

# Residual dynamics resolves recurrent contributions to neural computation

Aniruddh R. Galgali<sup>1,2</sup>, Maneesh Sahani<sup>3</sup> and Valerio Mante<sup>1,2</sup>

<sup>1</sup> Institute of Neuroinformatics, University of Zurich & ETH Zurich, Zurich, Switzerland

<sup>2</sup> Neuroscience Center Zurich, University of Zurich & ETH Zurich, Zurich, Switzerland

<sup>3</sup> Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom

Corresponding Authors: [galgalia@ini.uzh.ch](mailto:galgalia@ini.uzh.ch), [valerio@ini.uzh.ch](mailto:valerio@ini.uzh.ch)

## Abstract

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

## Introduction

14 Perception, decisions, and the resulting actions reflect neural computations implemented by  
15 large, interacting neuronal populations acting in concert. Inferring the nature of these  
16 interactions from recordings of neural activity is a key step towards uncovering the neural  
17 computations underlying behavior<sup>1-4</sup>. One promising approach is based on the premise that  
18 neural computations reflect the action of a dynamical system<sup>5-7</sup>, whereby the computations  
19 implemented by a neural population emerge from the interplay between external inputs into  
20 a distributed neural population and the internal dynamics resulting from the recurrent  
21 connections between neurons. The utility of such a “computation-through-dynamics”

22 framework hinges critically on our ability to characterize the nature of this interplay, and  
23 disentangle the individual contributions of inputs and recurrent dynamics. In practice,  
24 disentangling these two factors based on recordings of neural responses alone is challenging,  
25 as typically neither the exact properties of the inputs into a brain area, nor the nature of  
26 recurrent connectivity within and across areas, are known a priori<sup>8-11</sup>.

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

43 Our findings are organized in three sections. First, we illustrate the challenges in  
44 disentangling inputs and recurrent dynamics based on the simulations of a few, simple  
45 dynamical systems (Fig. 1-2). These dynamical systems are analogous to single-area,  
46 artificial recurrent neural networks (RNN) previously proposed for explaining the network-  
47 level mechanisms underlying sensory evidence integration<sup>36-42</sup> and movement generation in  
48 cortical areas<sup>43-47</sup>. We demonstrate that our estimates of residual dynamics can reveal the  
49 essential features of the computations implemented by these models, even when the time-  
50 course of the inputs are unknown. Second, we study neural population recordings from pre-  
51 frontal cortex (PFC) of macaque monkeys during decision-making and saccadic choices (Fig.  
52 3-5). While neural population trajectories in PFC are consistent with a number of previously  
53 proposed models of evidence integration and movement generation, we are able to rule out  
54 several candidate models based on the properties of the inferred residual dynamics. Third,  
55 we analyze simulated responses of a previously proposed multi-area RNN model of decision-  
56 making<sup>48</sup>, to illustrate how inferred residual dynamics can be used to deduce circuit-level  
57 implementations of distributed recurrent computations. (Fig. 6-8).

## Results

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

### Neural population trajectories poorly constrain recurrent computations

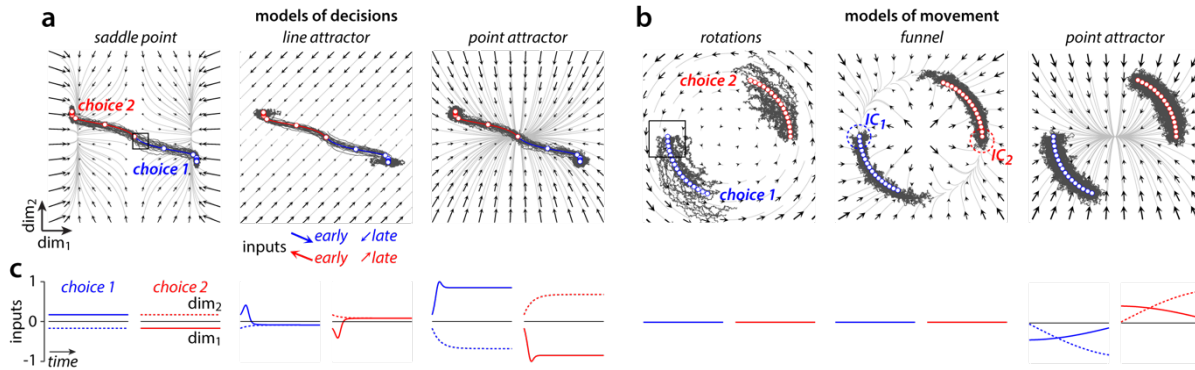
68 To illustrate the nature of this problem, we consider simulated responses of a number of  
69 distinct models of neural population dynamics during perceptual decision-making<sup>40</sup> and  
70 movement generation<sup>43</sup> (Fig. 1). While these hand-designed models are not meant to  
71 precisely reproduce neural recordings, they do capture the distinctive features of rather  
72 complex, non-linear RNNs trained to integrate sensory evidence towards a choice<sup>41,42,76</sup> (Fig.  
73 1a) or generate complex motor sequences<sup>46,47</sup> (Fig. 1b).

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

$$78 \quad \dot{\mathbf{z}}_t = \mathbf{F}(\mathbf{z}_t) + \mathbf{u}_t + \epsilon_t \quad (1)$$

79 Any solution to the above equation is also determined by the *initial condition*  $\mathbf{z}_0$  (the neural  
80 state at the start of the trial). Differences in responses across task-conditions (e.g., different  
81 choices or movements) are explained by allowing  $\mathbf{u}_t$  or  $\mathbf{z}_0$  to vary across conditions (Fig. 1c,  
82 red vs. blue; Fig. 1b, initial conditions IC<sub>1</sub> vs. IC<sub>2</sub>).

83 We simulated single-trial responses for two task-conditions and represented them as  
84 trajectories in a 2-dimensional neural state-space (Fig. 1a,b, choice 1 & 2; dark-gray curves).  
85 The recurrent dynamics ( $\mathbf{F}$ ) can be represented as a flow field (Fig. 1a,b, black arrows and  
86 light-gray curves), which describes how the instantaneous neural state ( $\mathbf{z}_t$ ) evolves from a  
87 given location in state-space in the absence of inputs and noise. The action of the external  
88 input ( $\mathbf{u}_t$ ) corresponds to injecting a pattern of activity into the neural population, and  
89 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.

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

95 Critically, the simulations show that very different combinations of these factors can result  
 96 in very similar trajectories. For example, the three models of decision-making differ in the  
 97 nature of their inputs and recurrent dynamics, each mimicking a specific behavioral  
 98 “strategy” for perceptual decision-making<sup>36–38,77–80</sup>, from unstable, impulsive decisions (Fig.  
 99 1a, saddle point), to optimal accumulation of evidence (Fig1a, line attractor), and leaky,  
 100 forgetful accumulation (Fig. 1a, point attractor). Yet, for the chosen inputs, which depending  
 101 on the model are either constant (Fig. 1c, saddle point) or transient (Fig. 1c, line and point  
 102 attractor), all three models result in similar single-trial trajectories (Fig. 1a, gray curves) and  
 103 essentially indistinguishable condition-averaged trajectories (Fig. 1a, blue and red curves).

104 Analogous observations hold for the models of movement generation (Fig. 1b). Two of the  
 105 models have no inputs, and are driven entirely by recurrent dynamics starting from

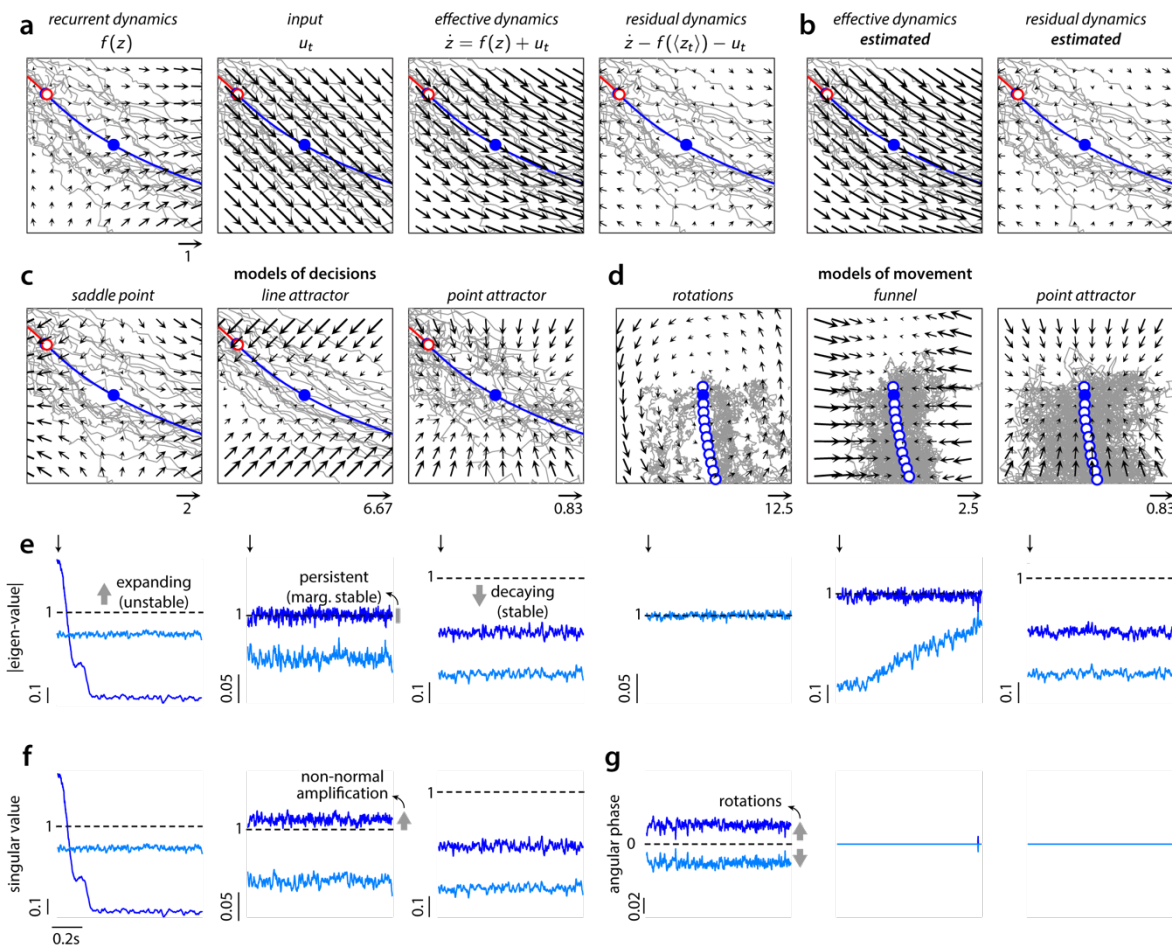
106 condition-dependent initial conditions—one model implements rotational dynamics<sup>45,46</sup>,  
107 implying that any variability in the initial condition on a given trial is reflected throughout  
108 the entire trajectory (Fig. 1b rotations; gray curves); the other implements what we refer to  
109 as “funnel” dynamics<sup>47,61</sup>, whereby neural activity is pushed through a narrow channel in  
110 state space, and any variability along directions orthogonal to the channel is suppressed (Fig.  
111 1b, funnel). In the third model, the recurrent dynamics implements a point attractor, and  
112 responses are mostly input driven<sup>11</sup> (Fig. 1b, point attractor). The simulated condition-  
113 averages can neither distinguish between the models with or without inputs, nor between  
114 the different recurrent dynamics associated with models that lack an input.

### Residual dynamics as a window onto recurrent dynamics

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

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

129 The residual dynamics describes how a perturbation away from the condition-averaged  
130 neural state would evolve relative to the trajectory over the course of one time-step. In Fig.  
131 2c,d, the blue dot indicates the unperturbed, “reference” neural state, which lies along the  
132 average trajectory. The tail of each arrow indicates the residual (the perturbed state), and  
133 the arrow-head shows how this residual evolves over one time-step. For the saddle point  
134 model (Fig. 2c, saddle point), perturbations along the horizontal direction, away from the  
135 trajectory, expand over time (arrows point away from the reference state), whereas  
136 perturbations along the vertical direction decay back to the trajectory (arrows point towards  
137 the reference state). These dynamics correctly reflect the influence of a saddle point in the  
138 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.

139 correctly reveals line attractor and point attractor dynamics in the other two models of

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

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

169 We summarize the residual dynamics through the main properties of the estimated local  
170 linear dynamical systems, specifically the magnitude of the eigen-values (EV), the singular  
171 values (SV), and the rotation frequency associated with the EV (Fig. 2e-g). For locations close  
172 to the saddle point in the model of decision-making, one of the EV is larger than 1, implying  
173 that perturbations along the associated eigen-vector (the horizontal direction in Fig. 1a, left)  
174 *expand* over time; the other EV is smaller than one, corresponding to *decay* along the vertical  
175 direction (Fig. 1a, left; center of flow field; Fig. 2e, left-most panel; early times). A line  
176 attractor results in a single EV of 1 (Fig. 2e, second from left) as horizontal perturbations are

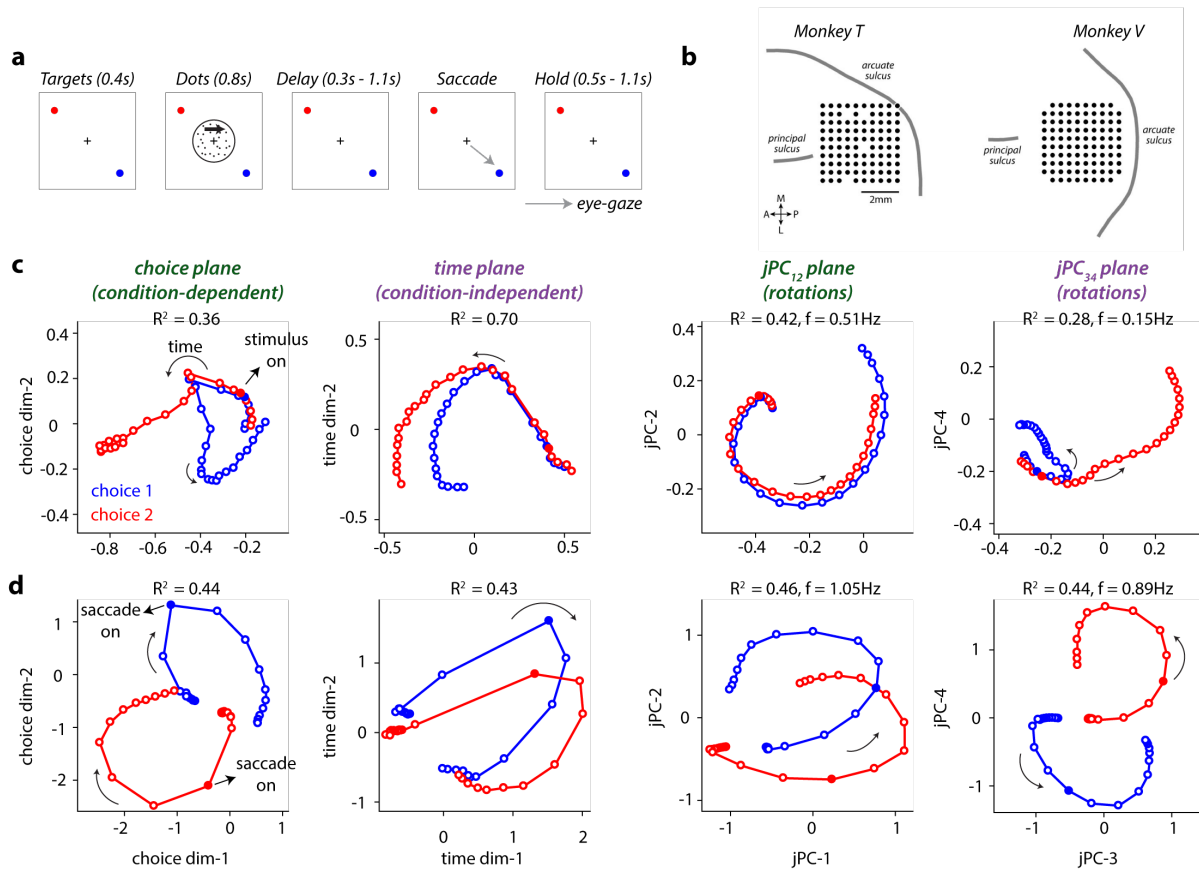
177 *persistent*, i.e. neither expand nor decay, and a point attractor in all EV smaller than 1 (Fig.  
178 2e, third from left; all directions decay). Rotational dynamics results in EV that are complex-  
179 valued and thus associated with a non-zero rotation frequency (Fig. 2g). Finally, differences  
180 between the magnitude of SV and EV reflect non-normal dynamics, a critical feature of a  
181 number of previous models of neural computation<sup>82-87</sup>. The SV larger than 1 in the line  
182 attractor model implies that small perturbations along the corresponding right singular  
183 vector transiently expand, even though they are persistent (EV=1) or decay (EV<1) over  
184 longer time-scales (Fig. 2e,f).

### Neural population responses of decisions and movements in PFC

185 We compared these model dynamics to neural population responses recorded in the pre-  
186 frontal cortex (PFC; area 8Ar) of two macaque monkeys performing a saccade-based  
187 perceptual decision-making task<sup>39,81,88,89</sup> (Fig 3a,b; Extended Data Fig. 4). To increase the  
188 statistical power of our analyses, we employed a dimensionality reduction technique to  
189 “align” the task-related subspaces of neural activity from different experiments with a  
190 similar task-configuration (Extended Data Fig. 4; 14-61 experiments per configuration; 150-  
191 200 units per experiment). This alignment yielded a shared, 20-dimensional neural state-  
192 space explaining >90% of task-related variance in the average neural responses measured  
193 across different experiments<sup>90</sup> (Extended Data Fig. 5). All the analyses below are performed  
194 within this aligned subspace, although the main results can be reproduced from sufficiently  
195 long single experiments (Extended Data Fig. 6).

196 The condition-averaged population trajectories in PFC shared important features with the  
197 average trajectories of the models in Fig. 1. We visualized the population trajectories through  
198 projections onto four distinct, two-dimensional activity subspaces: a “choice” plane,  
199 emphasizing choice-related activity; a “time” plane, emphasizing time-varying activity  
200 common to both choices; and two “jPC” planes<sup>45</sup>, emphasizing rotational dynamics (Fig. 3c,d;  
201 left to right). We estimated these planes separately during a decision-epoch, which coincided  
202 with the presentation of a random-dots stimulus (Fig 3c), and during a movement-epoch  
203 aligned to the execution of the saccade (Fig. 3d). As in the decision-models (Fig. 1a), PFC  
204 responses started in an undifferentiated state prior to stimulus onset (Fig 3c; choice plane;  
205 filled dots mark stimulus onset) and gradually diverged based on the upcoming choice of the  
206 animal (Fig. 3c, red vs. blue). PFC responses during the movement period showed  
207 pronounced rotational components (Fig. 3d, jPC<sub>12</sub> plane; filled dots mark movement onset)  
208 similar to those in the movement models (Fig. 1b). Prior to saccade-onset, PFC responses fell  
209 into largely stationary, choice-dependent states and then transitioned into rotational  
210 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<sub>12</sub>, and jPC<sub>34</sub> 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).

211 The measured PFC responses also differed from the model responses in several ways.  
 212 Consistent with past reports of population dynamics during decisions, working memory and  
 213 movements, PFC responses reflected strong condition-independent components during both  
 214 task-epochs (e.g. Fig. 3c,d, time-plane) <sup>23,25,51,91-95</sup>. Such condition-independent components  
 215 were not implemented in the models in Fig 1. Unlike in the models, pronounced choice-  
 216 related activity occurred along more than one state-space direction (Fig. 3c, choice plane)  
 217 and rotational dynamics within more than one plane. Moreover, rotational dynamics was

218 observed also during the decision-epoch (Fig. 3c, jPC planes). As for the models in Fig. 1, it is  
219 not clear which of these features of the condition-averaged trajectories reflect the influence  
220 of inputs, recurrent dynamics, or both.

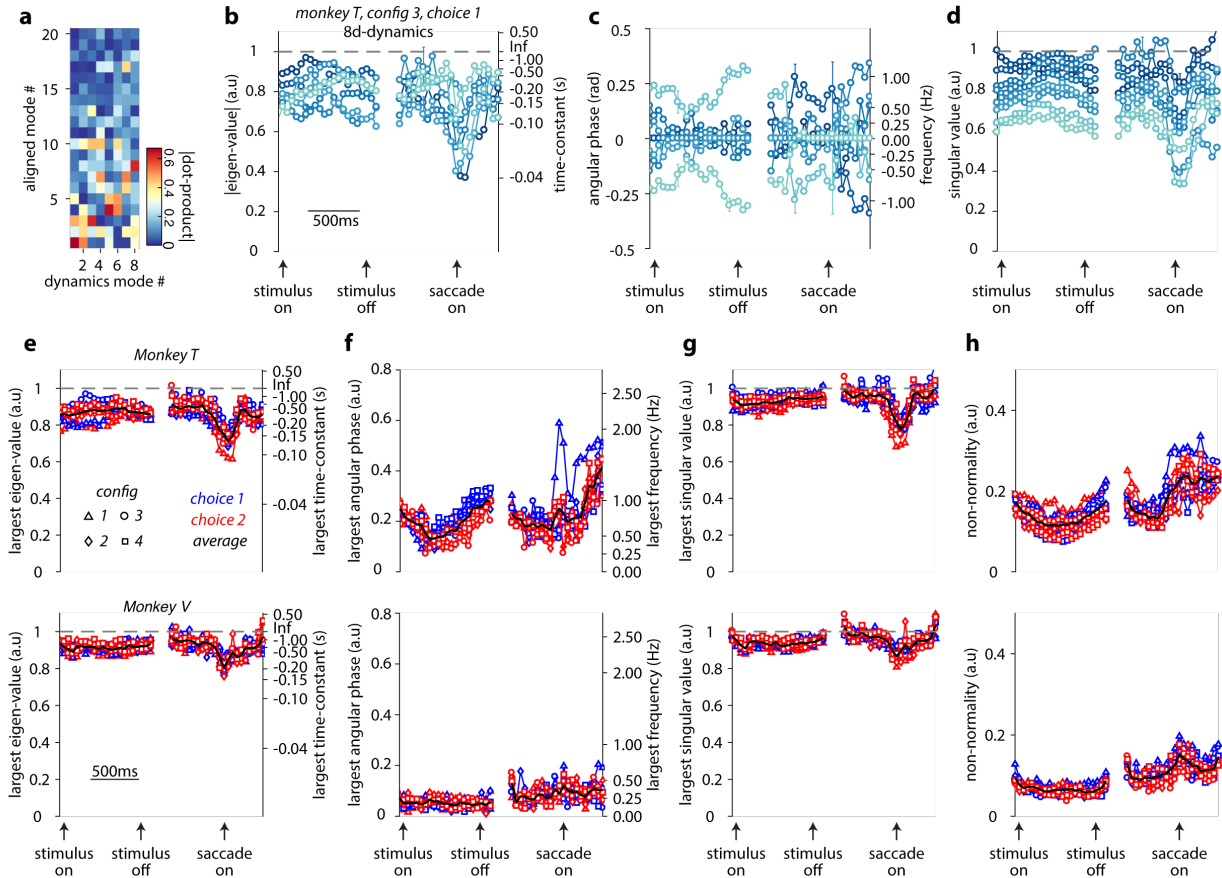
### Residual dynamics in PFC

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

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

239 The EV magnitudes were strongly time-dependent. For all task configurations, the largest EV  
240 were attained during the decision epoch or the delay period preceding the saccade. These  
241 EV were associated with decay time-constants in the range 187-745ms during the decision  
242 period (0s to +0.8s following stimulus onset) and 110-913ms during the delay period (-0.5s  
243 to +0.3s relative to saccade onset) for monkey T (95% CI, medians = 352ms and 293ms; Fig.  
244 4e, top), and 309-1064ms and 192-3586ms for monkey V (95% CI, medians = 489ms and  
245 491ms; Fig. 4e, bottom). Concurrently with the saccade onset, the EV consistently underwent  
246 a strong contraction—the largest measured time constants at saccade onset fell to median  
247 values of 159ms in monkey T and 310ms in monkey V (Fig. 4e), implying that perturbations  
248 away from the average trajectory during movement quickly fall back to the trajectory.

249 These findings alone rule out several models of recurrent dynamics in PFC. Even the largest  
250 EV in PFC during the decision epoch are inconsistent with unstable dynamics (EV>1, Figs.  
251 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.

252 from persistent dynamics ( $EV \approx 1$ , Figs. 1a,2e; line attractor). Likewise, the small EV around  
 253 the time of the saccade are inconsistent with purely rotational or funnel dynamics, which  
 254 would both result in directions with very slow decay ( $EV \approx 1$ , Figs. 1b,2e; rotations and

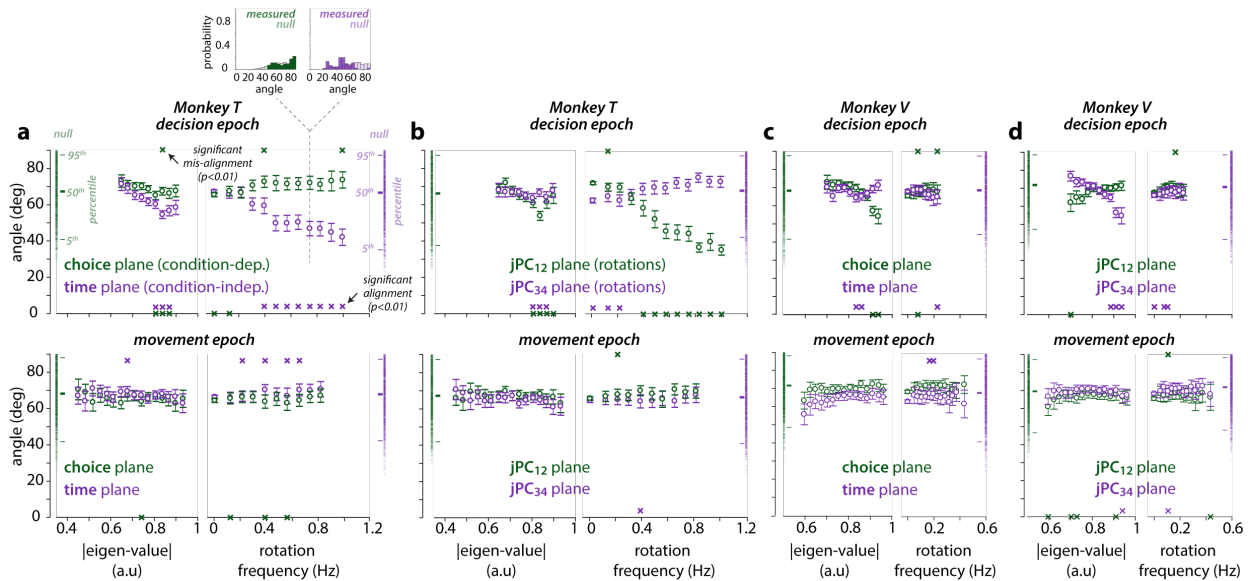
255 funnel). Rather, the inferred EV are consistent with quickly decaying recurrent dynamics  
256 (Figs. 1b,2e; point attractor).

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

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

### **Condition-averaged trajectories reflect time-dependent input contributions**

272 Additional insights into the relative strengths of recurrent dynamics and inputs can be  
273 gained by comparing the properties of residual dynamics and condition-averaged  
274 trajectories. When inputs are weak, the trajectories mostly reflect the properties of the  
275 recurrent dynamics, which in turn results in distinct relations between trajectories and  
276 residual dynamics. For example, in the saddle-point and line-attractor models, the condition-  
277 averaged trajectories for the two choices diverge along a direction that is closely aligned with  
278 the eigenvector associated with the largest EV in the residual dynamics (Fig. 1a, left-most  
279 panels; horizontal direction; Fig. 2e). Similarly, in the funnel and rotation models, the  
280 condition-averaged trajectories rotate in the plane containing residual dynamics with EV  
281 close to 1 (Fig. 1b, left-most panels; Fig. 2e), or EV with large angular phase (Fig. 1b, left  
282 panels; Fig. 2g). When such relations are absent, two scenarios are possible. First, the neural  
283 trajectories may mostly be driven by a strong input (Fig. 1b,2e, point-attractor model:  
284 trajectories rotate, whereas residual dynamics is decaying and non-rotational). Second, the  
285 recurrent dynamics may implement strong non-normal amplification, where population  
286 trajectories can display pronounced excursions along directions that are largely orthogonal  
287 to the eigenvectors associated with the largest EV<sup>82,85,96,97</sup>. While the latter scenario is ruled  
288 out by the properties of the residual dynamics (Fig. 4g,h,  $SV \leq 1$ ; no non-normal  
289 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<sub>12</sub>, and jPC<sub>34</sub> 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<sub>12</sub> and jPC<sub>34</sub> planes (**b**), in monkey T. **c-d**, same as **a-b**, for monkey V.

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

299 During the saccade epoch, subspace angles with the jPC planes showed no dependency on  
 300 either the rotation frequency or the magnitude of the associated EV in both monkeys (Fig.  
 301 5a-d, bottom). In fact, the mean subspace angle obtained for any given EV magnitude or  
 302 rotation frequency closely matched that expected from the null distribution (Fig. 5a-d,  
 303 bottom; vertically aligned green and purple points on the left and right). The prominent  
 304 rotations in the condition-averages during the saccade epoch (Fig. 3d) are thus not

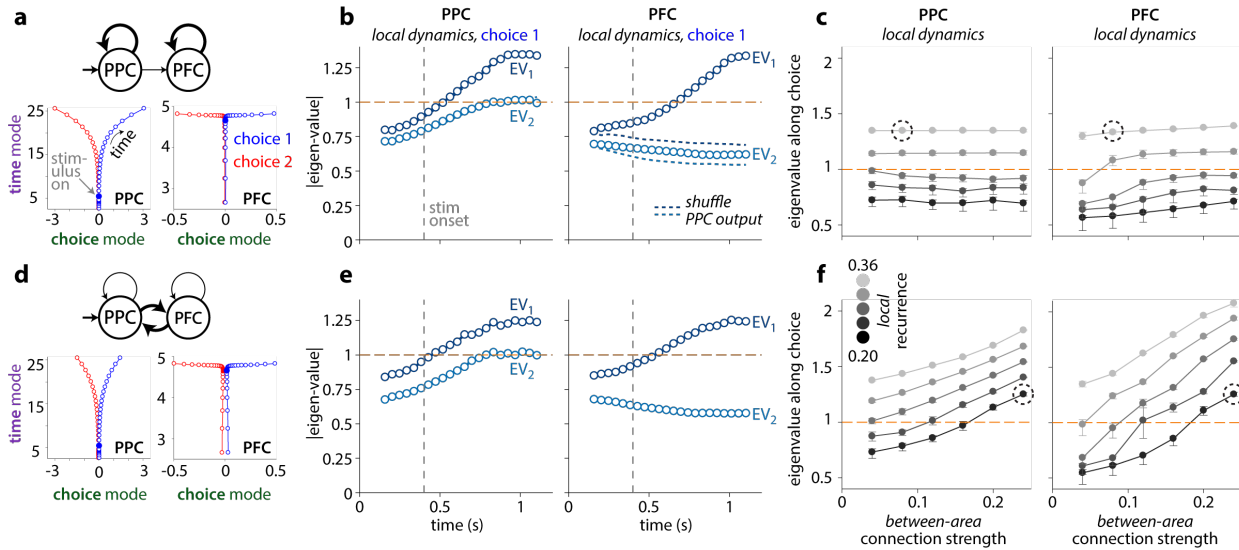
305 preferentially aligned with eigenvectors associated with EV of large magnitude or large  
306 rotation frequency. This finding is inconsistent with the predictions of the rotation and  
307 funnel models, and instead suggests a prominent role of inputs in driving saccade-related  
308 activity (Fig. 1b, point attractor).

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

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

### **Interpreting local residual dynamics in distributed cortical circuits**

328 These above conclusions, however, are based on a comparison to simplified models of neural  
329 dynamics, for which inputs and recurrent contributions are well defined (Fig. 1). Biological  
330 circuits tend to be modular, i.e. are subdivided into areas, with both local recurrence within  
331 areas, as well as long-range, feedforward or feedback connections between areas<sup>98,99</sup>. At the  
332 level of any single area, a clear distinction between inputs and recurrent dynamics may then  
333 be challenging, raising the question of how residual dynamics should be interpreted when  
334 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. **f**, Same conventions as in **a-c**, but for networks with between-area feedback.

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

344 Simulated responses of a random-dots task show choice-dependent and condition-  
 345 independent components, both in PPC and PFC (Fig. 6a,d; choice and time modes). The EV of  
 346 the residual dynamics, estimated *locally* in PPC or PFC, are typically time-dependent (Fig.  
 347 6b,e, Extended Data Fig. 9). In particular, the dynamics can change from stable (EV<1) to

348 unstable ( $EV > 1$ ) after the input is turned on, reflecting the non-linear nature of these  
349 networks.

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

358 In networks lacking feedback between areas, the residual dynamics in PPC naturally only  
359 reflects the local recurrence, whereby the largest EV gradually increases with stronger local  
360 recurrence. (Fig. 6c, PPC). The residual dynamics in PFC closely resembles that in PPC (Fig.  
361 6c, PFC), but this resemblance conceals a critical difference between the two areas. In PPC,  
362 the residual dynamics reflects the properties of the local recurrent dynamics. The same is  
363 not true in PFC, where any  $EV > 1$  mostly reflects recurrent dynamics implemented upstream,  
364 in PPC. Indeed, if the output of PPC is “shuffled” to remove any temporal correlations, while  
365 retaining its time-varying mean, the EV estimated in PFC fall below 1, indicating that  
366 recurrent dynamics in PFC is actually decaying in these networks (Fig. 6b, dashed). We refer  
367 to this effect as an “inflation” of the EV in PFC, due to the correlated input from PPC.

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

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



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

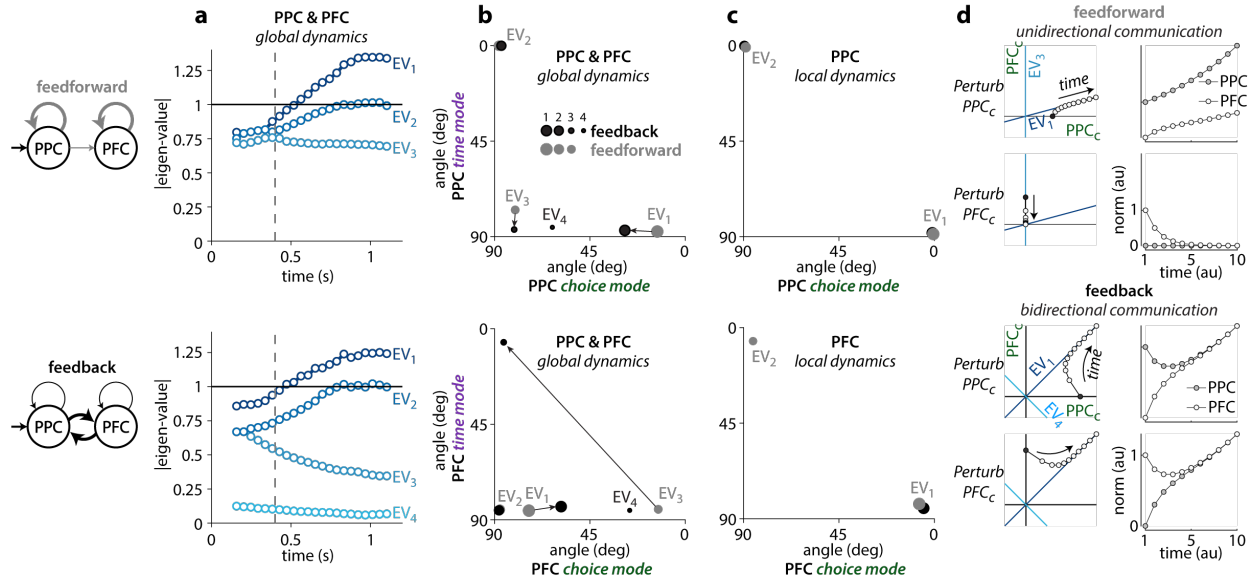
### Global residual dynamics resolves local and global recurrent computations

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

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

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

410 Global residual dynamics can distinguish between the two networks because variability  
411 evolves differently within and across areas depending on the connectivity between areas. To  
412 explore the possible nature of these differences, we first consider the effect of perturbations  
413 in two simple models implementing time-independent, linear dynamics (Fig. 7d), which  
414 mimics key properties of the inferred global dynamics (Fig. 7b). We considered activity that  
415 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 (95<sup>th</sup> 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.

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

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

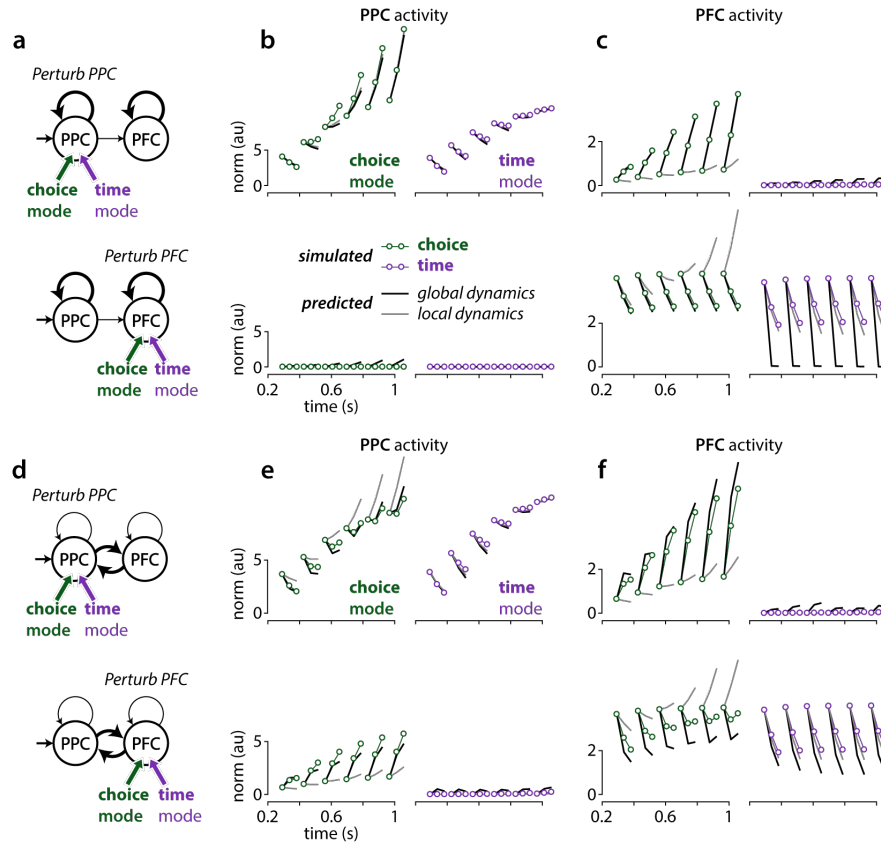
426 These simple models exemplify how the arrangement of global eigenvectors determines the  
427 directionality of the communication between areas. In the feedforward model, a PPC  
428 perturbation causes expanding activity in PPC that propagates to PFC, whereas a PFC  
429 perturbation decays in PFC,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

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

### **Inferring global dynamics with local causal perturbations**

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

453 For perturbations applied late in the trial, when dynamics is unstable (Fig. 7a,  $EV > 1$ ),  
454 perturbations of the choice modes result in activity that largely matches the dynamics of the  
455 simple models above (Fig. 7d). In the feedforward network, PPC perturbations lead to  
456 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).

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

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

## Discussion

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

488 The properties of global residual dynamics, based on recordings distributed across a  
489 network of inter-connected areas, can potentially resolve contributions of local, within-area  
490 recurrence and long-range, between-area connections (Fig. 7). The resulting description of  
491 dynamics in terms of modes (i.e., eigenvectors) that are shared across areas<sup>98</sup>, or private to  
492 a single area, appears plausible based on the past identification of communication- and null-  
493 subspaces between areas<sup>20,102,103</sup>—an eigenvector that is shared between two areas lies  
494 within their communication subspace, whereas one that is private lies outside of it, and  
495 potentially within the null-space of either area. Global residual dynamics, however, goes  
496 beyond a static description based on such subspaces, as it can capture also the dynamics of  
497 the responses (Fig. 8) resulting from unidirectional or bidirectional communication between

498 areas (Fig. 7d, top vs. bottom). Local residual dynamics in a single area, of the kind we  
499 describe for PFC, instead cannot readily distinguish between local and global contributions  
500 to observed neural responses. Any recurrent dynamics unfolding within the communication  
501 subspace of two areas will be reflected in the local dynamics of both areas, irrespective of  
502 the directionality of the communication between the areas (feedforward or feedback, Fig.  
503 6c,f).

504 Nonetheless, even our local estimates of PFC residual dynamics provide constraints on the  
505 properties of recurrent dynamics in PFC, and on the nature of the computations underlying  
506 decision-making and movement generation. For one, the largest estimated time constants  
507 provide an upper bound on the time-constants of the local recurrent dynamics in PFC (Fig.  
508 4e; 324ms and 510ms in monkeys T and V; medians), as any upstream contribution to PFC  
509 responses would have inflated these estimates (Fig. 6b; Extended Data Fig. 10,11). Recurrent  
510 dynamics in PFC is thus slow<sup>98,104</sup>, but stable throughout the decision and movement epochs.

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

528 Around the time of the saccade, PFC residual dynamics is quickly decaying, largely non-  
529 rotational, and only weakly non-normal, implying that PFC does not implement  
530 rotational<sup>45,46</sup>, funnel<sup>47,61</sup>, or strongly non-normal<sup>82,85</sup> recurrent dynamics of the kind  
531 previously proposed to explain movement activity in cortex. Rotational and funnel dynamics  
532 are also unlikely to be implemented in an upstream area driving PFC movement responses  
533 through a communication subspace, since the signatures of those dynamics would then also

534 appear in PFC residuals (Fig 6, Extended Data Fig. 10). Strong non-normal dynamics in an  
535 upstream area, however, could explain the residual dynamics and condition-averages  
536 observed in PFC. Non-normal systems can generate large activity transients along directions  
537 with only a small projection onto the activity subspace containing the slowest dynamics<sup>97</sup>. If  
538 the output from such an upstream area was partially aligned with the activity transients, but  
539 orthogonal to the slow dynamics, it could drive strong “input-driven” movement-related  
540 activity in PFC without revealing the signatures of the strongly non-normal dynamics that  
541 created it. Notably, the apparent absence of rotations in PFC recurrent dynamics during  
542 saccades does not rule out that such dynamics occurs in premotor and motor areas involved  
543 in hand-reaches<sup>43,45</sup>. The neural mechanisms underlying saccades and reaches may well be  
544 distinct, considering the substantial differences in the anatomy of the involved  
545 structures<sup>88,89</sup>.

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

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

## Author Contributions

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

## Acknowledgements

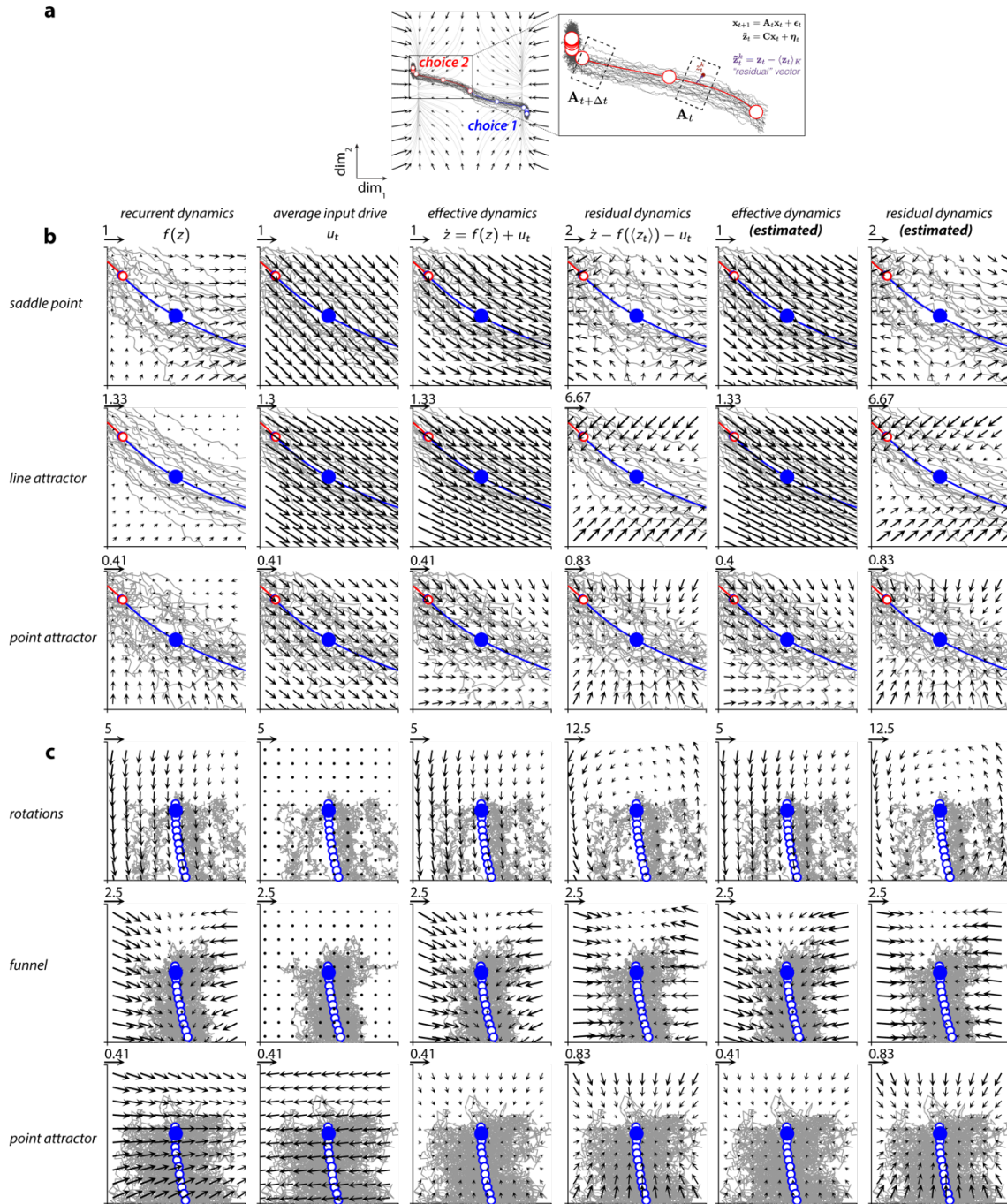
570 We thank John Reppas and William Newsome for the data collection. We thank Kevan Martin  
571 and all members of the Mante Lab for their valuable feedback, as well as Nicolas Meirhaeghe,  
572 Lea Duncker and Mehrdad Jazayeri for discussions and comments on the manuscript.

## Funding

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



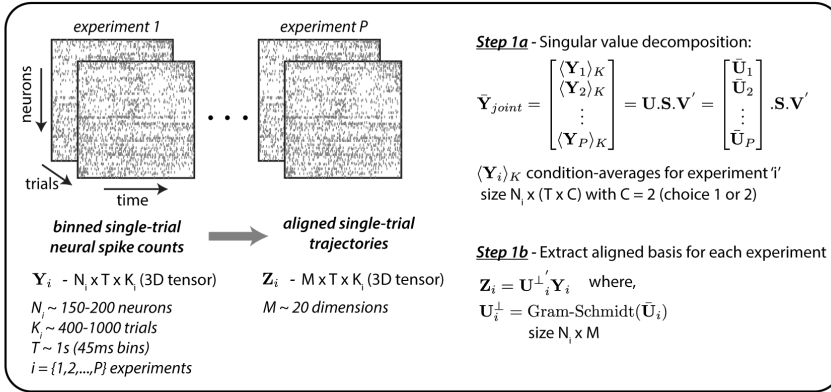
## Extended Data Figures



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

579 **a**, Variability in responses across trials from the same task condition are interpreted as perturbations away from the condition-  
 580 averaged trajectory. The evolution of these perturbations reflects the properties of the underlying recurrent dynamics (flow  
 581 field, same conventions as in Fig 1a). Inset on right shows a magnified view of the condition-averaged trajectory (red, choice  
 582 2) and corresponding single trials (dark gray) simulated from the saddle point model. Residual vectors at each time (shown in  
 583 purple for a single trial and time) are computed by subtracting the condition-averaged response at that time from the  
 584 corresponding single-trial response (purple equation). Time-varying dynamics matrices ( $A_t$ ) of a linear time-varying,

585 *autonomous state-space model (black equations, top-right) are fit to the residuals. These matrices approximate the dynamics*  
586 *in distinct 'local' regions of state space (e.g. dashed boxes) and are indexed according to time and condition. **b-c**, Components*  
587 *of the dynamics for the models of decisions (**b**) and movement (**c**) for an example reference time (blue dot) along the condition-*  
588 *averaged trajectory for choice 1. Same conventions as in Fig 2a. Dynamics are shown for a local state-space region close to the*  
589 *corresponding initial condition (boxes in Fig. 1a, b; left). For all models, the estimated effective and residual dynamics (columns*  
590 *5 and 6) closely match the true effective and residual dynamics (columns 3 and 4). In these models, the residual dynamics*  
591 *(column 4) reflects only the recurrent dynamics (column 1), but is not identical to it. For one, the fixed point of the residual*  
592 *dynamics by definition is located at the location of the reference state (the blue dot), which in general does not match the*  
593 *position of fixed points of the recurrent dynamics (e.g. the red circle in the first row and first column, corresponding to the*  
594 *position of the unstable fixed point in the saddle point model). The position of fixed points of the recurrent dynamics can only*  
595 *be inferred if the external inputs are known, a requirement that is not fulfilled in many experimental settings. For another,*  
596 *consistent drifts resulting from the recurrent dynamics (e.g. the drift along the limit cycle in the funnel model) are not reflected*  
597 *in the residual dynamics. Such drifts are "subtracted" from the variability in the computation of residuals.*



### Step 1: Session alignment

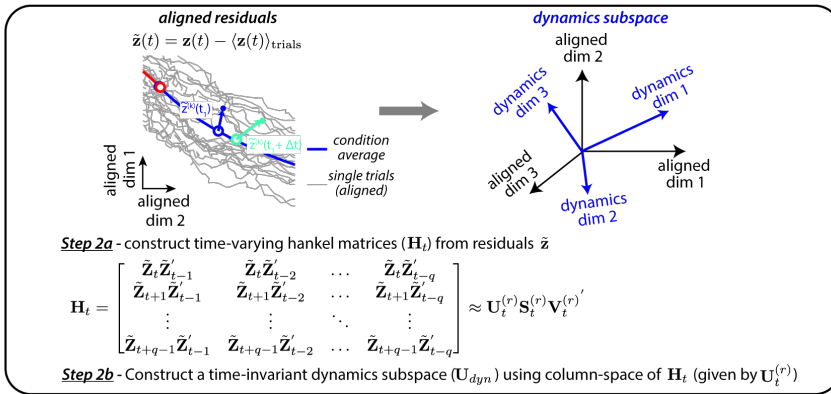
**Goal:** identify common, aligned modes across different experiments.

**Approach:** singular value decomposition and Gram-Schmidt orthogonalization

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

**Output:** "aligned" single-trial trajectories

**Hyperparameters:** # of aligned modes (M)



### Step 2: Dynamics subspace estimation

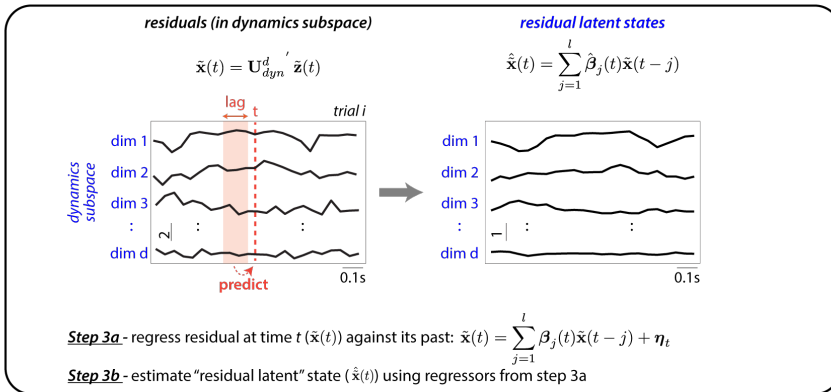
**Goal:** identify dimensions that are predictive of the "future" based on the "past".

**Approach:** Hankel matrix decomposition

**Input:** "aligned" residual responses

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

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



### Step 3: Residual latent state estimation

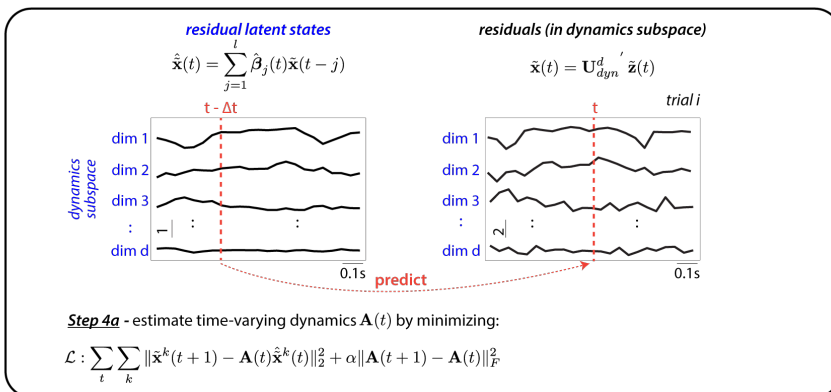
**Goal:** estimate the residual latent state

**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"

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



### Step 4: Time-varying dynamics estimation

**Goal:** estimate time-varying residual dynamics

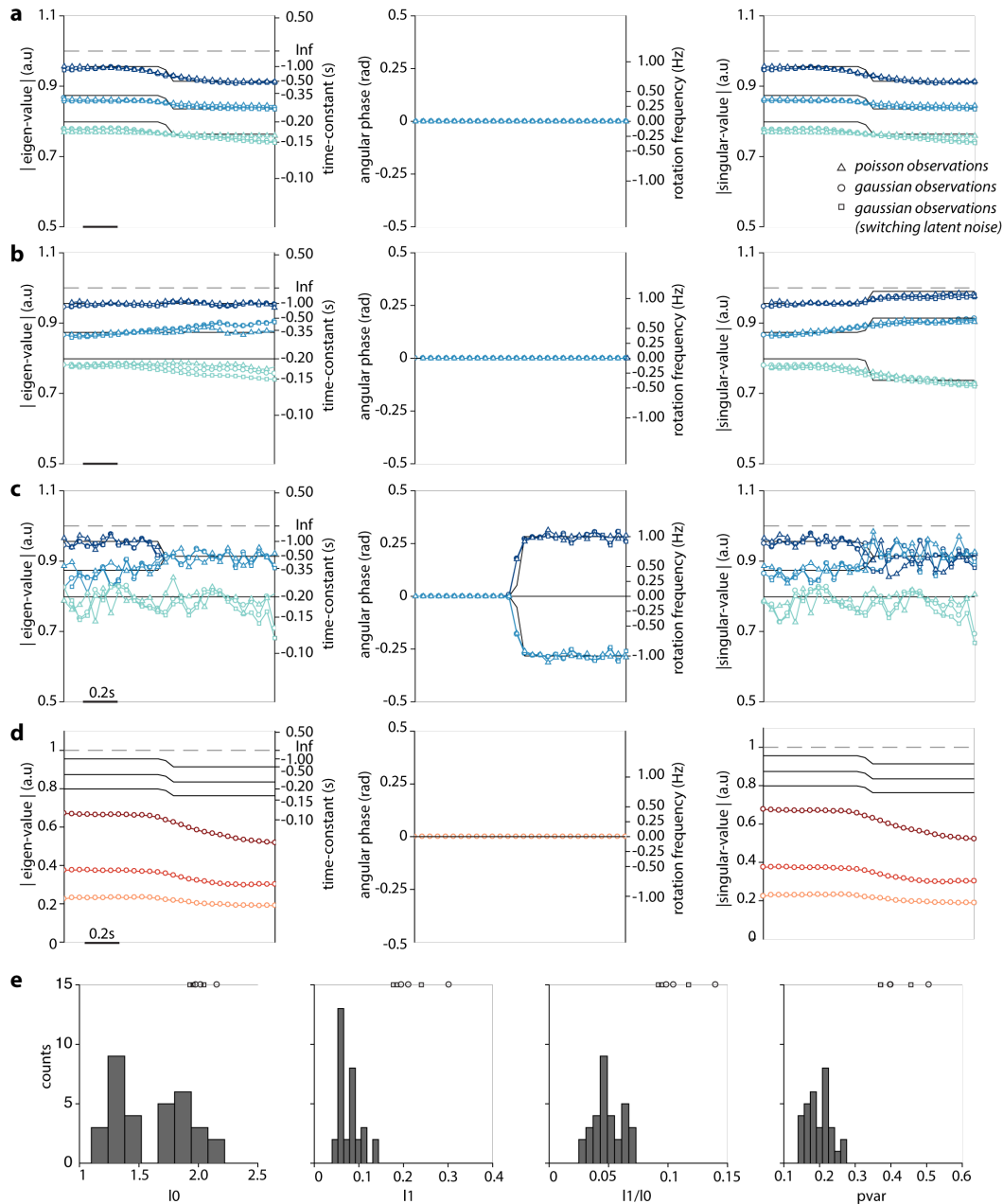
**Approach:** Second stage of a two-stage least squares (2SLS) regression (with regularization)

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

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

**Hyperparameters:** smoothness penalty  $\alpha$  (ascertained using cross-validation)

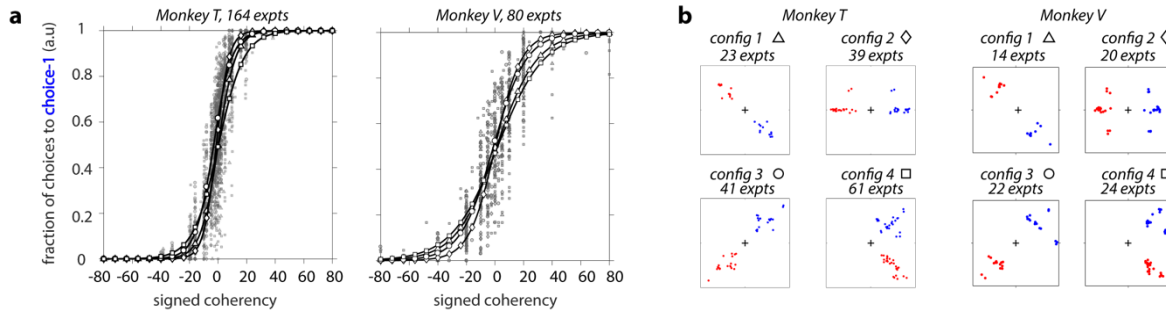
599 *Schematic depicting the complete data analysis pipeline for inferring residual dynamics from noisy neural population*  
600 *recordings. The pipeline involves four key sequential steps. Step 1: session alignment; involves pooling single trials from*  
601 *different recording sessions in order to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



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

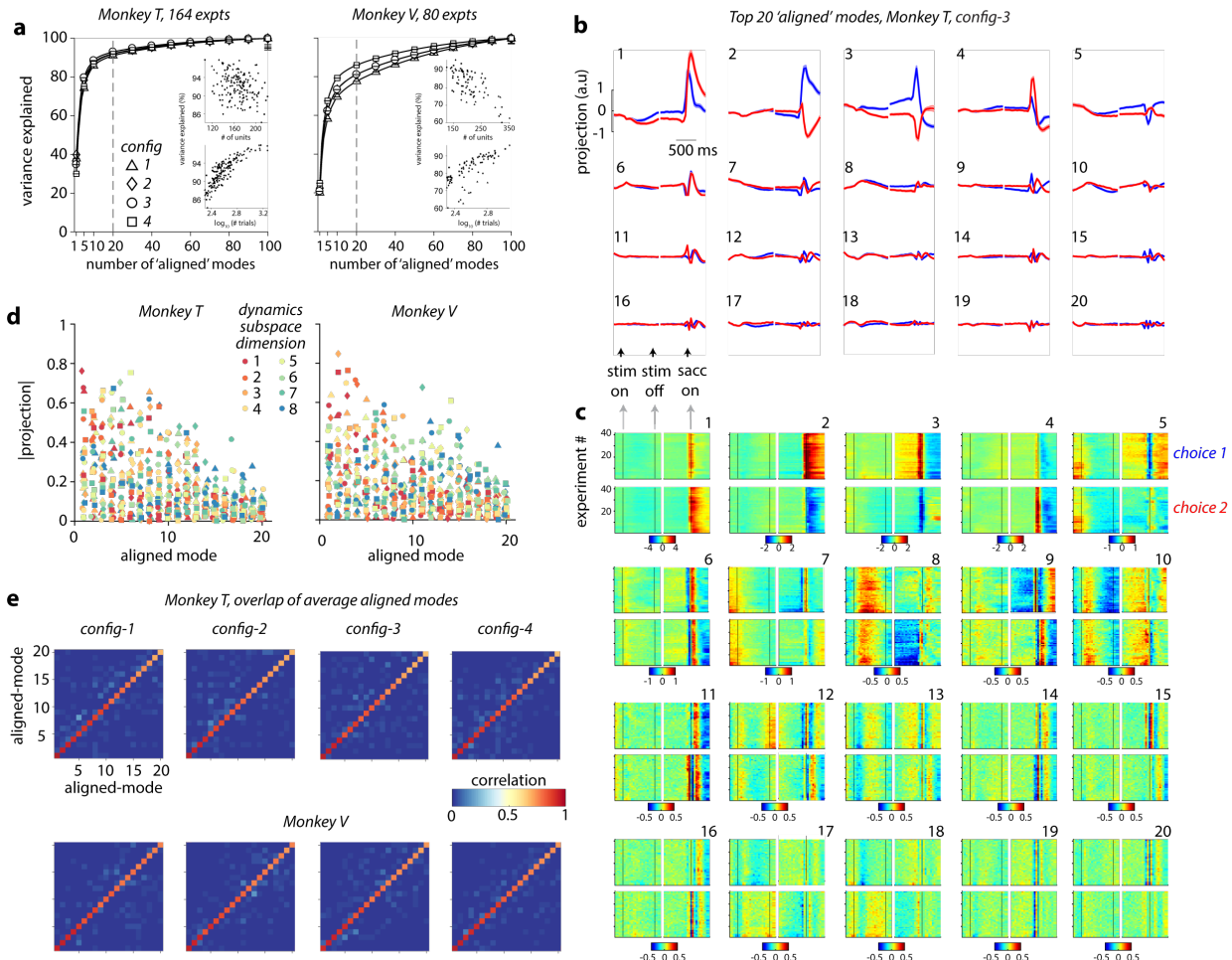
608 **a-c**, Validation of the estimation procedure on simulations of time-varying, linear dynamical systems. Simulations are based  
 609 on a latent variable dynamical system with 3 latent dimensions and 20 observed dimensions. Residual responses are generated  
 610 using a gaussian (circle markers: fixed latent noise variance; square markers: latent noise variance switches mid-way through  
 611 the trial) or poisson (triangle markers) observation process. In all simulations, the properties of the dynamics switch mid-way  
 612 through the simulated time window, from slowly decaying to quickly decaying (**a**); from non-rotational to rotational (**b**); or  
 613 from normal to non-normal (**c**). As in Fig. 4b-d, we characterize dynamics with the magnitude of the eigenvalues (left), the  
 614 rotational frequency (middle), and the singular values (right). Markers correspond to the estimated residual dynamics, black  
 615 curves to the ground-truth values. The estimated residual dynamics accurately matches the ground-truth for all types of  
 616 dynamics and observation models, before and after the switch, and also reveals the time of the switch. We observed this match  
 617 even when the latent noise variance of gaussian observations was switched at the same time as the eigenvalues/eigenvectors  
 618 of the dynamics (square markers), demonstrating that estimates of residual dynamics are robust to changes in latent noise

619 variance (see also Extended Data Fig. 11a-b vs e-f). **d**, Analogous to **c**, but for residual dynamics (circles) estimated using  
620 ordinary least squares (OLS) instead of two-stage least squares (2SLS) as in **c**. Results are only shown for data simulated using  
621 a gaussian observation process. Unlike the 2SLS estimates, the OLS estimates are strongly biased, i.e. the magnitude of the  
622 eigenvalues and the singular values are consistently underestimated. These biases are expected—they arise because both the  
623 regressors and the dependent variables are corrupted by observation noise (see Methods). The 2SLS instead produces unbiased  
624 estimates, as the first stage of 2SLS results in a denoising of the regressors (see also Extended Data Fig. 8). **e**, Parameters of the  
625 latent noise and observation noise for the simulations in **a-d** were chosen to approximately match the variability in the  
626 measured PFC responses. The variability in the measured responses were quantified in terms of four statistics ( $l_0$ ,  $l_1$ ,  $l_1/l_0$  and  
627  $pvar$ , x-axis; see Methods). Histograms indicate the respective values of these statistics in the neural data (one data point per  
628 task configuration, choice condition and monkey; see Extended Data Fig. 4). The open markers (top, same conventions as **a-c**)  
629 indicate the values of the statistics in the simulations for each of the three models.



630 **Extended Data Fig. 4: Behavioral task and configurations.**

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

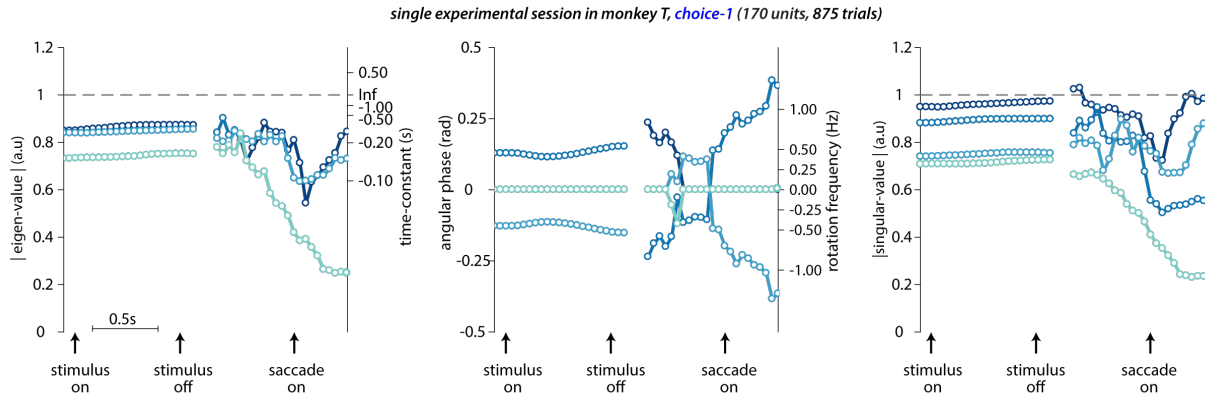


642 **Extended Data Fig. 5: Alignment of neural population responses from different experiments**

643 Validation of our session alignment procedure, Step 1 of the analysis pipeline (Extended Data Fig. 2). We aligned neural  
 644 population responses of all experiments belonging to the same task configuration and then pooled the aligned single trial  
 645 responses across experiments before computing the residuals used in estimating the dynamics. The outcome of the session  
 646 alignment procedure is a set of 20 'aligned' modes for each experiment, defined such that the activity of each mode has the  
 647 same dependency on time and choice across experiments. **a**, Cumulative variance explained in condition-averaged population  
 648 responses as a function of the number of aligned modes in both monkeys (left: Monkey T, right: Monkey V). We show the mean  
 649 across experiments of the cumulative variance explained for each task configuration (symbols as in Extended Data Fig. 4b).  
 650 Error-bars indicating twice the standard error of the mean are mostly smaller than the markers. The cumulative variance  
 651 explained by the first 20 aligned modes for all 164 experiments in Monkey T and 80 experiments in Monkey V showed a strong  
 652 positive trend with number of trials (inset, bottom) and a weak negative trend with the number of units (inset, top). **b**, Activity  
 653 of the first 20 aligned modes (numbered from top-left to bottom-right) for config-3 in monkey T (15,524 trials across 41  
 654 experiments) ordered according to the amount of variance explained. Activity is defined as the projection of the population  
 655 condition averages onto each mode. The projection was computed separately across experiments for choice 1 and choice 2  
 656 (blue and red) with responses aligned either to stimulus onset or saccade onset (black arrows). The resulting projections were  
 657 then averaged across experiments (shading: twice the standard error of the mean across experiments). **c**, Same data as in **b**  
 658 but showing the time-course of each aligned mode (numbered from 1 to 20) for each individual experiment (y-axis) separately  
 659 for the two choice conditions (choice 1 and choice 2, top and bottom sub-panels). Differences in the activation of a given mode  
 660 across experiments (i.e. across rows in each sub-panel) are much smaller than the differences in the activations across modes  
 661 (i.e. across sub-panels), demonstrating the success of the alignment procedure. **d**, Absolute value of the projection (y-axis) of



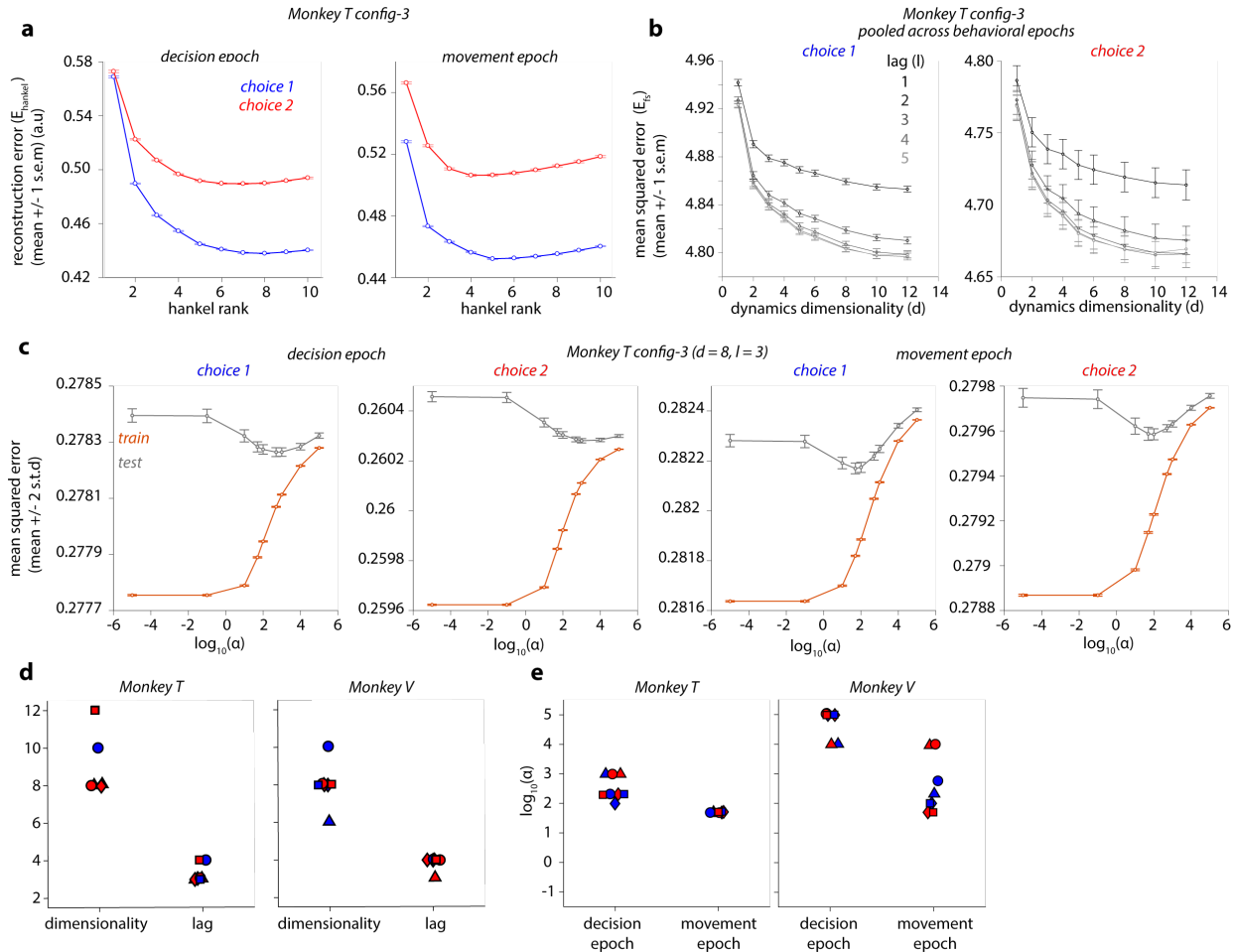
662 the 8 basis vectors (dim-1 through dim-8; red to blue) that span the dynamics subspace ( $\mathbf{U}_{dyn}$ , estimated in Step 2 of the  
663 analysis pipeline; Extended Data Fig. 2) onto the 20 aligned modes, indicating the relative alignment of the aligned and  
664 dynamics subspace. The dynamics subspace is computed separately for each task configuration (symbols as in Extended Data  
665 Fig. 4) in each monkey (left: Monkey T, right: Monkey V) and projects most strongly onto the first few aligned components (i.e.  
666 large projection values for smaller aligned mode number). The dynamics subspace thus largely overlaps with the subspace of  
667 activity that capture most of the task-related variance in the responses. **e**, Evaluation of the alignment procedure for all task  
668 configurations (columns) in both animals (rows). Each element of the matrix is obtained from the correlation coefficient  
669 between the time-courses of two aligned modes (i.e. positions along horizontal and vertical axes). We show the median  
670 correlation coefficient across all pairs of dissimilar experiments. Values close to 1 along the diagonal and close to 0 off-diagonal  
671 indicate that the time-courses are much more similar across experiments than across modes, indicating successful alignment.



672

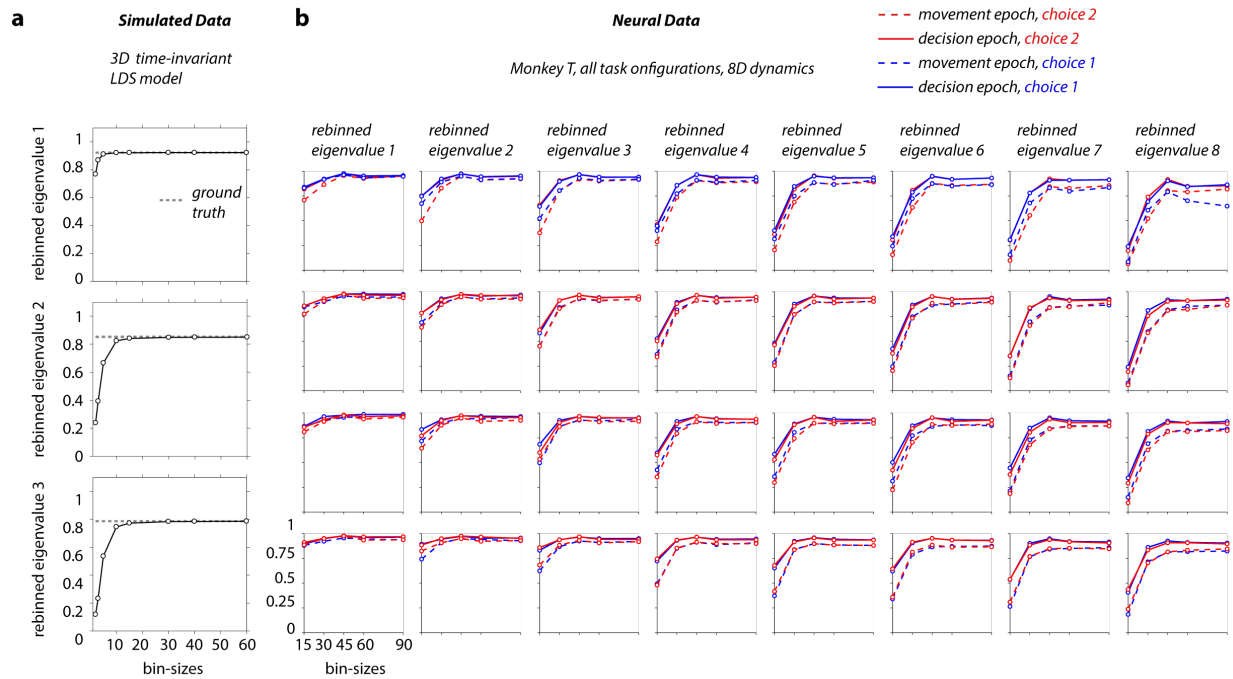
673 **Extended Data Fig. 6: Estimated residual dynamics for neural data from a single unaligned session**

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



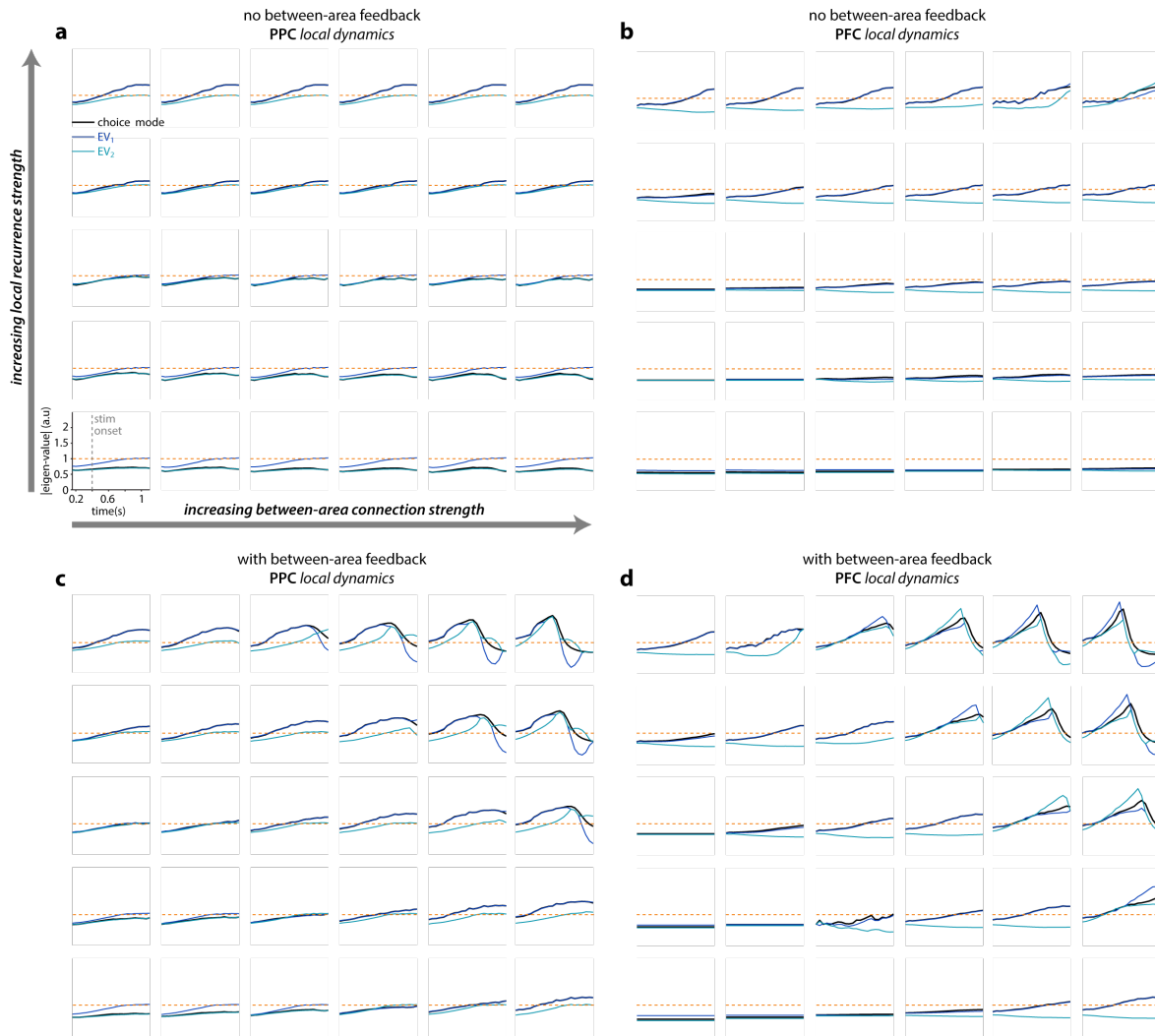
684 **Extended Data Fig. 7: Cross-validation of hyper-parameters used for estimating residual dynamics**

685 **a-c**, Representative results of the cross-validation procedure used to determine the various hyper-parameters of the analysis  
686 pipeline (see Extended Data Fig. 2) for neural data from a single task configuration in monkey T (config-3, see Extended Data  
687 Fig. 4). **a**, Cross-validated hankel matrix reconstruction error ( $E_{\text{hankel}}$ ; circle: mean over 20 repeats of hold-out cross validation;  
688 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  
689 the two epochs (left: decision; right: movement) and two choices (blue: choice 1; red: choice 2). The reconstruction error for  
690 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**,  
691 5-fold cross-validated mean squared error (circles: mean over 5 folds; error bars: 1 s.e.m) of the denoised residual predictions  
692 obtained from the first stage of the two-stage least squares regression (2SLS; step 3 in Extended Data Fig. 2), plotted as a  
693 function of the hyper-parameters:  $d$  (dimensionality of dynamics subspace); and  $l$  (number of past lags). For each cross-  
694 validation fold, a single mean squared error measure was computed by pooling the denoised predictions across time points in  
695 both epochs (left: choice 1; right: choice 2). **c**, 5-fold cross-validated mean squared error (circle: average across 5 'repeats' of  
696 the 5-fold cross validation; error bars: 2 standard deviations across repeats) of the residual predictions obtained from the  
697 second stage of the 2SLS regression (step 4 in Extended Data Fig. 2), plotted as a function of the smoothness hyper-  
698 parameter  $\alpha$  for different epochs (left: decision; right: movement) and choice (choice 1 and 2). Both the train (orange)  
699 error 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  
700 all task configurations and monkeys (symbols as in Extended Data Fig. 4b). A dimensionality of 8 and a lag of 3 was deemed  
701 optimal for both monkeys and task configurations (used in Fig. 4). **e**, Summary showing the optimal smoothness hyper-  
702 parameter  $\alpha$  (step 4 in Extended Data Fig. 2) for all task configurations and monkeys. Final values of  $\alpha$  were chosen to be the  
703 same across monkeys in Fig. 4 (decision epoch:  $\alpha = 200$ ; movement epoch:  $\alpha = 50$ ) despite a small degree of variability across  
704 the two monkeys. Same conventions as in **d**.



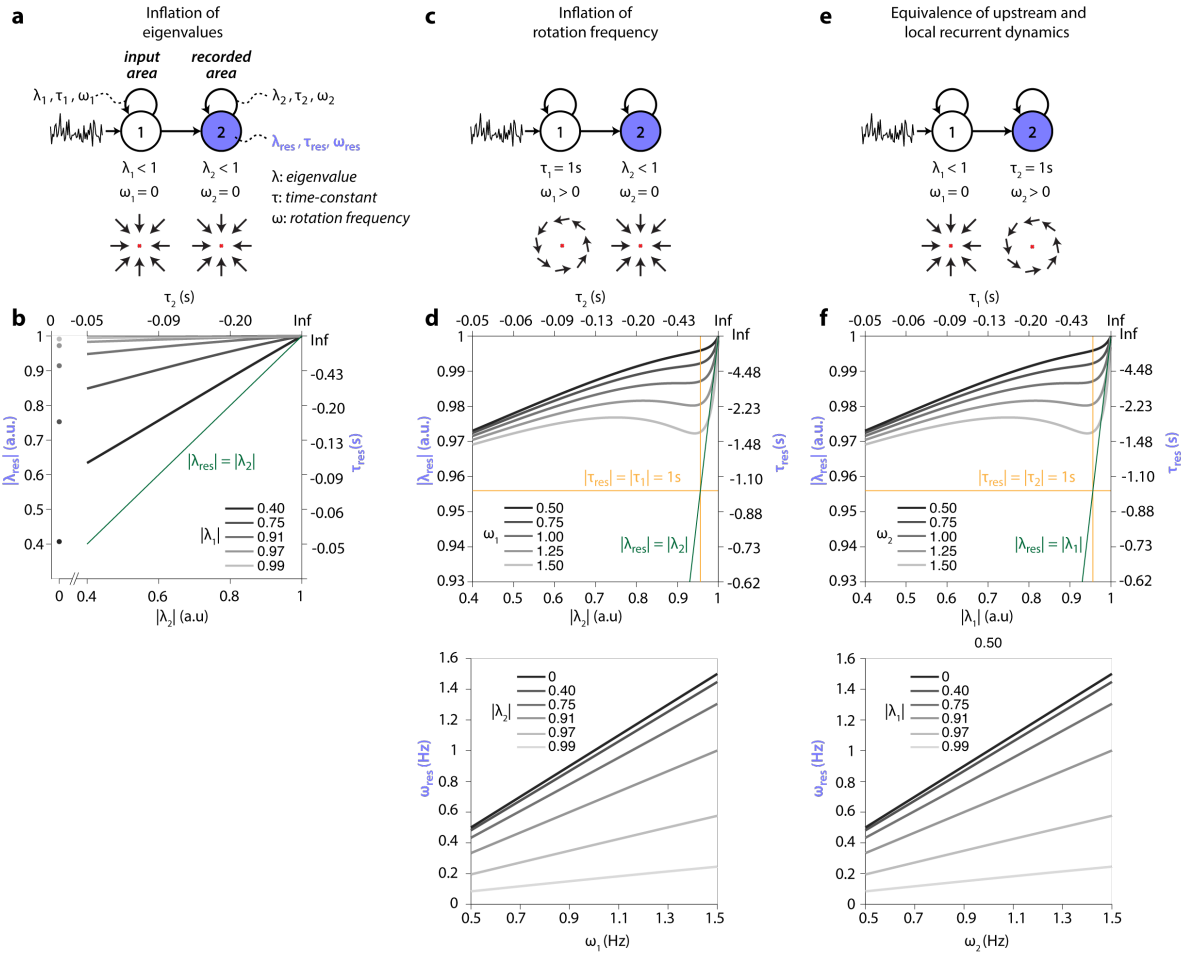
705 **Extended Data Fig. 8: Effect of bin size on the estimated residual dynamics**

706 We estimated the residual dynamics for different choices of bin size, to identify the smallest bin size resulting in unbiased  
 707 estimates. In the discrete time formulation of a linear dynamical system, like the one we use here, re-binning of the responses  
 708 trivially results in a scaling of the estimated eigenvalues of the residual dynamics. To compensate for this rescaling, here we  
 709 “mapped” the estimated eigenvalues onto a common, reference bin size (see Methods, Effects of bin size). In the absence of  
 710 statistical biases, the resulting “re-binned eigenvalue” would be independent of bin size. **a**, Re-binned eigenvalues for  
 711 simulations of a time-invariant, latent-variable (3 latent dimensions), LDS model (reference bin size = 40ms) as a function of  
 712 bin-size (dashed line: ground truth; black curve: estimate). Estimates of the residual dynamics are biased for small bin sizes,  
 713 but become unbiased when bin size is sufficiently large. **b**, Estimated re-binned eigenvalues (reference bin size = 15ms) as a  
 714 function of bin size for all configurations in monkey T. Columns correspond to the 8 distinct eigenmodes of the estimated 8-  
 715 dimensional residual dynamics (left to right, largest to smallest EV), rows correspond to task configurations (top to bottom,  
 716 config-1 to 4; Extended Data Fig. 4b). Here the re-binned eigenvalues were computed separately for each choice (red vs blue)  
 717 and averaged in small temporal windows specific to each epoch: 0.2-0.4s relative to stimulus onset (solid lines) and -0.15 to  
 718 0.25s relative to saccade onset (dashed lines). All main analyses of recorded neural responses are based on a bin size of 45ms,  
 719 for which estimates of residual dynamics appear to converge to an asymptote. Note that the re-binned eigenvalues for a bin  
 720 size of 45ms are larger than the corresponding eigenvalues reported in other figures (e.g. Fig. 4b), because the former were  
 721 mapped onto a reference bin size of 15ms.



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

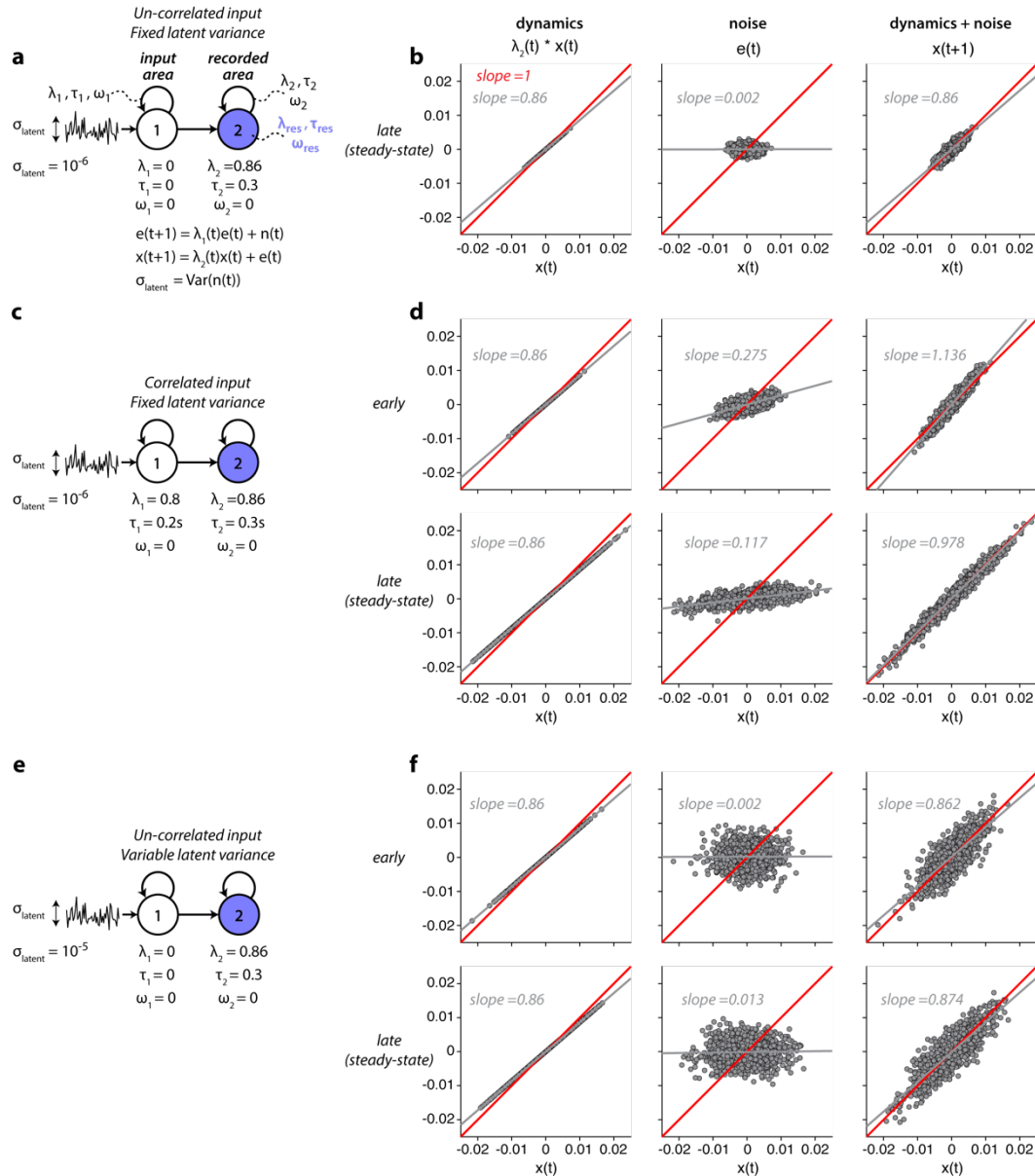
723 We estimated local residual dynamics separately in PPC or PFC of the two-area network models, as in Fig. 6 (PPC: **a,c**; PFC:  
 724 **b,d**). We estimated either 2-dimensional dynamics based on the residuals in a given area, resulting in two time-varying  
 725 eigenvalues (cyan and blue curves, analogous to Fig. 6b,e); or 1-dimensional dynamics based on the projection of the residuals  
 726 onto the choice mode in each area (Fig. 6a,d), resulting in a single time-varying eigenvalue (black curves). The maximum value  
 727 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  
 728 from PFC to PPC (Fig. 6c, left), for increasing strength of local recurrence (bottom to top) and between-area connectivity (left  
 729 to right). **b**, Local residual dynamics in PFC, same networks as in **a** (Fig. 6c, right). **c**, Local residual dynamics in PPC for the  
 730 networks with feedback from PFC to PPC (Fig. 6f, left), same conventions as in **a**. **d**, Local residual dynamics in PFC, same  
 731 networks as in **c** (Fig. 6f, right). For networks with strong local recurrence and strong between-area connections, the neural  
 732 population activity falls into one of two point-attractors before the end of the trial, resulting in a drop in the eigenvalues. The  
 733 two point attractors implement the commitment to a choice by the network.



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

735 We systematically explored the effect of correlated input variability on estimates of residual dynamics in a two-area, linear  
 736 dynamical system. The input area implements 2d isotropic recurrent dynamics characterized by parameters  $\lambda_1$ ,  $\tau_1$ , and  $\omega_1$   
 737 (eigenvalue, time-constant, rotation frequency). Activity in the input area is externally driven by uncorrelated noise. Values of  
 738  $\lambda_1$  closer to 1 result in longer auto-correlation times in the variability of activity in the input area. This activity provides the  
 739 input into the recorded area, which implements 2d isotropic recurrent dynamics with parameters  $\lambda_2$ ,  $\tau_2$ ,  $\omega_2$ . Residual dynamics  
 740 at steady-state is estimated from activity of the recorded area. At steady state, estimates can be derived analytically (see  
 741 Supplementary Math Note B). Because of temporally correlated input variability, the properties of the residual dynamics ( $\lambda_{res}$ ,  
 742  $\tau_{res}$ ,  $\omega_{res}$ ) in general do not match those of the recurrent dynamics in the recorded area. **a-b**, Inflation of eigenvalues. **a**,  
 743 Schematic of the model (top) and recurrent dynamics in each area (bottom, flow fields). Recurrent dynamics is stable and non-  
 744 rotational in both areas. **b**, Residual dynamics ( $\lambda_{res}$ ) in the recorded area as a function of recurrent dynamics in the recorded  
 745 area ( $\lambda_2$ , x-axis) and in the input area ( $\lambda_1$ , gray lines). The eigenvalues of the residual dynamics are inflated, i.e.  $\lambda_{res}$  is larger  
 746 than  $\lambda_2$  for (all gray lines above the green line). Larger  $\lambda_1$  (longer input auto-correlations) lead to stronger inflation. For  $\lambda_2 =$   
 747 0 (no recurrent dynamics in the recorded area)  $\lambda_{res} = \lambda_1$  (gray circles). **c-d**, Inflation of rotation frequency. **c**, Recurrent  
 748 dynamics is rotational in the input area, but stable and non-rotational in the recorded area. **d**, Residual dynamics in the  
 749 recorded area, expressed as the magnitude of the eigenvalue ( $\lambda_{res}$ , top) and the rotation frequency ( $\omega_{res}$ , bottom). The  
 750 eigenvalues of the residual dynamics are generally inflated (top), but the relation with  $\lambda_2$  is non-monotonic and depends on  
 751  $\omega_1$ . The residual dynamics is rotational (bottom,  $\omega_{res} > 0$ ) even though the recurrent dynamics in the recorded area is not ( $\omega_2 =$   
 752 0). The inflation of rotation frequency is reduced for increasing  $\lambda_2$ . **e-f**, Equivalence of upstream and local recurrent dynamics.  
 753 **e**, Analogous to **c**, but dynamics is switched between input and recorded area. **f**, Analogous to **d**, but for the dynamics in **e**. The

754 *residual dynamics is identical to that in **d**. In general, residual dynamics in the recorded area reflects the combined effect of*  
755 *local and upstream recurrent dynamics.*

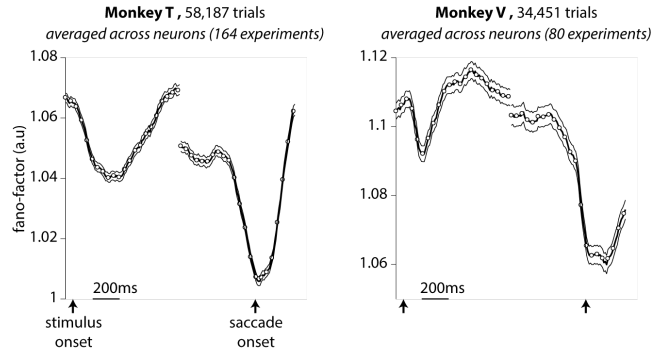


756 **Extended Data Fig. 11: Explanation of input driven inflation in residual dynamics**

757 To gain an intuitive understanding of inflation of eigenvalue magnitude, we consider simulations of two-area linear dynamical  
 758 systems similar to those in Extended Data Fig. 10a. For simplicity, here we simulate stable 1d-dynamics in each area, whereby  
 759 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  
 760 time-dependent latent noise variance (e-f). The variability injected into the input area is always uncorrelated. Recurrent  
 761 dynamics in the recorded area is identical in all simulations. **a**, Model parameters for the case of uncorrelated input ( $\lambda_1 = 0$ ). **b**,  
 762 Contributions to activity  $x$  in the recorded area at steady-state. Activity  $x(t)$  ( $x$ -axis) is propagated through the recurrent  
 763 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  
 764  $e(t)$  corresponds to activity/output of the input area, and is shaped by dynamics determined by  $\lambda_1$ . Points in the scatter plots  
 765 correspond to different simulated trials. Estimating the eigenvalue of the residual dynamics in the absence of observation noise  
 766 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  
 767 to that obtained if the latent noise had not been added to the activity (left, gray line), meaning that residual dynamics correctly  
 768 reflects the effect of the recurrent dynamics in the recorded area (slope < 0, reflecting  $\lambda_2 < 0$ ; left). **c**, Model parameters for the



769 *case of correlated input ( $\lambda_1 > 0$  for  $t > 0$ ;  $\lambda_1 = 0$  at other times). **d**, Analogous to **b**, but for the model in **c**. Here activity and noise  
770 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  
771 times, residual dynamics is inflated, i.e. the regression slope between  $x(t)$  and  $x(t+1)$  (right) is larger than that obtained by  
772 applying only the recurrent dynamics (left), indicating inflation of the eigenvalues. Inflation occurs because the noise itself is  
773 correlated with activity in the recorded area (middle, slope  $> 0$ ), an effect that results indirectly from the correlation between  
774  $e(t)$  and  $e(t-1)$ . At steady state, even the inflated residual dynamics is still stable (bottom-right, slope  $< 1$ ; see also Extended  
775 Data 10b). However, immediately after the onset of the correlated input, residual dynamics erroneously reveals an instability  
776 (top-right, slope  $> 1$ ). **e**, Parameters for the case of uncorrelated noise but time-varying noise variance. The variance of the  
777 noise injected into the input area is increased at time  $t = 0$ , from  $\sigma_{\text{latent}} = 10^{-6}$  to  $10^{-5}$ . **f**, A change in noise variance does not  
778 result in inflation of the residual dynamics, neither early nor late after the change (right, top and bottom; same slope as on the  
779 left; see also Extended Data Fig. 3a-c, squares).*



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

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

## References:

1. Steinmetz, N. A., Zatzka-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).