-left) and to expanding activity in the non-perturbed area (Fig. 8f, top-left and Fig. 8e, bottom-left), as in the corresponding simple model (Fig. 7d, feedback). All these effects are specific to perturbation along the choice modes—perturbations along the time-mode, in either area, result in very different, consistently decaying dynamics (Fig. 8b-c,e-f; purple color).

These varied effects of causal perturbations can be predicted quite accurately based entirely on our estimates of the global residual dynamics (Fig. 7a-b). The predicted time-course of activity following a perturbation at least qualitatively matches the simulated one, for all types of perturbations (Fig. 8b-c,e-f, black). Predictions based on local estimates of residual dynamics fare much worse overall, but the failures are nonetheless informative about the underlying network (Fig. 8b-c,e-f, gray). For example, the inflation of local PFC residual dynamics in the feedforward network (Fig. 6b) leads to the erroneous prediction that PFC perturbations result in expanding, rather than decaying, PFC activity (Fig. 8c, bottom-left, gray). In the feedback model, predictions based on local residual dynamics instead fail to account for the dip in activity in the perturbed area (Fig. 8e, top-left, Fig. 8f, bottom-left) and underestimate the increase in activity in the unperturbed area (gray; Fig. 8f, top-left, Fig. 8e, bottom-left). Both failures reflect the existence of a global, shared unstable direction, which local residual dynamics cannot adequately capture.

Discussion

It has long been recognized that trial-by-trial variability in neural activity can provide insights into population-level computations[12–22]. Residual dynamics amounts to a complete, quantitative description of the dynamics of trial-by-trial variability at the level of a neural population. Residual dynamics tightly relates to the recurrent computations implemented in the underlying neural circuits, and is capable of resolving fine differences in dynamics across state-space locations and time. This fine resolution allows one to describe dynamics that are globally non-linear[100,101], through a series of local approximations. Unlike past statistical approaches that directly model single-trial dynamics[68,69,71], residual dynamics completely discounts the component of neural responses that is *repeatable* across trials of a given task condition. As a result, residual dynamics can be estimated with more easily interpretable models than the dynamics of the full, single-trial neural responses.

The properties of global residual dynamics, based on recordings distributed across a network of inter-connected areas, can potentially resolve contributions of local, within-area recurrence and long-range, between-area connections (Fig. 7). The resulting description of dynamics in terms of modes (i.e., eigenvectors) that are shared across areas[98], or private to a single area, appears plausible based on the past identification of communication- and null-subspaces between areas[20,102,103]—an eigenvector that is shared between two areas lies within their communication subspace, whereas one that is private lies outside of it, and potentially within the null-space of either area. Global residual dynamics, however, goes beyond a static description based on such subspaces, as it can capture also the dynamics of the responses (Fig. 8) resulting from unidirectional or bidirectional communication between areas (Fig. 7d, top vs. bottom). Local residual dynamics in a single area, of the kind we describe for PFC, instead cannot readily distinguish between local and global contributions to observed neural responses. Any recurrent dynamics unfolding within the communication subspace of two areas will be reflected in the local dynamics of both areas, irrespective of the directionality of the communication between the areas (feedforward or feedback, Fig. 6c,f).

Nonetheless, even our local estimates of PFC residual dynamics provide constraints on the properties of recurrent dynamics in PFC, and on the nature of the computations underlying decision-making and movement generation. For one, the largest estimated time constants provide an upper bound on the time-constants of the local recurrent dynamics in PFC (Fig. 4e; 324ms and 510ms in monkeys T and V; medians), as any upstream contribution to PFC responses would have inflated these estimates (Fig. 6b; Extended Data Fig. 10,11). Recurrent dynamics in PFC is thus slow[98,104], but stable throughout the decision and movement epochs.

This finding does not rule out that the decision-process leading to the monkeys' choices involves unstable or line-attractor dynamics (Fig. 1a), but those dynamics would have to unfold in areas upstream of PFC[80,105], and at least partly outside their communication subspace with PFC. The estimated time-constants would reflect the dynamics of the decision-process if that process unfolded either in PFC alone, or within its communication subspace with other areas (as for all networks in Fig. 6). In such scenarios, our estimates would imply a leaky decision-process, whereby late evidence affects choice more strongly than early evidence. In practice though, monkeys are thought to terminate the accumulation of evidence early in the trial, when a decision-threshold is reached[106], which would reduce the behavioral effects of any leaks in the accumulation. Notably, a recent study hypothesized that the termination of evidence accumulation coincides with the onset of rotational dynamics in PFC [107]. In our study, condition-independent, rotational dynamics during the decision-epoch also stands out, as in monkey T it is the component of the recorded activity that can be best explained as resulting from recurrent computations (Fig. 5). Irrespective of the possible contributions of PFC to the process underlying the monkeys' choices, this finding may be indicative of a broader role for PFC in governing transitions between cognitive states[107,108], e.g. the transition from an uncommitted to a committed state.

Around the time of the saccade, PFC residual dynamics is quickly decaying, largely non-rotational, and only weakly non-normal, implying that PFC does not implement rotational[45,46], funnel[47,61], or strongly non-normal[82,85] recurrent dynamics of the kind previously proposed to explain movement activity in cortex. Rotational and funnel dynamics are also unlikely to be implemented in an upstream area driving PFC movement responses through a communication subspace, since the signatures of those dynamics would then also appear in PFC residuals (Fig 6, Extended Data Fig. 10). Strong non-normal dynamics in an upstream area, however, could explain the residual dynamics and condition-averages observed in PFC. Non-normal systems can generate large activity transients along directions with only a small projection onto the activity subspace containing the slowest dynamics[97]. If the output from such an upstream area was partially aligned with the activity transients, but orthogonal to the slow dynamics, it could drive strong “input-driven" movement-related activity in PFC without revealing the signatures of the strongly non-normal dynamics that created it. Notably, the apparent absence of rotations in PFC recurrent dynamics during saccades does not rule out that such dynamics occurs in premotor and motor areas involved in hand-reaches[43,45]. The neural mechanisms underlying saccades and reaches may well be distinct, considering the substantial differences in the anatomy of the involved structures[88,89].

A complementary approach to distinguishing between the above interpretations of PFC function, beyond characterizing global residual dynamics, would involve combining local estimates of residual dynamics with targeted causal perturbations[11,23–30]. Residual dynamics naturally leads to predictions of the consequences of such perturbations, and failures of the predictions can be diagnostic of the underlying long-range connectivity (Fig. 8). Most useful in this respect are small perturbations that probe the intrinsic manifold explored by the neural variability[27,31,32].

Residual dynamics and the structure of variability may also speak to specific biological constraints at play in neural circuits. The observation of eigenvalues that are smaller, but close to 1 during the decision-epoch is consistent with the underlying neural circuit operating near a critical regime, resulting in large variability and sensitivity to inputs[109–112]. Variability at the level of single neurons is transiently reduced at the time of stimulus and movement onset (Extended Data Fig. 12), potentially reflecting the widespread quenching of variability across cortex in response to task events[13,113,114]. Near-critical dynamics, non-normality, and variability quenching are thought to emerge naturally in balanced excitation-inhibition (E-I) networks[115–117]. A disruption of E-I balance by the onset of an input could potentially lead to contracting dynamics, and thus reduced variability. Notably, the observed reduction in variability in PFC coincides with contracting dynamics at movement onset, but not at stimulus onset (Extended Data Fig. 12), suggesting that such E-I networks may have to be adapted to fully capture the interactions of internal dynamics, inputs, and variability we observed in PFC.

Author Contributions

A.R.G and V.M conceived and designed the study. A.R.G developed the methods and performed the analyses, with input from M.S. and V.M. A.R.G and V.M wrote the manuscript. All authors were involved in discussing the results and the manuscript.

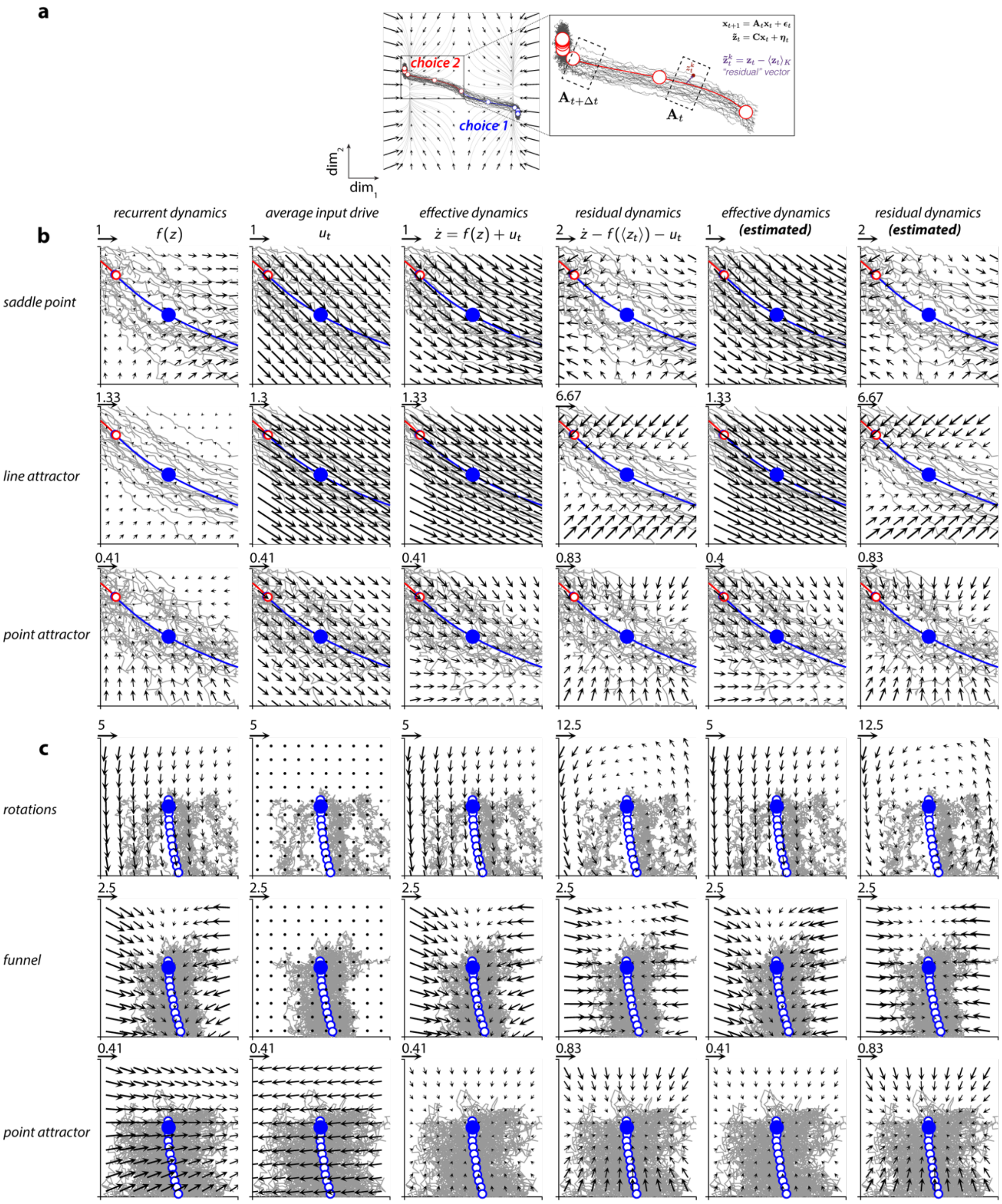
Acknowledgements

We thank John Reppas and William Newsome for the data collection. We thank Kevan Martin and all members of the Mante Lab for their valuable feedback, as well as Nicolas Meirhaeghe, Lea Duncker and Mehrdad Jazayeri for discussions and comments on the manuscript.

Funding

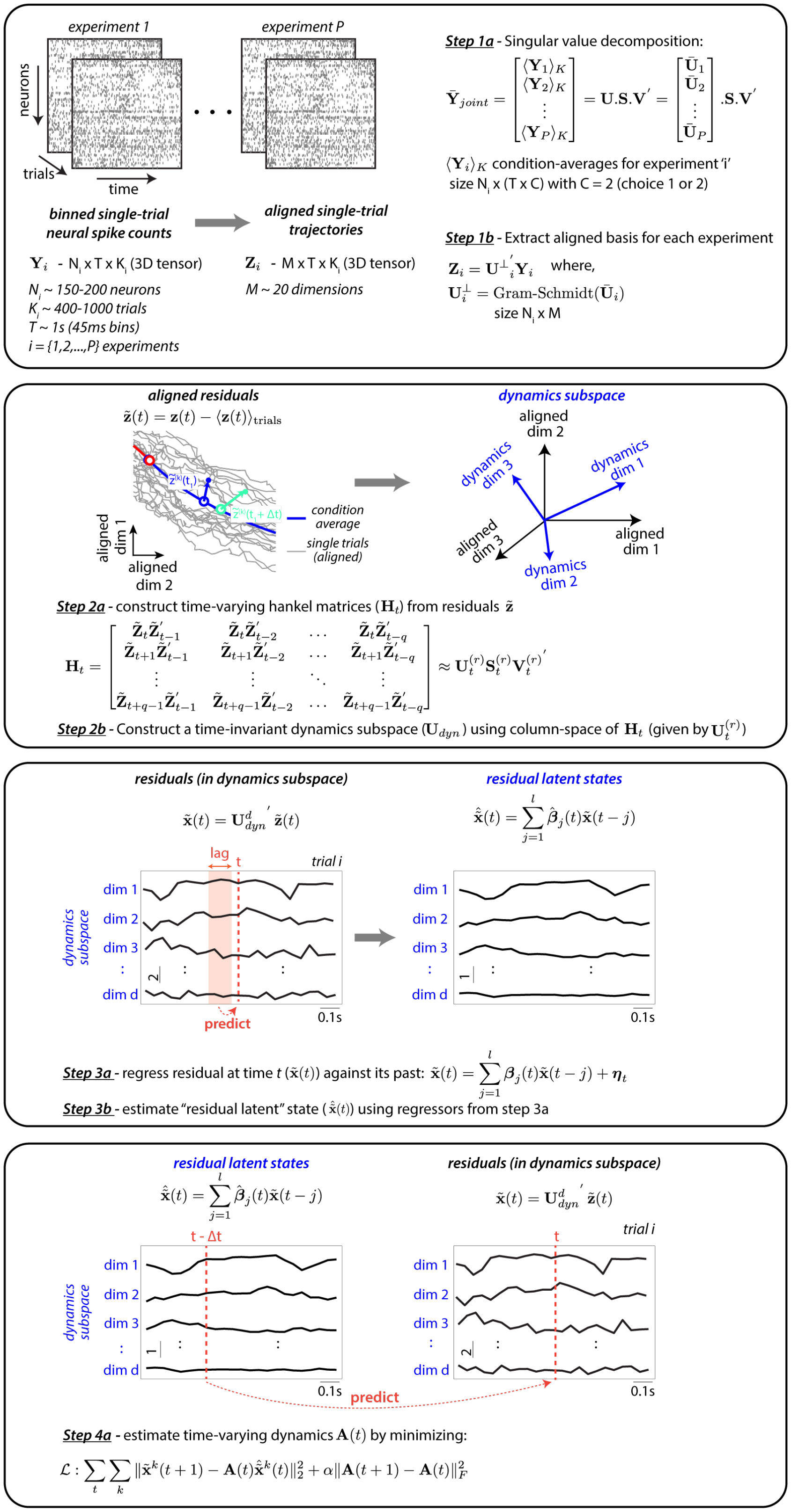
This work was funded by Swiss National Science Foundation (Award PP00P3-157539, VM), the Simons Foundation (SCGB 328189 and 543013, VM; SCGB 543039, MS), the Swiss Primate Competence Center in Research (VM), the Gatsby Charitable Foundation (MS), the Howard Hughes Medical Institute (William Newsome), and the Air Force Research Laboratory (William Newsome).

Extended Data Figures



Extended Data Fig. 1: Residual and effective dynamics in models of decisions and movement

a**, Variability in responses across trials from the same task condition are interpreted as perturbations away from the condition-averaged trajectory. The evolution of these perturbations reflects the properties of the underlying recurrent dynamics (flow field, same conventions as in Fig 1a). Inset on right shows a magnified view of the condition-averaged trajectory (red, choice 2) and corresponding single trials (dark gray) simulated from the saddle point model. Residual vectors at each time (shown in purple for a single trial and time) are computed by subtracting the condition-averaged response at that time from the corresponding single-trial response (purple equation). Time-varying dynamics matrices ($\boldsymbol{A}_t$) of a linear time-varying,* *autonomous state-space model (black equations, top-right) are fit to the residuals. These matrices approximate the dynamics in distinct 'local' regions of state space (e.g. dashed boxes) and are indexed according to time and condition.* ***b-c*, Components of the dynamics for the models of decisions (****b****) and movement (****c****) for an example reference time (blue dot) along the condition-averaged trajectory for choice 1. Same conventions as in Fig 2a. Dynamics are shown for a local state-space region close to the corresponding initial condition (boxes in Fig. 1a, b; left). For all models, the estimated effective and residual dynamics (columns 5 and 6) closely match the true effective and residual dynamics (columns 3 and 4). In these models, the residual dynamics (column 4) reflects only the recurrent dynamics (column 1), but is not identical to it. For one, the fixed point of the residual dynamics by definition is located at the location of the reference state (the blue dot), which in general does not match the position of fixed points of the recurrent dynamics (e.g. the red circle in the first row and first column, corresponding to the position of the unstable fixed point in the saddle point model). The position of fixed points of the recurrent dynamics can only be inferred if the external inputs are known, a requirement that is not fulfilled in many experimental settings. For another, consistent drifts resulting from the recurrent dynamics (e.g. the drift along the limit cycle in the funnel model) are not reflected in the residual dynamics. Such drifts are "subtracted" from the variability in the computation of residuals.*


experiment 1
experiment P
neurons
trials
time
binned single-trial neural spike counts
aligned single-trial trajectories
$\mathbf{Y}_i$ - N_i x T x K_i (3D tensor)
$\mathbf{Z}_i$ - M x T x K_i (3D tensor)
N_i ~ 150-200 neurons
K_i ~ 400-1000 trials
T ~ 1s (45ms bins)
i = {1,2,...,P} experiments
M ~ 20 dimensions
Step 1a - Singular value decomposition:
$\bar{\mathbf{Y}}_{joint} = \begin{bmatrix} \langle \mathbf{Y}_1 \rangle_K \\ \langle \mathbf{Y}_2 \rangle_K \\ \vdots \\ \langle \mathbf{Y}_P \rangle_K \end{bmatrix} = \mathbf{U}.\mathbf{S}.\mathbf{V}' = \begin{bmatrix} \bar{\mathbf{U}}_1 \\ \bar{\mathbf{U}}_2 \\ \vdots \\ \bar{\mathbf{U}}_P \end{bmatrix} .\mathbf{S}.\mathbf{V}'$
$\langle \mathbf{Y}_i \rangle_K$ condition-averages for experiment 'i'
size N_i x (T x C) with C = 2 (choice 1 or 2)
Step 1b - Extract aligned basis for each experiment
$\mathbf{Z}_i = \mathbf{U}^{\perp}{}_i' \mathbf{Y}_i$ where,
$\mathbf{U}_i^{\perp} = \text{Gram-Schmidt}(\bar{\mathbf{U}}_i)$
size N_i x M
aligned residuals
$\tilde{\mathbf{z}}(t) = \mathbf{z}(t) - \langle \mathbf{z}(t) \rangle_{\text{trials}}$
$\tilde{z}^{(k)}(t_i)$
$\tilde{z}^{(k)}(t_i+\Delta t)$
condition average
single trials (aligned)
aligned dim 1
aligned dim 2
dynamics subspace
aligned dim 2
dynamics dim 3
dynamics dim 1
aligned dim 1
aligned dim 3
dynamics dim 2
Step 2a - construct time-varying hankel matrices ($\mathbf{H}_t$) from residuals $\tilde{\mathbf{z}}$
$\mathbf{H}_t = \begin{bmatrix} \tilde{\mathbf{Z}}_t \tilde{\mathbf{Z}}'_{t-1} & \tilde{\mathbf{Z}}_t \tilde{\mathbf{Z}}'_{t-2} & \dots & \tilde{\mathbf{Z}}_t \tilde{\mathbf{Z}}'_{t-q} \\ \tilde{\mathbf{Z}}_{t+1} \tilde{\mathbf{Z}}'_{t-1} & \tilde{\mathbf{Z}}_{t+1} \tilde{\mathbf{Z}}'_{t-2} & \dots & \tilde{\mathbf{Z}}_{t+1} \tilde{\mathbf{Z}}'_{t-q} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{Z}}_{t+q-1} \tilde{\mathbf{Z}}'_{t-1} & \tilde{\mathbf{Z}}_{t+q-1} \tilde{\mathbf{Z}}'_{t-2} & \dots & \tilde{\mathbf{Z}}_{t+q-1} \tilde{\mathbf{Z}}'_{t-q} \end{bmatrix} \approx \mathbf{U}_t^{(r)} \mathbf{S}_t^{(r)} \mathbf{V}_t^{(r)}{}'$
Step 2b - Construct a time-invariant dynamics subspace ($\mathbf{U}_{dyn}$) using column-space of $\mathbf{H}_t$ (given by $\mathbf{U}_t^{(r)}$)
residuals (in dynamics subspace)
$\tilde{\mathbf{x}}(t) = \mathbf{U}^{d}_{dyn}{}' \tilde{\mathbf{z}}(t)$
residual latent states
$\hat{\tilde{\mathbf{x}}}(t) = \sum_{j=1}^{l} \hat{\boldsymbol{\beta}}_j(t) \tilde{\mathbf{x}}(t-j)$
lag
t
trial i
dynamics subspace
dim 1
dim 2
dim 3
:
dim d
predict
0.1s
Step 3a - regress residual at time t ($\tilde{\mathbf{x}}(t)$) against its past: $\tilde{\mathbf{x}}(t) = \sum_{j=1}^{l} \boldsymbol{\beta}_j(t) \tilde{\mathbf{x}}(t-j) + \boldsymbol{\eta}_t$
Step 3b - estimate "residual latent" state ($\hat{\tilde{\mathbf{x}}}(t)$) using regressors from step 3a
residual latent states
$\hat{\tilde{\mathbf{x}}}(t) = \sum_{j=1}^{l} \hat{\boldsymbol{\beta}}_j(t) \tilde{\mathbf{x}}(t-j)$
residuals (in dynamics subspace)
$\tilde{\mathbf{x}}(t) = \mathbf{U}^{d}_{dyn}{}' \tilde{\mathbf{z}}(t)$
t - Δt
t
trial i
dim 1
dim 2
dim 3
:
dim d
0.1s
predict
Step 4a - estimate time-varying dynamics $\mathbf{A}(t)$ by minimizing:
$\mathcal{L} : \sum_t \sum_k \|\tilde{\mathbf{x}}^k(t+1) - \mathbf{A}(t)\hat{\tilde{\mathbf{x}}}^k(t)\|_2^2 + \alpha \|\mathbf{A}(t+1) - \mathbf{A}(t)\|_F^2$

Step 1: Session alignment

Goal: identify common, aligned modes across different experiments.

Approach: singular value decomposition and Gram-Schimdt orthogonalization

Input: concatenated (across sessions) condition averaged trajectories $\bar{\mathbf{Y}}_{joint}$

Output: "aligned" single-trial trajectories

Hyperparamters: # of aligned modes (M)

Step 2: Dynamics subspace estimation

Goal: identify dimensions that are predictive of the "future" based on the "past".

Approach: Hankel matrix decomposition

Input: "aligned" residual responses

Output: ordered set of dimensions that constitute a "dynamics subspace"

Hyperparamters: rank of hankel matrix (r) (ascertained using cross-validation)

Step 3: Residual latent state estimation

Goal: estimate the residual latent state

Approach: First stage of a two-stage least squares(2SLS) regression approach

Input: "aligned" residual responses & dimensions of dynamics subspace (step 2)

Output: denoised estimate of residual latent state at each 't'

Hyperparamters: dimensionality of dynamics subspace (d) & number of regression lags (l)

Step 4: Time-varying dynamics estimation

Goal: estimate time-varying residual dynamics

Approach: Second stage of a two-stage least squares(2SLS) regression (with regularization)

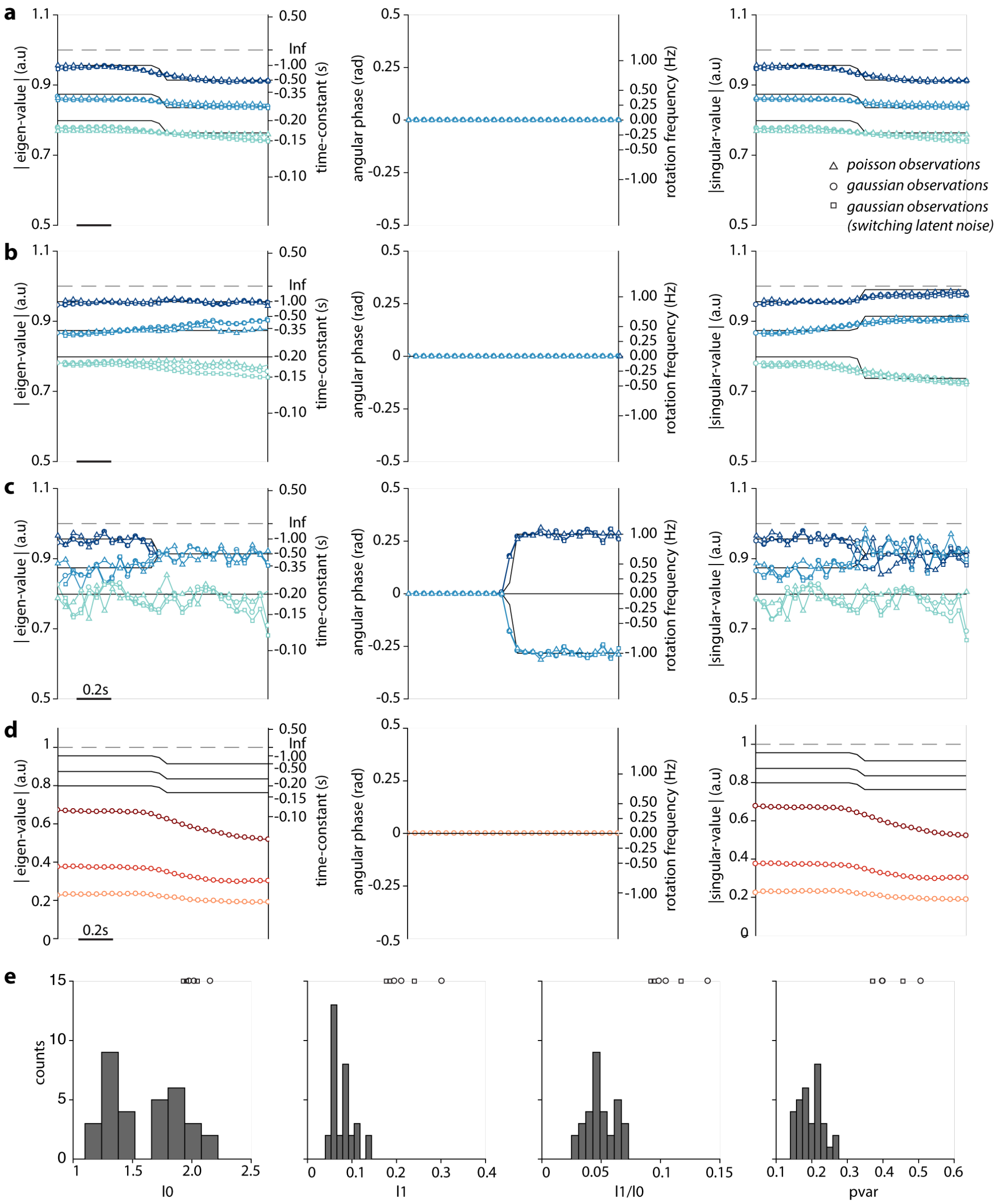
Input: "aligned" residual responses & first stage estimates of residual latent states (from step 3)

Output: time-varying dynamics matrices $\mathbf{A}_t$

Hyperparamters: smoothness penalty α (ascertained using cross-validation)

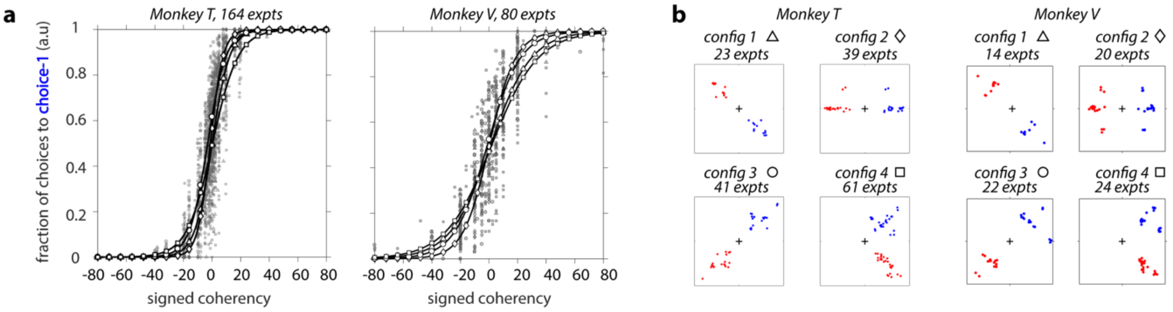
Extended Data Fig. 2: Schematic of analysis pipeline

Schematic depicting the complete data analysis pipeline for inferring residual dynamics from noisy neural population recordings. The pipeline involves four key sequential steps. Step 1: session alignment; involves pooling single trials from different recording sessions in order to increase the statistical power of the analyses Step 2: dynamics subspace estimation; involves using 'aligned' single-trial neural residuals to obtain estimates of a dynamics subspace ($\boldsymbol{U}_{dyn}$*) that effectively contains the residual dynamics; Step 3: residual latent state estimation; involves using the first stage of a two stage least squares (2SLS) approach to estimate a 'denoised' latent residual state; and Step 4: dynamics estimation; uses the denoised residual latent states (obtained in step 3) for the second stage of the 2SLS, in order to estimate the time-varying residual dynamics matrices (*$\mathbf{A}_t$*).*



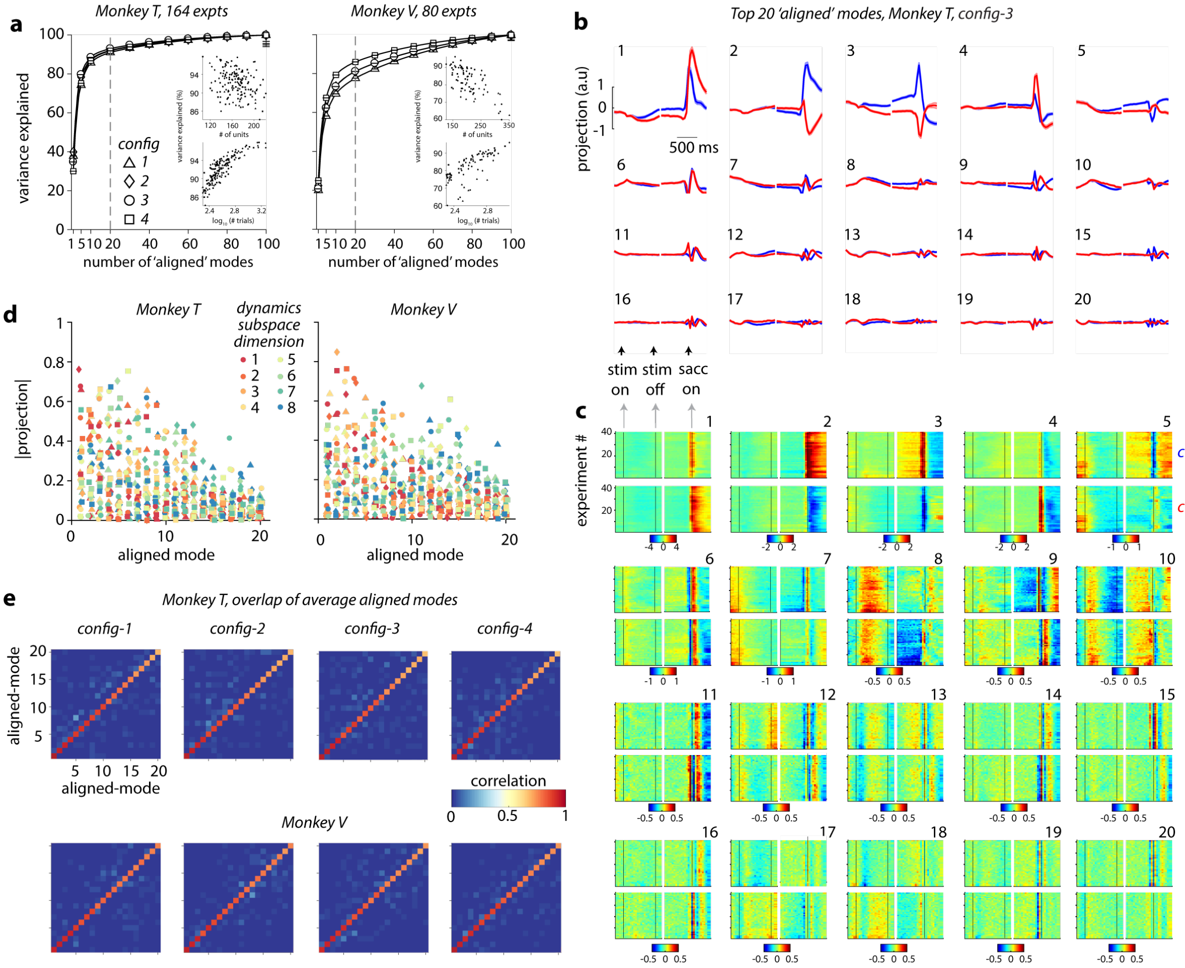
Extended Data Fig. 3: Residual dynamics of simulated, time-varying, linear dynamical systems.

a-c**, Validation of the estimation procedure on simulations of time-varying, linear dynamical systems. Simulations are based on a latent variable dynamical system with 3 latent dimensions and 20 observed dimensions. Residual responses are generated using a gaussian (circle markers: fixed latent noise variance; square markers: latent noise variance switches mid-way through the trial) or poisson (triangle markers) observation process. In all simulations, the properties of the dynamics switch midway through the simulated time window, from slowly decaying to quickly decaying (**a**); from non-rotational to rotational (**b**); or from normal to non-normal (**c**). As in Fig. 4b-d, we characterize dynamics with the magnitude of the eigenvalues (left), the rotational frequency (middle), and the singular values (right). Markers correspond to the estimated residual dynamics, black curves to the ground-truth values. The estimated residual dynamics accurately matches the ground-truth for all types of dynamics and observation models, before and after the switch, and also reveals the time of the switch. We observed this match even when the latent noise variance of gaussian observations was switched at the same time as the eigenvalues/eigenvectors of the dynamics (square markers), demonstrating that estimates of residual dynamics are robust to changes in latent noise* *variance (see also Extended Data Fig. 11a-b vs e-f).* ***d*, Analogous to* ***c****, but for residual dynamics (circles) estimated using ordinary least squares (OLS) instead of two-stage least squares (2SLS) as in* ***c****. Results are only shown for data simulated using a gaussian observation process. Unlike the 2SLS estimates, the OLS estimates are strongly biased, i.e. the magnitude of the eigenvalues and the singular values are consistently underestimated. These biases are expected—they arise because both the regressors and the dependent variables are corrupted by observation noise (see Methods). The 2SLS instead produces unbiased estimates, as the first stage of 2SLS results in a denoising of the regressors (see also Extended Data Fig. 8).* ***e****, Parameters of the latent noise and observation noise for the simulations in* ***a-d*** *were chosen to approximately match the variability in the measured PFC responses. The variability in the measured responses were quantified in terms of four statistics (l0, l1, l1/l0 and pvar, x-axis; see Methods). Histograms indicate the respective values of these statistics in the neural data (one data point per task configuration, choice condition and monkey; see Extended Data Fig. 4). The open markers (top, same conventions as* ***a-c****) indicate the values of the statistics in the simulations for each of the three models.*



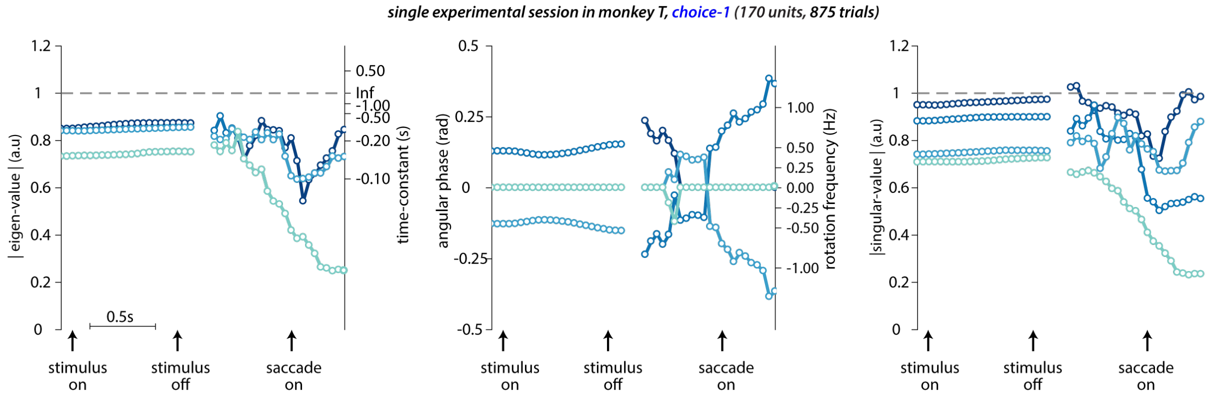
Extended Data Fig. 4: Behavioral task and configurations.

***a**, Psychometric curves for all experimental sessions in both monkeys (left: Monkey T, right: Monkey V), showing the fraction of saccades to choice 1 as a function of the signed motion coherency. Each gray data point is computed from trials belonging to a single experiment (see Methods). The employed values of signed coherency varied slightly across experiments, in an attempt to achieve a comparable overall performance in each experiment. Black curves show logistic functions fitted separately to data points from a given task configuration (different markers; see legends in b) and evaluated at logarithmically spaced levels of coherency (positions of the white markers along the x-axis). **b**, Definition of task configurations. We assigned each experiment to one of four target configurations based on the angular position of the targets (blue: choice 1; red: choice 2). The positions of the targets is similar, but not identical, for experiments assigned to the same task configuration. (left: Monkey T, right: Monkey V). The number of experiments belonging to each configuration are indicated on top of each panel. Estimates of residual dynamics (Fig. 4) are obtained separately for each configuration, after aligning the neural activity from experiments belonging to a given configuration (see Extended Data Fig. 5).*



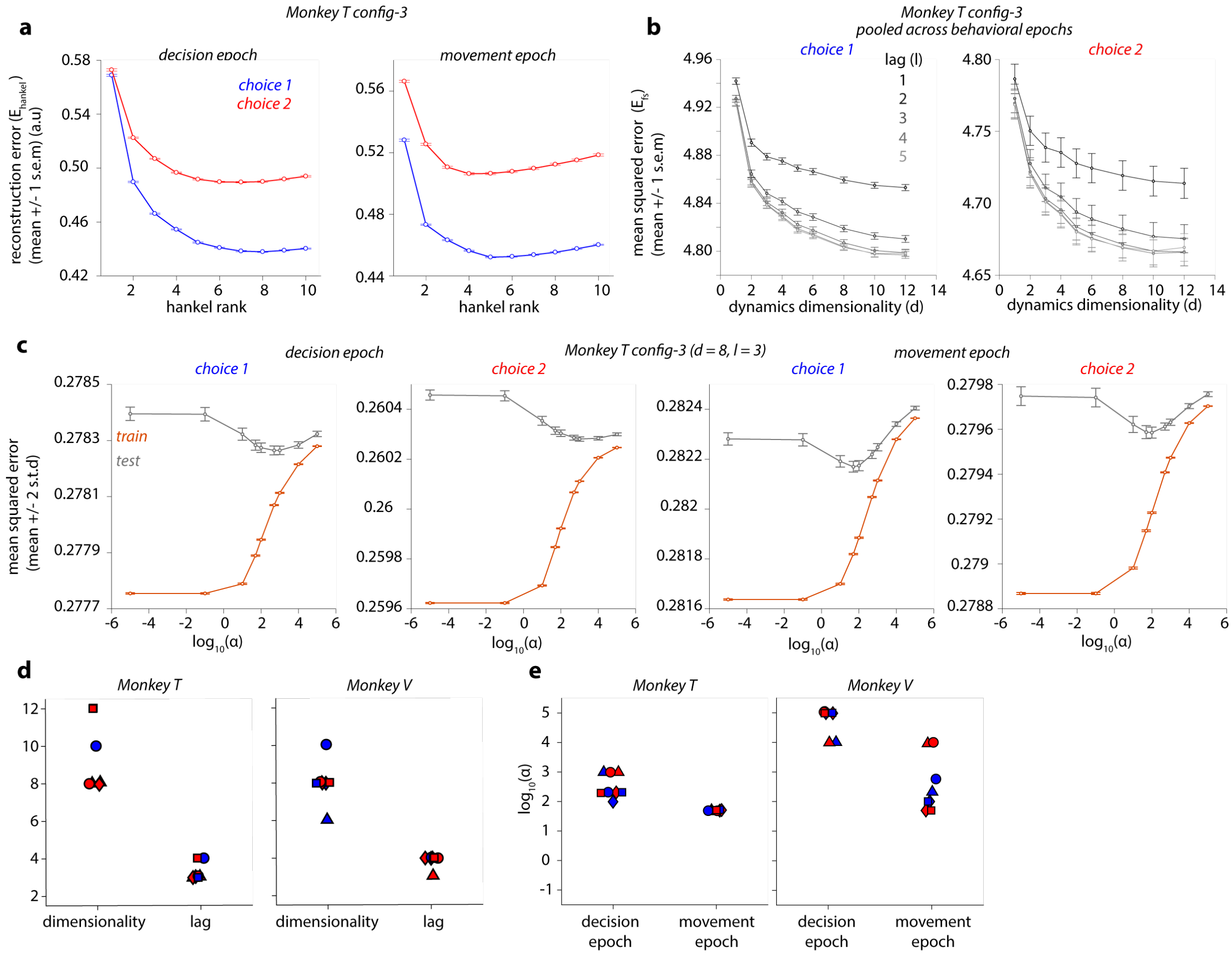
Extended Data Fig. 5: Alignment of neural population responses from different experiments

*Validation of our session alignment procedure, Step 1 of the analysis pipeline (Extended Data Fig. 2). We aligned neural population responses of all experiments belonging to the same task configuration and then pooled the aligned single trial responses across experiments before computing the residuals used in estimating the dynamics. The outcome of the session alignment procedure is a set of 20 'aligned' modes for each experiment, defined such that the activity of each mode has the same dependency on time and choice across experiments. **a**, Cumulative variance explained in condition-averaged population responses as a function of the number of aligned modes in both monkeys (left: Monkey T, right: Monkey V). We show the mean across experiments of the cumulative variance explained for each task configuration (symbols as in Extended Data Fig. 4b). Error-bars indicating twice the standard error of the mean are mostly smaller than the markers. The cumulative variance explained by the first 20 aligned modes for all 164 experiments in Monkey T and 80 experiments in Monkey V showed a strong positive trend with number of trials (inset, bottom) and a weak negative trend with the number of units (inset, top). **b**, Activity of the first 20 aligned modes (numbered from top-left to bottom-right) for config-3 in monkey T (15,524 trials across 41 experiments) ordered according to the amount of variance explained. Activity is defined as the projection of the population condition averages onto each mode. The projection was computed separately across experiments for choice 1 and choice 2 (blue and red) with responses aligned either to stimulus onset or saccade onset (black arrows). The resulting projections were then averaged across experiments (shading: twice the standard error of the mean across experiments). **c**, Same data as in **b** but showing the time-course of each aligned mode (numbered from 1 to 20) for each individual experiment (y-axis) separately for the two choice conditions (choice 1 and choice 2, top and bottom sub-panels). Differences in the activation of a given mode across experiments (i.e. across rows in each sub-panel) are much smaller than the differences in the activations across modes (i.e. across sub-panels), demonstrating the success of the alignment procedure. **d**, Absolute value of the projection (y-axis) of* *the 8 basis vectors (dim-1 through dim-8; red to blue) that span the dynamics subspace ($\boldsymbol{U}_{dyn}$, estimated in Step 2 of the analysis pipeline; Extended Data Fig. 2) onto the 20 aligned modes, indicating the relative alignment of the aligned and dynamics subspace. The dynamics subspace is computed separately for each task configuration (symbols as in Extended Data Fig. 4) in each monkey (left: Monkey T, right: Monkey V) and projects most strongly onto the first few aligned components (i.e large projection values for smaller aligned mode number). The dynamics subspace thus largely overlaps with the subspace of activity that capture most of the task-related variance in the responses.* ***e****, Evaluation of the alignment procedure for all task configurations (columns) in both animals (rows). Each element of the matrix is obtained from the correlation coefficient between the time-courses of two aligned modes (i.e. positions along horizontal and vertical axes). We show the median correlation coefficient across all pairs of dissimilar experiments. Values close to 1 along the diagonal and close to 0 off-diagonal indicate that the time-courses are much more similar across experiments than across modes, indicating successful alignment.*



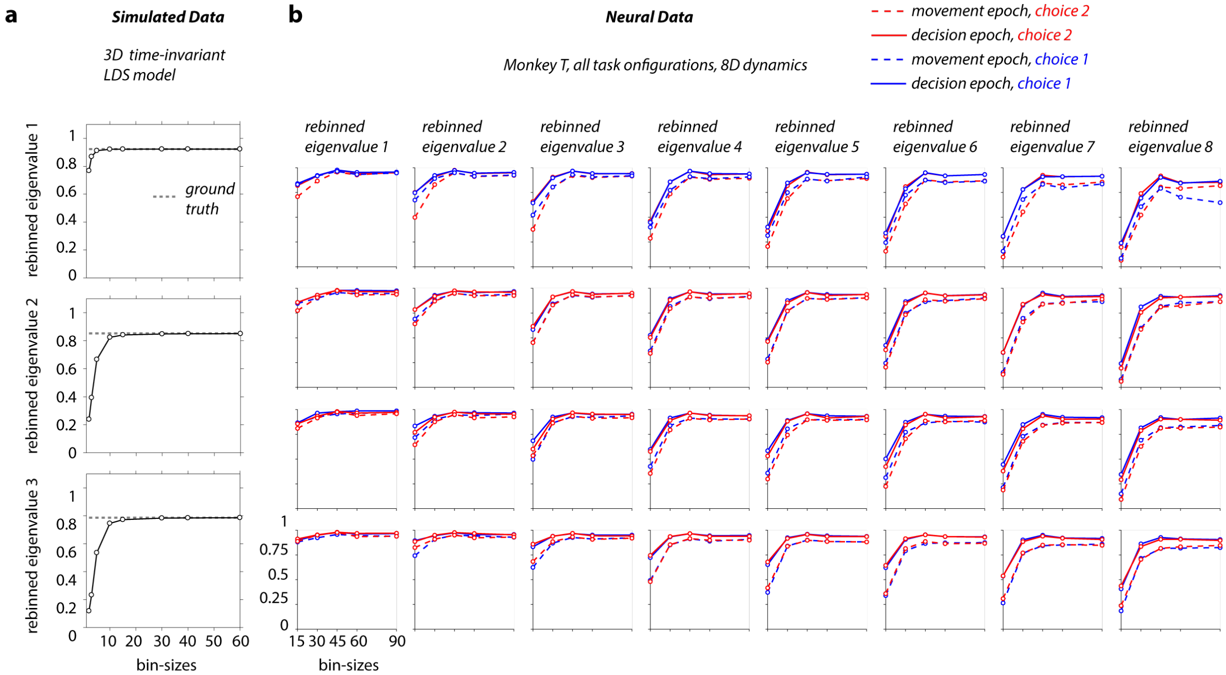
Extended Data Fig. 6: Estimated residual dynamics for neural data from a single unaligned session

Residual dynamics estimated using neural data for a single choice condition (choice-1, 875 trials) from a single experiment in monkey T. This experiment has the largest number of trials among all sessions in monkey T. Conventions as in Fig 4b-d. We estimated the residual dynamics directly from high-dimensional residual observations that corresponded to square-root transformed, binned spike-count vectors (dimensionality = number of units; 170 for this session), without performing the session alignment (step 1 in Extended Data Fig. 2). Overall, the properties of the residual dynamics estimated from this single session are similar to those obtained after pooling trials across sessions (Fig 4b-d, 8 dimensional), suggesting that the main features of the residual dynamics (Fig. 4) are not affected by the alignment procedure. The lower dimensionality of the estimated residual dynamics (4 dimensions, blue to cyan; compared to 8 dimensions in Fig. 4a-d) most likely is a consequence of the smaller number of available trials in the single session compared to the aligned sessions. The resulting smaller statistical power makes is harder to estimate, in particular, the faster decaying eigenmodes of the dynamics.



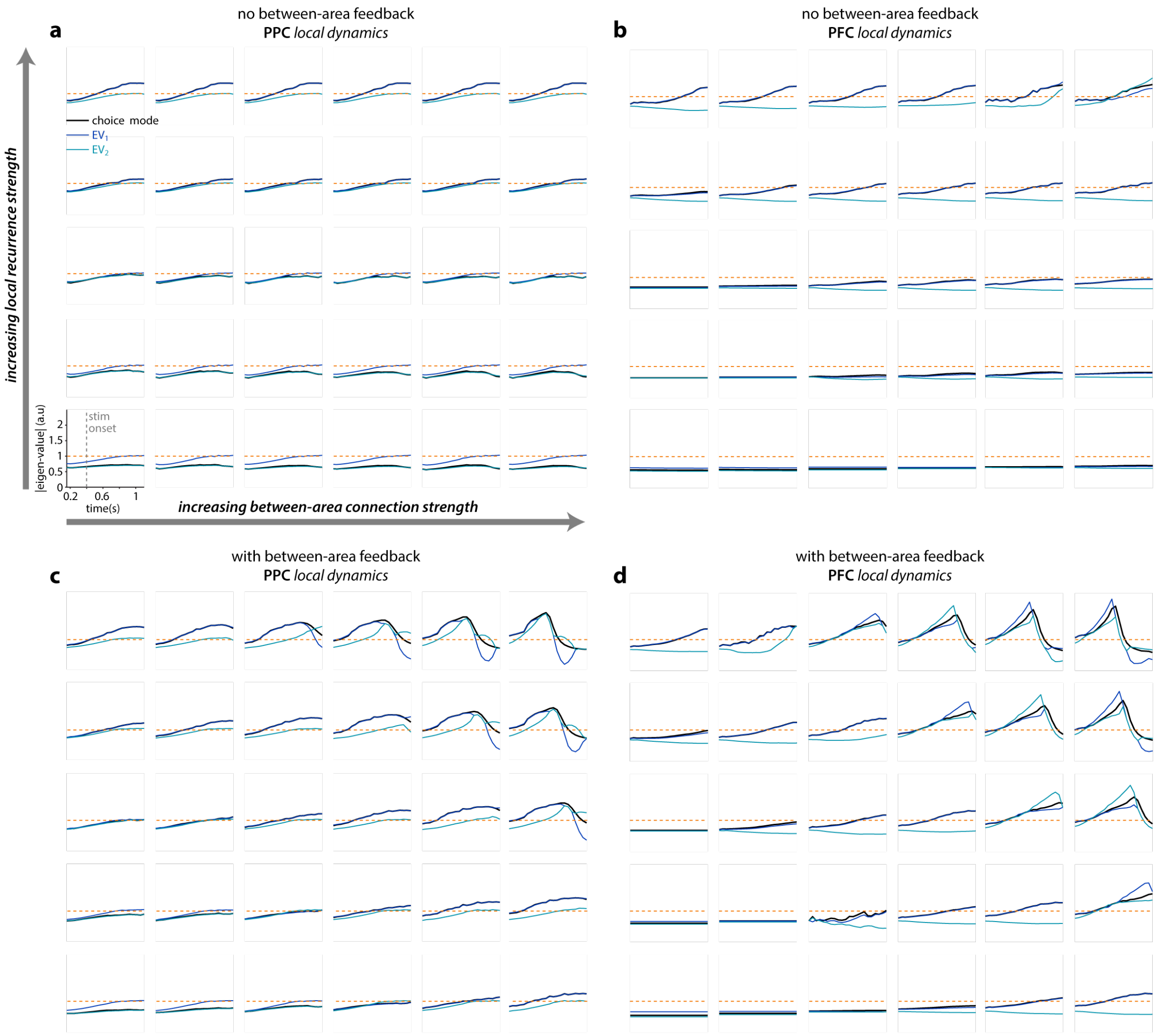
Extended Data Fig. 7: Cross-validation of hyper-parameters used for estimating residual dynamics

a*-*c, Representative results of the cross-validation procedure used to determine the various hyper-parameters of the analysis pipeline (see Extended Data Fig. 2) for neural data from a single task configuration in monkey T (config-3, see Extended Data Fig. 4). ***a***, Cross-validated hankel matrix reconstruction error (*E_{hankel}*; circle: mean over 20 repeats of hold-out cross validation; error bars: 1 s.e.m) plotted as a function of the rank of the hankel matrix (r, step 2 in Extended Data Fig. 2) for residuals from the two epochs (left: decision; right: movement) and two choices (blue: choice 1; red: choice 2). The reconstruction error for each of the 20 repeats was computed by assigning a random 50% of the trials as a "training" set and the rest as a "test" set. ***b***, 5-fold cross-validated mean squared error (circles: mean over 5 folds; error bars: 1 s.e.m) of the denoised residual predictions obtained from the first stage of the two-stage least squares regression (2SLS; step 3 in Extended Data Fig. 2), plotted as a function of the hyper-parameters: d (dimensionality of dynamics subspace); and l (number of past lags). For each cross-validation fold, a single mean squared error measure was computed by pooling the denoised predictions across time points in both epochs (left: choice 1; right: choice 2). ***c***, 5-fold cross-validated mean squared error (circle: average across 5 'repeats' of the 5-fold cross validation; error bars: 2 standard deviations across repeats) of the residual predictions obtained from the second stage of the 2SLS regression (step 4 in Extended Data Fig. 2), plotted as a function of the smoothness hyper-parameter α for different epochs (left: decision; right: movement) and choice (choice 1 and 2). Both the train (orange) and test (gray) error are shown. ***d***, Summary showing the optimal value for the dimensionality d and lag l (step 3 in Extended Data Fig. 2) for all task configurations and monkeys (symbols as in Extended Data Fig. 4b). A dimensionality of 8 and a lag of 3 was deemed optimal for both monkeys and task configurations (used in Fig 4). ***e***, Summary showing the optimal smoothness hyper-parameter α (step 4 in Extended Data Fig. 2) for all task configurations and monkeys. Final values of α were chosen to be the same across monkeys in Fig. 4 (decision epoch: α = 200; movement epoch: α = 50) despite a small degree of variability across the two monkeys. Same conventions as in ***d***.



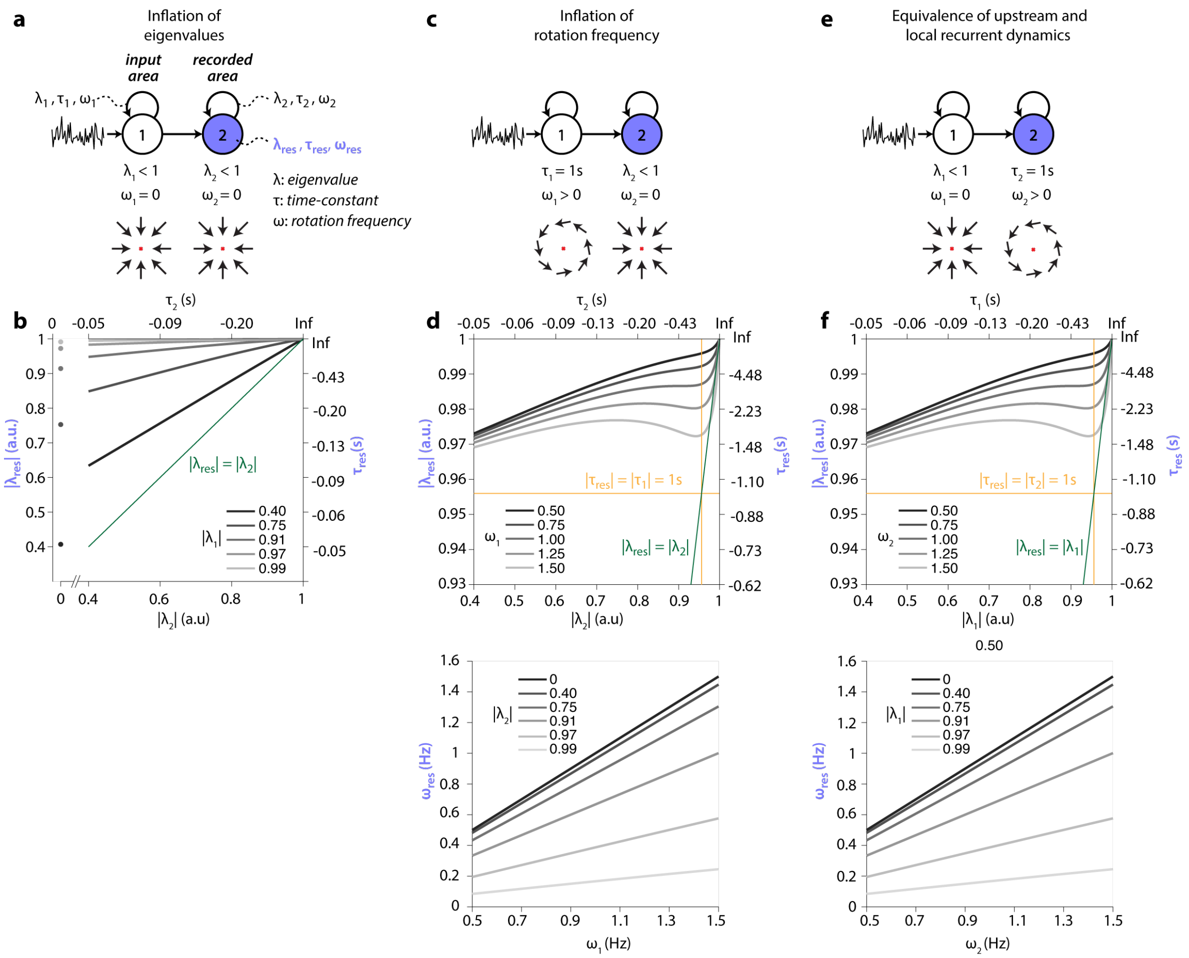
Extended Data Fig. 8: Effect of bin size on the estimated residual dynamics

We estimated the residual dynamics for different choices of bin size, to identify the smallest bin size resulting in unbiased estimates. In the discrete time formulation of a linear dynamical system, like the one we use here, re-binning of the responses trivially results in a scaling of the estimated eigenvalues of the residual dynamics. To compensate for this rescaling, here we "mapped" the estimated eigenvalues onto a common, reference bin size (see Methods, Effects of bin size). In the absence of statistical biases, the resulting "re-binned eigenvalue" would be independent of bin size. ***a****, Re-binned eigenvalues for simulations of a time-invariant, latent-variable (3 latent dimensions), LDS model (reference bin size = 40ms) as a function of bin-size (dashed line: ground truth; black curve: estimate). Estimates of the residual dynamics are biased for small bin sizes, but become unbiased when bin size is sufficiently large.* ***b****, Estimated re-binned eigenvalues (reference bin size = 15ms) as a function of bin size for all configurations in monkey T. Columns correspond to the 8 distinct eigenmodes of the estimated 8-dimensional residual dynamics (left to right, largest to smallest EV), rows correspond to task configurations (top to bottom, config-1 to 4; Extended Data Fig. 4b). Here the re-binned eigenvalues were computed separately for each choice (red vs blue) and averaged in small temporal windows specific to each epoch: 0.2-0.4s relative to stimulus onset (solid lines) and -0.15 to 0.25s relative to saccade onset (dashed lines). All main analyses of recorded neural responses are based on a bin size of 45ms, for which estimates of residual dynamics appear to converge to an asymptote. Note that the re-binned eigenvalues for a bin size of 45ms are larger than the corresponding eigenvalues reported in other figures (e.g. Fig. 4b), because the former were mapped onto a reference bin size of 15ms.*



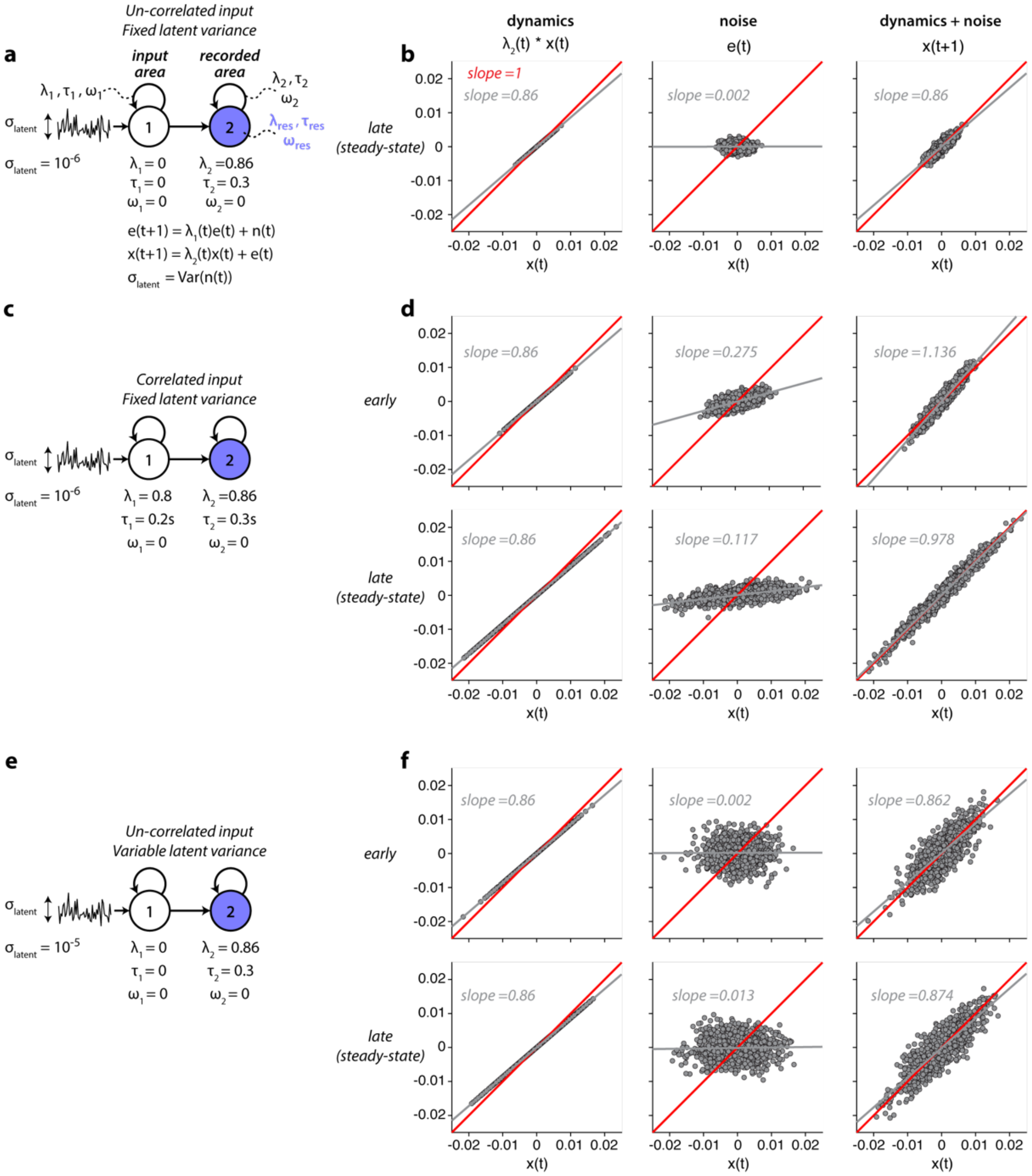
Extended Data Fig. 9: Time-varying local residual dynamics for all multi-area networks

*We estimated local residual dynamics separately in PPC or PFC of the two-area network models, as in Fig. 6 (PPC: **a**,**c**; PFC: **b**,**d**). We estimated either 2-dimensional dynamics based on the residuals in a given area, resulting in two time-varying eigenvalues (cyan and blue curves, analogous to Fig. 6b,e); or 1-dimensional dynamics based on the projection of the residuals onto the choice mode in each area (Fig. 6a,d), resulting in a single time-varying eigenvalue (black curves). The maximum value of the black curve in each sub-panel is shown in Fig. 6c,f. **a**, Local residual dynamics in PPC for the networks without feedback from PFC to PPC (Fig. 6c, left), for increasing strength of local recurrence (bottom to top) and between-area connectivity (left to right). **b**, Local residual dynamics in PFC, same networks as in **a** (Fig. 6c, right). **c**, Local residual dynamics in PPC for the networks with feedback from PFC to PPC (Fig. 6f, left), same conventions as in **a**. **d**, Local residual dynamics in PFC, same networks as in **c** (Fig. 6f, right). For networks with strong local recurrence and strong between-area connections, the neural population activity falls into one of two point-attractors before the end of the trial, resulting in a drop in the eigenvalues. The two point attractors implement the commitment to a choice by the network.*



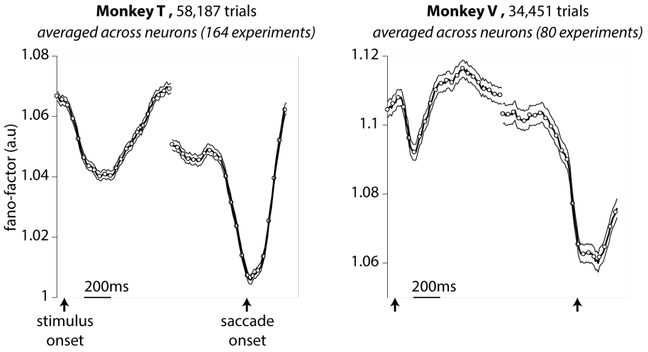
Extended Data Fig. 10: Inflation of local residual dynamics in a linear two-area dynamical system

We systematically explored the effect of correlated input variability on estimates of residual dynamics in a two-area, linear dynamical system. The input area implements 2d isotropic recurrent dynamics characterized by parameters λ_1, τ_1, and ω_1 (eigenvalue, time-constant, rotation frequency). Activity in the input area is externally driven by uncorrelated noise. Values of λ_1 closer to 1 result in longer auto-correlation times in the variability of activity in the input area. This activity provides the input into the recorded area, which implements 2d isotropic recurrent dynamics with parameters λ_2, τ_2, ω_2. Residual dynamics at steady-state is estimated from activity of the recorded area. At steady state, estimates can be derived analytically (see Supplementary Math Note B). Because of temporally correlated input variability, the properties of the residual dynamics (λ_{res}, τ_{res}, ω_{res}) in general do not match those of the recurrent dynamics in the recorded area. ***a-b****, Inflation of eigenvalues.* ***a****, Schematic of the model (top) and recurrent dynamics in each area (bottom, flow fields). Recurrent dynamics is stable and non-rotational in both areas.* ***b****, Residual dynamics (λ_{res}) in the recorded area as a function of recurrent dynamics in the recorded area (λ_2, x-axis) and in the input area (λ_1, gray lines). The eigenvalues of the residual dynamics are inflated, i.e. λ_{res} is larger than λ_2 for (all gray lines above the green line). Larger λ_1 (longer input auto-correlations) lead to stronger inflation. For λ_2 = 0 (no recurrent dynamics in the recorded area) λ_{res} = λ_1 (gray circles).* ***c-d****, Inflation of rotation frequency.* ***c****, Recurrent dynamics is rotational in the input area, but stable and non-rotational in the recorded area.* ***d****, Residual dynamics in the recorded area, expressed as the magnitude of the eigenvalue (λ_{res}, top) and the rotation frequency (ω_{res}, bottom). The eigenvalues of the residual dynamics are generally inflated (top), but the relation with λ_2 is non-monotonic and depends on ω_1. The residual dynamics is rotational (bottom, $\omega_{res} > 0$) even though the recurrent dynamics in the recorded area is not (ω_2= 0). The inflation of rotation frequency is reduced for increasing λ_2.* ***e-f****, Equivalence of upstream and local recurrent dynamics.* ***e****, Analogous to* ***c****, but dynamics is switched between input and recorded area.* ***f****, Analogous to* ***d****, but for the dynamics in* ***e****. The* *residual dynamics is identical to that in **d**. In general, residual dynamics in the recorded area reflects the combined effect of local and upstream recurrent dynamics.*



756 ***Extended Data Fig. 11: Explanation of input driven inflation in residual dynamics***

*To gain an intuitive understanding of inflation of eigenvalue magnitude, we consider simulations of two-area linear dynamical systems similar to those in Extended Data Fig. 10a. For simplicity, here we simulate stable 1d-dynamics in each area, whereby variability of the input into the recorded area is either correlated (**c-d**) or uncorrelated (**a-b**, **e-f**), and has fixed (**a-b**, **c-d**) or time-dependent latent noise variance (**e-f**). The variability injected into the input area is always uncorrelated. Recurrent dynamics in the recorded area is identical in all simulations. **a**, Model parameters for the case of uncorrelated input ($\lambda_1 = 0$). **b**, Contributions to activity x in the recorded area at steady-state. Activity x(t) (x-axis) is propagated through the recurrent dynamics (left, y-axis) and added to the noise e(t) (middle, y-axis) to obtain activity x(t+1) at time t+1 (right, y-axis). The noise e(t) corresponds to activity/output of the input area, and is shaped by dynamics determined by λ_1. Points in the scatter plots correspond to different simulated trials. Estimating the eigenvalue of the residual dynamics in the absence of observation noise amounts to measuring the slope of the regression line relating x(t) to x(t+1) (right, gray line). In this case, this slope is identical to that obtained if the latent noise had not been added to the activity (left, gray line), meaning that residual dynamics correctly reflects the effect of the recurrent dynamics in the recorded area (slope < 0, reflecting $\lambda_2 < 0$; left). **c**, Model parameters for the* *case of correlated input (λ_1 > 0 for t > 0; λ_1 = 0 at other times).* ***d****, Analogous to* ***b****, but for the model in* ***c****. Here activity and noise are shown at two times in the trial: early, when steady-state is not yet reached (top) and late, at steady-state (bottom). At both times, residual dynamics is inflated, i.e. the regression slope between x(t) and x(t+1) (right) is larger than that obtained by applying only the recurrent dynamics (left), indicating inflation of the eigenvalues. Inflation occurs because the noise itself is correlated with activity in the recorded area (middle, slope > 0), an effect that results indirectly from the correlation between e(t) and e(t-1). At steady state, even the inflated residual dynamics is still stable (bottom-right, slope < 1; see also Extended Data 10b). However, immediately after the onset of the correlated input, residual dynamics erroneously reveals an instability (top-right, slope > 1).* ***e****, Parameters for the case of uncorrelated noise but time-varying noise variance. The variance of the noise injected into the input area is increased at time t = 0, from σ_{latent} = 10^{-6} to 10^{-5}.* ***f,*** *A change in noise variance does not result in inflation of the residual dynamics, neither early nor late after the change (right, top and bottom; same slope as on the left; see also Extended Data Fig. 3a-c, squares).*



Extended Data Fig. 12: Quenching of variability in single PFC neurons

Trial-by-trial variability in single neurons is transiently reduced at the onset of specific task-events. We quantified single neuron variability as the time-varying mean-matched Fano-factor computed by pooling each recorded unit (100ms long time bins), across all experiments in a monkey (empty circles; dashed curve: 95% normal confidence intervals; left: Monkey T, right: Monkey V). In both monkeys, the mean-matched Fano factor undergoes a transient reduction locked to the onset of the stimulus and the onset of the saccade. The reduction in variability around the time of saccade onset coincides with a contraction of the eigenvalues of the residual dynamics (Fig. 4b,e), suggesting that more quickly decaying dynamics may underlie variability quenching at that time. A contraction of eigenvalues, however, does not appear necessary to explain variability quenching, as an analogous contraction is not observed at the time of stimulus onset, despite the consistent reduction in variability at stimulus onset.

References:

1. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).
2. Ahrens, M. B. *et al.* Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–477 (2012).
3. Saxena, S. & Cunningham, J. P. Towards the neural population doctrine. *Curr. Opin. Neurobiol.* **55**, 103–111 (2019).
4. Gao, P. *et al.* A theory of multineuronal dimensionality, dynamics and measurement. *bioRxiv* (2017) doi:10.1101/214262.
5. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. Computation Through Neural Population Dynamics. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
6. Brody, C. D., Romo, R. & Kepecs, A. Basic mechanisms for graded persistent activity: Discrete attractors, continuous attractors, and dynamic representations. *Curr. Opin. Neurobiol.* **13**, 204–211 (2003).
7. Rabinovich, M. I., Varona, P., Selverston, A. I. & Abarbanel, H. D. I. Dynamical principles in neuroscience. *Rev. Mod. Phys.* **78**, 1213–1265 (2006).
8. Brinkman, B. A. W., Rieke, F., Shea-Brown, E. & Buice, M. A. Predicting how and when hidden neurons skew measured synaptic interactions. *PLOS Comput. Biol.* **14**, e1006490 (2018).
9. Wilting, J. & Priesemann, V. Inferring collective dynamical states from widely unobserved systems. *Nat. Commun.* **9**, 2325 (2018).
10. Das, A. & Fiete, I. R. Systematic errors in connectivity inferred from activity in strongly recurrent networks. *Nat. Neurosci.* **23**, 1286–1296 (2020).
11. Sauerbrei, B. A. *et al.* Cortical pattern generation during dexterous movement is input-driven. *Nature* **577**, 386–391 (2020).
12. Churchland, A. K. *et al.* Variance as a Signature of Neural Computations during Decision Making. *Neuron* **69**, 818–831 (2011).
13. Churchland, M. M. *et al.* Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat. Neurosci.* **13**, 369–378 (2010).
14. Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
15. Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).
16. Goris, R. L. T., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nat. Neurosci.* **17**, 858–865 (2014).
17. Ecker, A. S. *et al.* State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–248 (2014).
18. Lin, I. C., Okun, M., Carandini, M. & Harris, K. D. The Nature of Shared Cortical Variability. *Neuron* **87**, 644–656 (2015).
19. Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E. & Doiron, B. The spatial structure of correlated neuronal variability. *Nat. Neurosci.* **20**, 107–114 (2017).
20. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical Areas Interact through a Communication Subspace. *Neuron* **102**, 249-259.e4 (2019).
21. Cowley, B. R. *et al.* Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron* **108**, 551-567.e8 (2020).
22. Rumyantsev, O. I. *et al.* Fundamental bounds on the fidelity of sensory cortical coding. *Nature* **580**, 100–105 (2020).
23. Li, N., Daie, K., Svoboda, K. & Druckmann, S. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* **532**, 459–464 (2016).

24. Chettih, S. N. & Harvey, C. D. Single-neuron perturbations reveal feature-specific competition in V1. *Nature* **567**, 334–340 (2019).
25. Inagaki, H. K., Fontolan, L., Romani, S. & Svoboda, K. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* **566**, 212–217 (2019).
26. Daie, K., Svoboda, K. & Druckmann, S. Targeted photostimulation uncovers circuit motifs supporting short-term memory. *Nat. Neurosci.* **24**, 259–265 (2021).
27. Jazayeri, M. & Afraz, A. Navigating the Neural Space in Search of the Neural Code. *Neuron* **93**, 1003–1014 (2017).
28. Salzman, C. D., Britten, K. H. & Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. *Nature* **346**, 174–177 (1990).
29. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in Neural Systems. *Neuron* **71**, 9–34 (2011).
30. Packer, A. M., Russell, L. E., Dalgleish, H. W. P. & Häusser, M. Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. *Nat. Methods* **12**, 140–146 (2015).
31. Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* **512**, 423–426 (2014).
32. Sadeh, S. & Clopath, C. Theory of neuronal perturbome in cortical networks. *Proc. Natl. Acad. Sci.* **117**, 26966–26976 (2020).
33. Buesing, L., Macke, J. H. & Sahani, M. Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. *Adv. Neural Inf. Process. Syst. NIPS* 1–9 (2012) doi:10.3109/0954898X.2012.677095.
34. Sani, O. G., Abbaspourazad, H., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nat. Neurosci.* **24**, 140–149 (2021).
35. Angrist, J. D. & Krueger, A. B. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J. Econ. Perspect.* **15**, 69–85 (2001).
36. Gold, J. I. & Shadlen, M. N. The Neural Basis of Decision Making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
37. Hanks, T. D. & Summerfield, C. Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron* **93**, 15–31 (2017).
38. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
39. Shadlen, M. N. & Newsome, W. T. Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
40. Wang, X.-J. Decision Making in Recurrent Neuronal Circuits. *Neuron* **60**, 215–234 (2008).
41. Wong, K.-F. & Wang, X.-J. A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *J. Neurosci.* **26**, 1314–1328 (2006).
42. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
43. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical Control of Arm Movements: A Dynamical Systems Perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
44. Svoboda, K. & Li, N. Neural mechanisms of movement planning: motor cortex and beyond. *Curr. Opin. Neurobiol.* **49**, 33–41 (2018).
45. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
46. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18**, 1025–1033 (2015).
47. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**, 925–933 (2013).
48. Murray, J. D., Jaramillo, J. & Wang, X.-J. Working Memory and Decision-Making in a Frontoparietal Circuit Model. *J. Neurosci.* **37**, 12167–12186 (2017).

49. Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48**, 661–673 (2005).

50. Yu, B. M. *et al.* Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. *J. Neurophysiol.* **102**, 614–635 (2009).

51. Machens, C. K., Romo, R. & Brody, C. D. Functional, But Not Anatomical, Separation of 'What' and 'When' in Prefrontal Cortex. *J. Neurosci.* **30**, 350–360 (2010).

52. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).

53. Brunton, B. W., Johnson, L. A., Ojemann, J. G. & Kutz, J. N. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci. Methods* **258**, 1–15 (2016).

54. Kobak, D. *et al.* Demixed principal component analysis of neural population data. *eLife* **5**, 1–36 (2016).

55. Williams, A. H. *et al.* Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron* **98**, 1099-1115.e8 (2018).

56. Amit, D. J. & Brunel, N. Dynamics of a recurrent network of spiking neurons before and following learning. *Netw. Comput. Neural Syst.* **8**, 373–404 (1997).

57. Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).

58. Machens, C. K. Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination. *Science* **307**, 1121–1124 (2005).

59. Rigotti, M., Ben Dayan Rubin, D., Wang, X.-J. & Fusi, S. Internal Representation of Task Rules by Recurrent Dynamics: The Importance of the Diversity of Neural Responses. *Front. Comput. Neurosci.* **4**, 1–29 (2010).

60. Mastrogiuseppe, F. & Ostojic, S. Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks. *Neuron* **99**, 609–623 (2018).

61. Chaisangmongkon, W., Swaminathan, S. K., Freedman, D. J. & Wang, X.-J. Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. *Neuron* **93**, 1504-1517.e4 (2017).

62. Sussillo, D. & Abbott, L. F. Generating Coherent Patterns of Activity from Chaotic Neural Networks. *Neuron* **63**, 544–557 (2009).

63. Carnevale, F., de Lafuente, V., Romo, R., Barak, O. & Parga, N. Dynamic Control of Response Criterion in Premotor Cortex during Perceptual Detection under Temporal Uncertainty. *Neuron* **86**, 1067–1077 (2015).

64. Wang, J., Narain, D., Hosseini, E. A. & Jazayeri, M. Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* **21**, 102–112 (2018).

65. Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X.-J. Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* **22**, 297 (2019).

66. Sohn, H., Narain, D., Meirhaeghe, N. & Jazayeri, M. Bayesian Computation through Cortical Latent Dynamics. *Neuron* **103**, 934-947.e5 (2019).

67. Paninski, L. *et al.* A new look at state-space models for neural data. *J. Comput. Neurosci.* **29**, 107–126 (2010).

68. Lawhern, V., Wu, W., Hatsopoulos, N. & Paninski, L. Population decoding of motor cortical activity using a generalized linear model with hidden states. *J. Neurosci. Methods* **189**, 267–280 (2010).

69. Macke, J. H. *et al.* Empirical models of spiking in neural populations. *Adv. Neural Inf. Process. Syst.* **24**, 1350–1358 (2011).

70. Smith, A. C. & Brown, E. N. Estimating a State-Space Model from Point Process Observations. *Neural Comput.* **15**, 965–991 (2003).

71. Kao, J. C. *et al.* Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat. Commun.* **6**, 1–12 (2015).

72. Gao, Y., Archer, E. W., Paninski, L. & Cunningham, J. P. Linear dynamical neural population models through nonlinear embeddings. in *Advances in Neural Information Processing Systems* vol. 29 (2016).

73. Linderman, S. W. *et al.* Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems. *Proc. 20th Int. Conf. Artif. Intell. Stat.* **54**, 914–922 (2017).

74. Zhao, Y. & Park, I. M. Variational Latent Gaussian Process for Recovering Single-Trial Dynamics from Population Spike Trains. *Neural Comput.* **29**, 1293–1316 (2017).

75. Pandarinath, C. *et al.* Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* **15**, 805–815 (2018).

76. Seung, H. S. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci.* **93**, 13339–13344 (1996).

77. Smith, P. L. & Ratcliff, R. Psychology and neurobiology of simple decisions. *Trends Neurosci.* **27**, 161–168 (2004).

78. Usher, M. & McClelland, J. L. The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.* **108**, 550–592 (2001).

79. Brunton, B. W., Botvinick, M. M. & Brody, C. D. Rats and humans can optimally accumulate evidence for decision-making. *Science* **340**, 95–98 (2013).

80. Hanks, T. D. *et al.* Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* **520**, 220–223 (2015).

81. Kiani, R. *et al.* Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* **85**, 1359–1373 (2015).

82. Murphy, B. K. & Miller, K. D. Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns. *Neuron* **61**, 635–648 (2009).

83. Goldman, M. S. Memory without Feedback in a Neural Network. *Neuron* **61**, 621–634 (2009).

84. Hennequin, G., Vogels, T. P. & Gerstner, W. Non-normal amplification in random balanced neuronal networks. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* **86**, 1–12 (2012).

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

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

87. Joglekar, M. R., Mejias, J. F., Yang, G. R. & Wang, X.-J. Inter-areal Balanced Amplification Enhances Signal Propagation in a Large-Scale Circuit Model of the Primate Cortex. *Neuron* **98**, 222-234.e8 (2018).

88. Schall, J. D. The neural selection and control of saccades by the frontal eye field. *Philos. Trans. R. Soc. B Biol. Sci.* **357**, 1073–1082 (2002).

89. Schall, J. D. Visuomotor Functions in the Frontal Lobe. *Annu. Rev. Vis. Sci.* **1**, 469–498 (2015).

90. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).

91. Murray, J. D. *et al.* Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci.* **114**, 394–399 (2017).

92. Hunt, L. T., Behrens, T. E. J., Hosokawa, T., Wallis, J. D. & Kennerley, S. W. Capturing the temporal evolution of choice across prefrontal cortex. *eLife* **4**, 1–25 (2015).

93. Thura, D. & Cisek, P. Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron* **81**, 1401–1416 (2014).

94. Druckmann, S. & Chklovskii, D. B. Neuronal Circuits Underlying Persistent Representations Despite Time Varying Activity. *Curr. Biol.* **22**, 2095–2103 (2012).

95. Kaufman, M. T. *et al.* The Largest Response Component in the Motor Cortex Reflects Movement Timing but Not Movement Type. *eneuro* **3**, ENEURO.0085-16.2016 (2016).

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

97. Duncker, L., O'Shea, D., Shenoy, K. V. & Sahani, M. A dynamical model with E/I balance explains robustness to optogenetic stimulation in motor cortex. in *Cosyne Abstracts 2020*.

98. Chaudhuri, R., Knoblauch, K., Gariel, M. A., Kennedy, H. & Wang, X.-J. A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. *Neuron* **88**, 419–431 (2015).

99. Markov, N. T. *et al.* A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex. *Cereb. Cortex* **24**, 17–36 (2014).

100. Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* **22**, 1512–1520 (2019).

101. Nieh, E. H. *et al.* Geometry of abstract learned knowledge in the hippocampus. *Nature* **595**, 80–84 (2021).

102. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: Permitting preparation without movement. *Nat. Neurosci.* **17**, 440–448 (2014).

103. Semedo, J. D., Gokcen, E., Machens, C. K., Kohn, A. & Yu, B. M. Statistical methods for dissecting interactions between brain areas. *Curr. Opin. Neurobiol.* **65**, 59–69 (2020).

104. Murray, J. D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).

105. Hernández, A. *et al.* Decoding a Perceptual Decision Process across Cortex. *Neuron* **66**, 300–314 (2010).

106. Kiani, R., Hanks, T. D. & Shadlen, M. N. Bounded Integration in Parietal Cortex Underlies Decisions Even When Viewing Duration Is Dictated by the Environment. *J. Neurosci.* **28**, 3017–3029 (2008).

107. Aoi, M. C., Mante, V. & Pillow, J. W. Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nat. Neurosci.* **23**, 1410–1420 (2020).

108. Libby, A. & Buschman, T. J. Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* **24**, 715–726 (2021).

109. Durstewitz, D. & Seamans, J. K. Beyond bistability: Biophysics and temporal dynamics of working memory. *Neuroscience* **139**, 119–133 (2006).

110. Deco, G. & Jirsa, V. K. Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors. *J. Neurosci.* **32**, 3366–3375 (2012).

111. Dahmen, D., Grün, S., Diesmann, M. & Helias, M. Second type of criticality in the brain uncovers rich multiple-neuron dynamics. *Proc. Natl. Acad. Sci.* **116**, 13051–13060 (2019).

112. Kozachkov, L., Lundqvist, M., Slotine, J.-J. & Miller, E. K. Achieving stable dynamics in neural circuits. *PLOS Comput. Biol.* **16**, e1007659 (2020).

113. Purcell, B. A., Heitz, R. P., Cohen, J. Y. & Schall, J. D. Response variability of frontal eye field neurons modulates with sensory input and saccade preparation but not visual search salience. *J. Neurophysiol.* **108**, 2737–2750 (2012).

114. Chang, M. H., Armstrong, K. M. & Moore, T. Dissociation of Response Variability from Firing Rate Effects in Frontal Eye Field Neurons during Visual Stimulation, Working Memory, and Attention. *J. Neurosci.* **32**, 2204–2216 (2012).

115. Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M. & Miller, K. D. The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability. *Neuron* **98**, 846-860.e5 (2018).

116. Litwin-Kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* **15**, 1498–1505 (2012).

117. Deco, G. & Hugues, E. Neural network mechanisms underlying stimulus driven variability reduction. *PLoS Comput. Biol.* **8**, (2012).