

Supplementary Information

Interactions across hemispheres in prefrontal cortex reflect global cognitive processing

Megan E. McDonnell^{1,2,†}, Akash Umakantha^{1,2,3,†}, Ryan C. Williamson^{1,2,3,4,†}
Matthew A. Smith^{1,2,5,‡,*} Byron M. Yu^{1,2,5,6,‡,*}

¹Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA.

²Center for the Neural Basis of Cognition, Carnegie Mellon University & University
of Pittsburgh, Pittsburgh, PA, USA.

³Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA.

⁴School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA.

⁵Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh,
PA, USA.

⁶Department of Electrical and Computer Engineering, Carnegie Mellon University,
Pittsburgh, PA, USA.

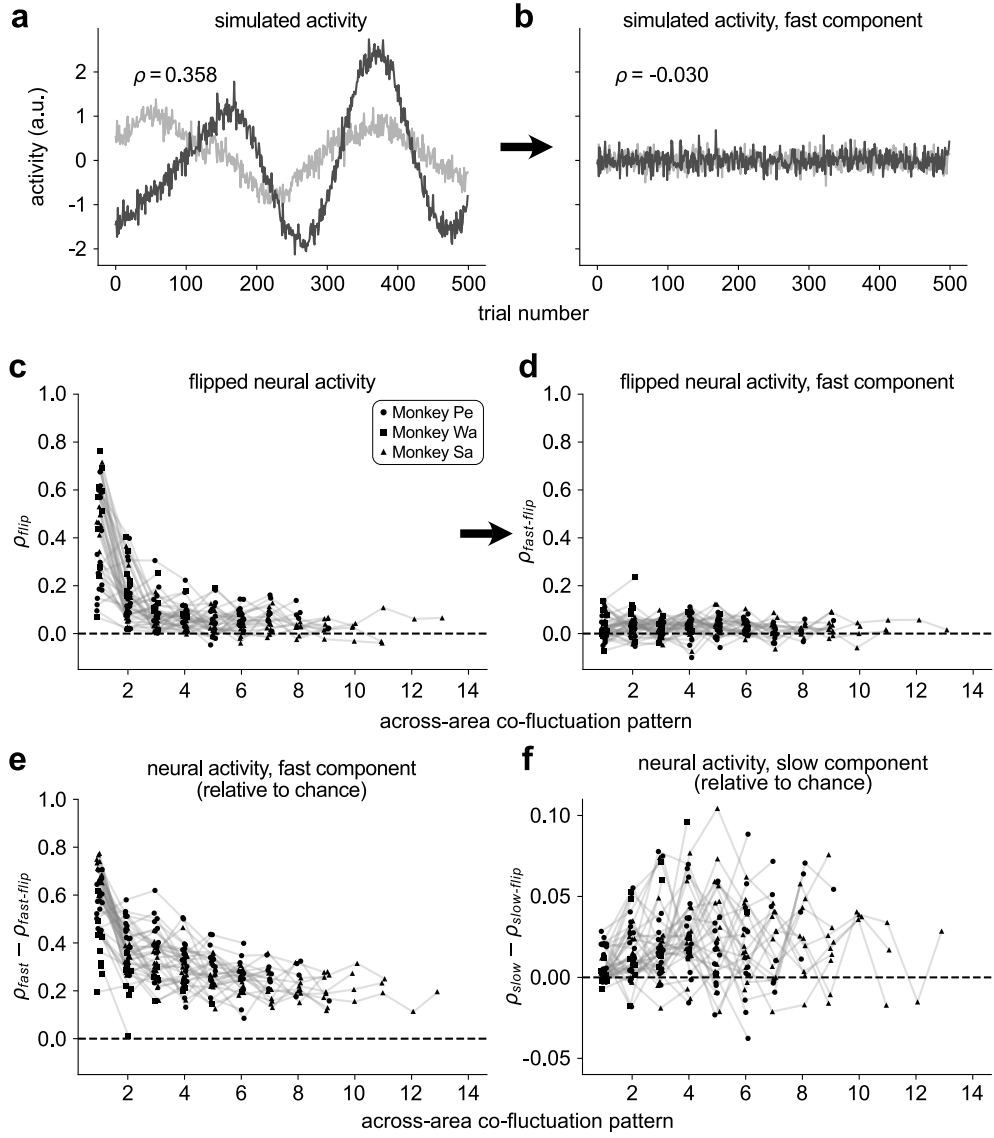
† These authors contributed equally.

‡ These authors jointly supervised this work.

*Corresponding author(s). E-mail(s): mattsmith@cmu.edu, byronyu@cmu.edu

Contents

1. Supplementary Fig. 1 with legend
2. Supplementary Fig. 2 with legend
3. Supplementary Fig. 3 with legend
4. Supplementary Fig. 4 with legend
5. Supplementary Fig. 5 with legend
6. Supplementary Fig. 6 with legend
7. Supplementary Fig. 7 with legend
8. Supplementary Fig. 8 with legend
9. Supplementary Fig. 9 with legend
10. Supplementary Fig. 10 with legend
11. Supplementary Information References



Supplementary Fig. 1: Spurious correlations induced by slow-timescale fluctuations.

In this work, we were interested in characterizing trial-to-trial variability shared among neurons. However, neural activity can exhibit multiple timescales of variability that can influence interpretation. For example, fluctuations on the timescale of minutes to hours have been observed in many brain areas, including PFC [1]. It is important to control for these slow timescales in correlation analyses such as spike count correlation

(r_{sc}), canonical correlation analysis (CCA), and our method, pCCA-FA (a combination of probabilistic canonical correlation analysis and factor analysis), to prevent the measurement of potentially spurious correlations [2, 3].

a, To illustrate how slow-timescale activity fluctuations can yield spurious correlations, consider a simulation where activity from two brain areas is generated as two independently drawn Gaussian Processes (GPs, [4]). Each GP consists of a time series of 500 values, representing the activity of a single brain area over 500 trials. We generated two GPs representing two brain areas with corresponding trial indices, each including a slow-timescale fluctuation with time constant $\tau = 50$ trials. Since the GPs were independently drawn, we would expect them to have a correlation across trials that is close to zero. However, given a limited number of trials, two independent GPs with slow-timescale fluctuations can have substantial non-zero correlation. Here, the two time series have a trial-to-trial correlation of $\rho = 0.358$.

b, As the number of trials grows to infinity, the correlation between the GPs would shrink to zero. However, this trial count inflation is not feasible in animal experiments. Instead, we can remove the slow-timescale components by subtracting a moving average (i.e., retain only the fast component; see Methods). After this subtraction, the correlation between the two GPs is now close to zero ($\rho = -0.030$). Thus, whenever fluctuations in the activity occur on timescales that extend across a substantial portion of the trials, spurious correlations can arise.

c, The same phenomenon could be seen in our neural recordings. We ran a control analysis to demonstrate the importance of removing slow components before performing any correlation analysis in neural activity. To generate a chance distribution, we performed a flip control where the trial order of neural activity in one area (right PFC) was flipped. Our assumption was that this flipping would break any trial-to-trial correspondence and should in principle result in correlations close to zero. Any recovered correlations would be spurious (induced by slow-timescale fluctuations).

For each experimental session, we fit a pCCA-FA model to the flipped neural activity (left PFC remained the same, right PFC was flipped). We selected the number of across- and within-area latent variables of each model to be the same as what was identified in the non-flipped neural activity (i.e., in the models fit for analyses in Figs. 4-6). We then computed the cross-validated canonical correlations by splitting the trials of flipped neural activity into training and testing folds. For each fold, we estimated pCCA-FA model parameters using the training trials. Then, we computed the held-out posterior means ($\mathbb{E}[\mathbf{z}|\mathbf{x}_1]$ and $\mathbb{E}[\mathbf{z}|\mathbf{x}_2] \in \mathbb{R}^d$, see Methods and Equation (7)) using the testing trials. Across folds, this yielded one posterior mean per brain area per trial. The cross-validated canonical correlation is the trial-by-trial correlation of the posterior means. Note that this is equivalent to canonical correlations that would be identified by CCA, when applied to the same data.

When fitting pCCA-FA to the flipped neural activity, we recovered large cross-validated canonical correlations (ρ_{flip}), indicating that the across-area variance of models fit to neural activity that includes slow-timescale fluctuations could be representing spurious correlations rather than true across-area interactions. Each plotted point indicates the cross-validated canonical correlation (i.e., the correlation between held-out posterior means) for a pair of across-area co-fluctuation patterns (one in each

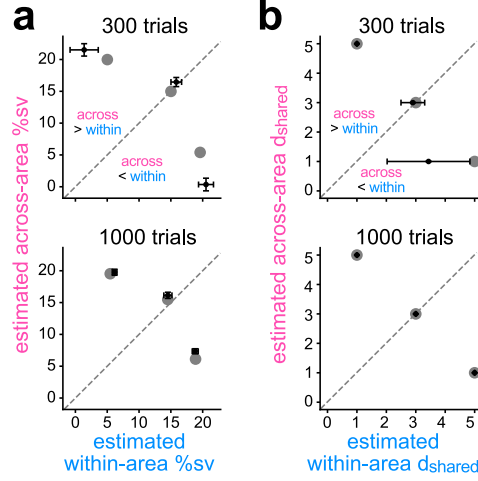
area) in a single session. The horizontal axis positions of each point are jittered slightly for visual clarity.

d, If the slow component were inducing spurious correlations, we would expect that removing this component from the flipped neural activity would reduce the correlations seen in panel **c**. To test this, we subtracted the slow component from the neural activity using a moving average (see Methods). When fit to the flipped fast component, pCCA-FA recovered cross-validated canonical correlations close to zero ($\rho_{\text{fast-flip}}$), indicating that spurious correlations were successfully removed from the neural activity. Therefore, the analyses in this work (r_{sc} in Fig. 2 and pCCA-FA in Figs. 4-6) focus on the fast component. We used the same model fitting procedure and computation of cross-validated canonical correlations as in panel **c**, but on the flipped fast component of neural activity. Each point indicates cross-validated canonical correlation for a pair of across-area co-fluctuation patterns as in panel **c**.

e, Throughout this work, we fit pCCA-FA to the fast component of neural activity. Here, we sought to assess whether the cross-validated canonical correlations (ρ_{fast}) in this case were meaningful. To do this, we defined a chance distribution using the canonical correlations recovered when fitting pCCA-FA to the flipped fast component ($\rho_{\text{fast-flip}}$, i.e., the values plotted in panel **d**). We found that the cross-validated canonical correlations in the (non-flipped) fast component of our PFC recordings were significantly above chance ($\rho_{\text{fast}} - \rho_{\text{fast-flip}} > 0$) in all sessions (Monkey Pe: $p = 1.1\text{e-}127$, $n = 103$ co-fluctuation patterns; Monkey Wa: $p = 1.0\text{e-}32$, $n = 28$; Monkey Sa: $p = 2.8\text{e-}156$, $n = 132$; paired one-tailed t-tests). This indicated that substantial trial-to-trial correlation existed between the hemispheres of PFC.

f, Though this work focuses on trial-to-trial variability, we also wondered whether any meaningful (i.e., not spurious) correlations existed in the slow component of neural activity. We fit pCCA-FA to the estimated slow component of neural activity. To determine whether the estimated canonical correlations (ρ_{slow}) were meaningful, we performed the same analysis as in panel **e**, but using the slow component instead of the fast component. Similar to the analysis in our previous work [1], we found here that there were substantial correlations present in the slowly varying neural activity. ρ_{slow} was above $\rho_{\text{slow-flip}}$ ($\rho_{\text{slow}} - \rho_{\text{slow-flip}} > 0$) across sessions for most across-area co-fluctuation patterns (Monkey Pe: $p = 1.0\text{e-}14$, $n = 103$ co-fluctuation patterns; Monkey Wa: $p = 0.0010$, $n = 28$; Monkey Sa: $p = 1.0\text{e-}17$, $n = 132$; paired one-tailed t-tests). This indicates that slow-timescale across-area interactions do indeed exist in PFC activity (perhaps similar to the correlations observed in our previous analysis of V4 and PFC [1]), in addition to the trial-to-trial effects that we focused on in this work.

* Source data are provided as a Source Data file.



Supplementary Fig. 2: pCCA-FA accurately recovered %sv and d_{shared} in simulations.

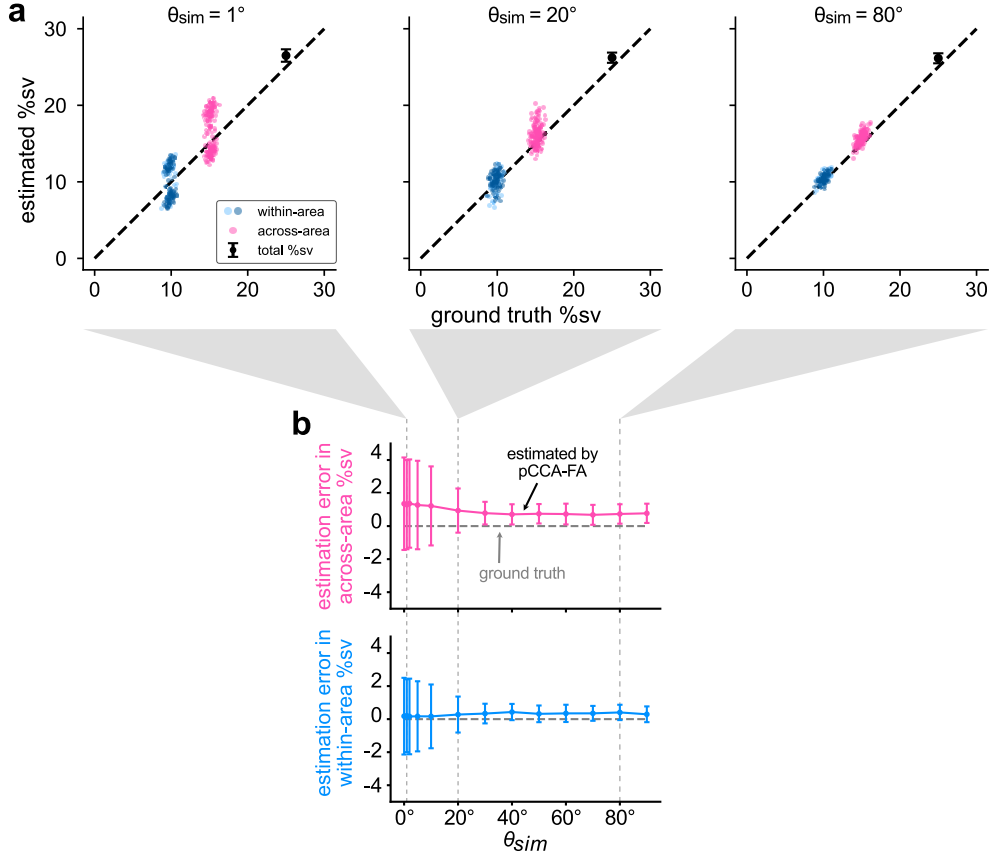
To validate the model fitting procedure, we asked whether pCCA-FA could recover ground truth values of %sv and d_{shared} using realistic numbers of simulated neurons and trials. We systematically varied the ground truth values of across- and within-area %sv (panel **a**) and d_{shared} (panel **b**) over values in the real-data regime. In each simulation, we used 30 neurons per area and generated random loading matrices W_m and L_m for each area m using independent draws from a standard normal distribution. To scale the loading matrices to enforce a particular level of %sv, we adjusted the singular values of the across- and within-area loading matrices (W_m and L_m) relative to each other to achieve the desired ratios. In this analysis, we used a flat distribution of singular values for each loading matrix. Although this is not physiologically realistic (see Supplementary Fig. 6), it was appropriate for this analysis to ensure that each co-fluctuation pattern described sufficient variance in cases of low %sv. The independent variance values (i.e., entries of Ψ_m) were then scaled to achieve a level of independent variance similar to what is estimated from our neural recordings (in this analysis, 75%). For each parameter configuration, we generated 30 independent datasets. We fit pCCA-FA to each dataset and estimated the number of across- and within-area latent variables in a cross-validated manner (see Methods). Plotted error bars indicate one standard deviation computed across the 30 datasets.

a, We tested three simulated %sv scenarios (marked with gray points), in which the across-area %sv was less than (below the dashed line), equal to (on the dashed line), or greater than (above the dashed line) the within-area %sv. This allowed us to assess whether pCCA-FA was able to accurately recover the across- and within-area components in scenarios in which one component dominated the other. Each simulated dataset included 5 across-area latent variables ($d = 5$) and 3 within-area latent variables per area ($d_m = 3$, for each area m). These values were chosen to be within the range observed in our neural recordings. To understand how trial count limitations impacted the estimation of %sv, we fit pCCA-FA to the three %sv scenarios

using either 300 (top) or 1,000 (bottom) trials. These trial counts represented the lower and upper end of the number of trials in our experimental sessions. With a small number of trials (top), when %sv was low pCCA-FA tended to underestimate the value of %sv (points to the left or below gray points indicating the ground truth configurations). With sufficient trial counts (bottom), the estimated values of %sv became accurate (estimates on top of gray points with low spread), regardless of the relative proportions of across- and within-area %sv. For visual clarity, we plotted the estimated %sv values from only one simulated brain area, and the other area yielded similar results.

b, We next asked if pCCA-FA was similarly able to identify the ground truth d_{shared} under scenarios in which across-area d_{shared} was less than (below the dashed line), equal to (on the dashed line), or larger than (above the dashed line) within-area d_{shared} . We manipulated the ground truth d_{shared} while keeping the %sv constant ($\approx 15\%$ across-area and $\approx 10\%$ within-area). When fitting pCCA-FA with 300 trials (top), we found that pCCA-FA underestimated within-area d_{shared} in some cases. These were cases in which each within-area co-fluctuation pattern explained only $\approx 2\%$ of the overall variance, making them difficult to identify with limited trials. Across-area d_{shared} was accurately estimated in all cases, perhaps because each across-area co-fluctuation pattern explained a greater amount of %sv than each within-area co-fluctuation pattern (because the ground truth %sv is higher). With increased trial counts (1,000, bottom), pCCA-FA made no errors in estimating across- or within-area d_{shared} . As in panel **a**, we plotted the estimated d_{shared} values from only one area, and the other area yielded similar results. Taken together, these results demonstrate that pCCA-FA was able to identify and partition across- and within-area shared variability even with modest trial counts.

* Source data are provided as a Source Data file.



Supplementary Fig. 3: Recovery of %sv depends on angle between across- and within-area co-fluctuation patterns.

In Fig. 5, we found that the angle between across- and within-area co-fluctuation patterns tended to be large (mean across sessions = 68.2°). We asked how well pCCA-FA could partition across- and within-area variance as the across- and within-area co-fluctuation patterns varied in their similarity. To test this, we conducted simulations in which we systematically varied the angle between the ground truth across- and within-area co-fluctuation patterns. We then asked how accurately pCCA-FA estimated across- and within-area %sv. In each simulation, we specified the angle between the top across- and within-area co-fluctuation pattern (i.e., the angle between the first column of the across- and within-area loading matrices; θ_{sim} , see Methods). We wished to test the dependence of the estimation procedure only on the value of θ_{sim} , so we matched the level of %sv across all simulations and all values of θ_{sim} . To do this, we generated loading matrices W_m and L_m for each area m as in Supplementary Fig. 2, and scaled their columns to ensure that the ground truth value of across-area %sv was $\approx 15\%$ and within-area %sv was $\approx 10\%$. In this analysis, we scaled the columns of each loading matrix using singular values that followed an exponential distribution (i.e.,

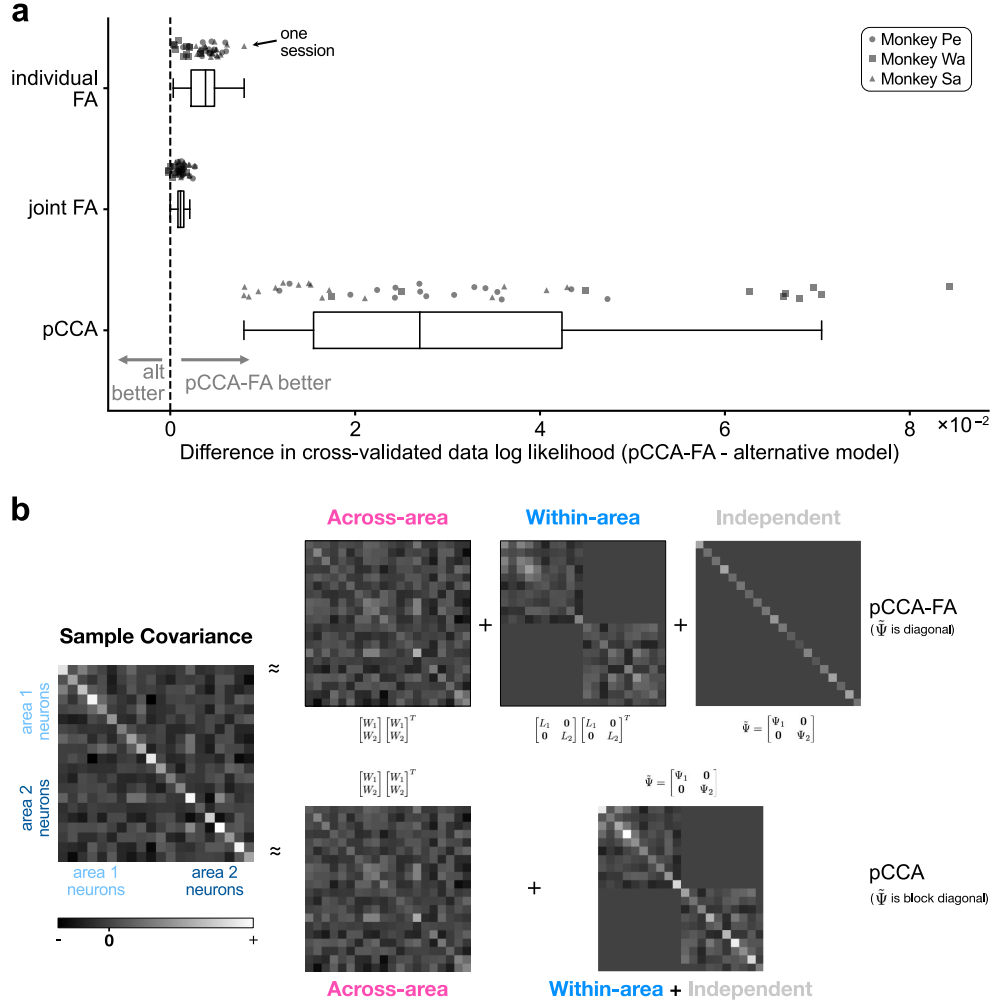
the first co-fluctuation pattern had the highest singular value, and each subsequent co-fluctuation pattern had smaller singular values). This ensured that application of the singular value decomposition did not change the order of the columns, which would lead to an incorrect value of θ_{sim} . The independent variance values (i.e., entries of Ψ_m) were then scaled to achieve 75% independent variance. We selected other parameters to be within the range of values observed in neural recordings: 30 neurons per area, 5 across-area latent variables ($d = 5$) and 3 within-area latent variables for each area ($d_m = 3$, for each area m). We repeated the simulation procedure 100 times for each value of θ_{sim} , using 1,000 trials per simulated dataset. In this analysis, we were interested in the partitioning of across- and within-area shared variance, so we fit a single model using all trials at the ground truth values of across- and within-area dimensionality, then computed the estimated values of across- and within-area %sv. We report the ground truth and estimated values of across-area %sv (pink points, one per simulated brain area) and within-area %sv (blue points, one per simulated brain area; light and dark blue for areas 1 and 2, respectively), as well as the total %sv, which is computed as the across-area %sv plus the within-area %sv (black point, mean $\pm$ standard deviation, pooled across simulated brain areas, 2 per simulated dataset, $n = 200$).

a, When the across- and within-area co-fluctuation patterns were aligned ($\theta_{\text{sim}} = 1^\circ$; left), the estimated %sv was within ≈ 5 %sv of the ground truth value (each point represents a simulated brain area; dashed line indicates zero error). Although the individual across-area (pink) and within-area (light and dark blue for areas 1 and 2, respectively) points sometimes deviated from the ground truth (indicated by their vertical offset from the dashed line), the total %sv (i.e., across-area %sv plus within-area %sv, indicated by black points) was well estimated (close to dashed line). Thus, the estimation errors of across- and within-area %sv involved overestimates of one type of %sv (across- or within-area) with corresponding underestimates of the other, thereby providing an accurate estimate of the total ground truth level of %sv. Each individual simulation contributed four points to the scatter, one per simulated brain area for across-area (pink) and within-area (blue) values. As the co-fluctuation patterns became more distinct ($\theta_{\text{sim}} = 20^\circ, 80^\circ$; middle and right, respectively), pCCA-FA became more accurate in estimating across- and within-area %sv (pink and blue points approaching dashed line).

b, To provide a fuller picture of the dependence of estimated %sv on θ_{sim} , we systematically varied the alignment of the across- and within-area co-fluctuation patterns (θ_{sim}). We then asked how the estimated partitioning of variance depended on the alignment of the co-fluctuation patterns. Estimated %sv settled very close to the ground truth value and variance across estimates decreased quickly as co-fluctuation patterns became more distinct (from left to right). Note that the results shown at $\theta_{\text{sim}} \in [1^\circ, 20^\circ, 80^\circ]$ (marked by dashed lines) match the panels in **a**. The error bars indicate the mean $\pm$ one standard deviation across 100 independent simulations (pooled across simulated brain areas, 2 per simulated dataset, $n = 200$). Together, these results indicate that across- and within-area co-fluctuation patterns were well estimated even when they were quite similar in the ground truth. As the across- and within-area co-fluctuation patterns became more distinct, the estimation errors of

across- and within-area %sv also decreased.

* Source data are provided as a Source Data file.



Supplementary Fig. 4: pCCA-FA outperformed alternative models on neural recordings.

Here we compared pCCA-FA to alternative statistical models that can be used to describe neural activity recorded simultaneously from two areas. We focused on probabilistic models tailored for studying trial-to-trial variability (without a time series model). Overall, we found that pCCA-FA outperformed alternative models on our dual-hemisphere PFC recordings (measured in terms of higher cross-validated data likelihood). In addition to the greater generalization performance, pCCA-FA distinguishes between across- and within-area interactions, providing a more thorough dissection of neural activity than the alternative models that we considered.

a, For each recording session and model, we performed 10-fold cross-validation to identify the number of across- and within-area latent variables, and used the cross-validated data log likelihood of the selected model for comparison with other models (see Equation (6) in Methods for the pCCA-FA definition). For each experimental session, we computed the difference in data log likelihood between pCCA-FA and that of each alternative model under consideration. A value above zero indicates that pCCA-FA outperformed the alternative model. For plotting purposes, we normalized the log likelihood of each session by the number of trials and neurons so that the log likelihood values would be comparable across sessions. Paired statistical tests (see below) were performed using non-normalized values. Each box plot covers the first to third quartiles of the differences in log likelihood, with the median across sessions marked with a line. Each whisker covers an additional $1.5\times$ the interquartile range. Values for each individual session and animal are shown above the box plot for each candidate model.

(top) For two populations of neural activity, we can fit a separate factor analysis (FA) model to each area (termed individual FA model). Although this model does not identify across-area interactions, it is a simple model whose performance served as a baseline for interpreting the performance of pCCA-FA and other models. Note that this model is mathematically equivalent to pCCA-FA when the number of across-area latent variables is set to zero. We fit a FA model to each area, and calculated the data log likelihood for both areas by adding together the log likelihood from each single-area model. We found that pCCA-FA outperformed the individual FA model (one point per session; Monkey Pe: $p = 5.5e-8$, $n = 16$; Monkey Wa: $p = 0.0018$, $n = 10$; Monkey Sa: $p = 1.4e-6$, $n = 16$; paired one-tailed t-tests) because pCCA-FA captures activity that is correlated across areas, whereas the individual FA model does not.

(middle) We next fit FA jointly to all recorded neurons, without consideration that they were recorded from different areas (termed joint FA model). The joint FA model captures both across- and within-area interactions together, and is not designed to separate these two types of interactions as pCCA-FA is. Again, we found that pCCA-FA outperformed the joint FA model (Monkey Pe: $p = 1.1e-7$, $n = 16$; Monkey Wa: $p = 0.014$, $n = 10$; Monkey Sa: $p = 6.6e-7$, $n = 16$; paired one-tailed t-tests).

(bottom) Both formulations of FA considered so far are unable to identify across-area interactions separately from within-area interactions. To explicitly identify across-area interactions, we fit a pCCA model (see Methods, [5]) to the two areas. We found that pCCA-FA outperformed pCCA for all sessions (Monkey Pe: $p = 6.8e-12$, $n = 16$; Monkey Wa: $p = 4.5e-7$, $n = 10$; Monkey Sa: $p = 5.4e-9$, $n = 16$; paired one-tailed t-tests). The average number of trials per session ranged from 450 – 850, which is in the regime in which pCCA-FA outperformed pCCA in simulated data (Fig. 3e-f). pCCA-FA outperforms pCCA in this regime because pCCA-FA has fewer parameters than pCCA (see panel **b**, Methods).

b, We illustrate how pCCA-FA (top) has fewer parameters than pCCA (bottom). Suppose that we record from 10 neurons in each of areas 1 and 2. Together, the 20 neurons have a joint spike count covariance matrix (left). The color of each square in the heatmap indicates the corresponding value in the covariance matrix. The color

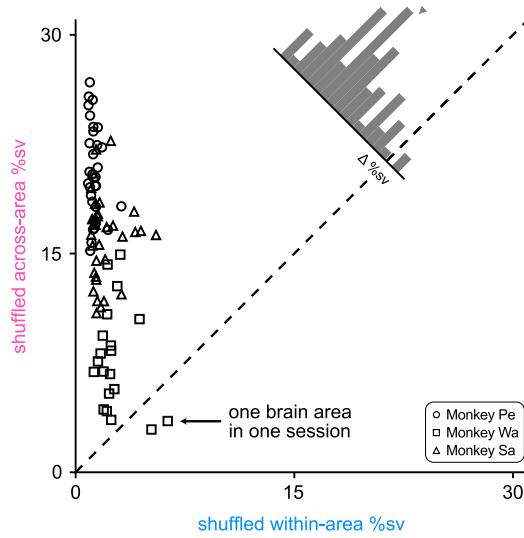
scale is indicated below the joint covariance matrix, and applies to all heatmaps. Precise numerical values are arbitrary and for illustration purposes only.

pCCA-FA approximates this joint spike count covariance matrix as a sum of a low-rank across-area component (top left), a block diagonal within-area component where each block is low-rank (top middle), and a diagonal independent neuron component (top right). This results in a parameter count of $(n_1 + n_2) \times d$ for the across-area component, $(n_1 \times d_1)$ for the within-area 1 component, $(n_2 \times d_2)$ for the within-area 2 component, and $(n_1 + n_2)$ for the independent neuron component.

In contrast, pCCA approximates the joint spike count covariance matrix as the sum of a low-rank across-area component (bottom left) and a block diagonal matrix where each block is full rank (bottom right). The block diagonal matrix includes all variance that is not captured by the across-area component, including any variance shared among neurons solely within each area and independent neuron variances. This results in a parameter count of $(n_1 + n_2) \times d$ for the across-area component and $(n_1 \times n_1) + (n_2 \times n_2)$ for the block diagonal matrix.

The number of parameters for the across-area component is the same for pCCA-FA and pCCA. However, pCCA-FA can have a substantially smaller number of parameters than pCCA in the remaining components, especially if the number of within-area latent variables in pCCA-FA is small relative to the number of neurons in each population. This explains the practical benefits of pCCA-FA over pCCA in the limited data regime (see panel **a**, Fig. 3e-f).

* Source data are provided as a Source Data file.

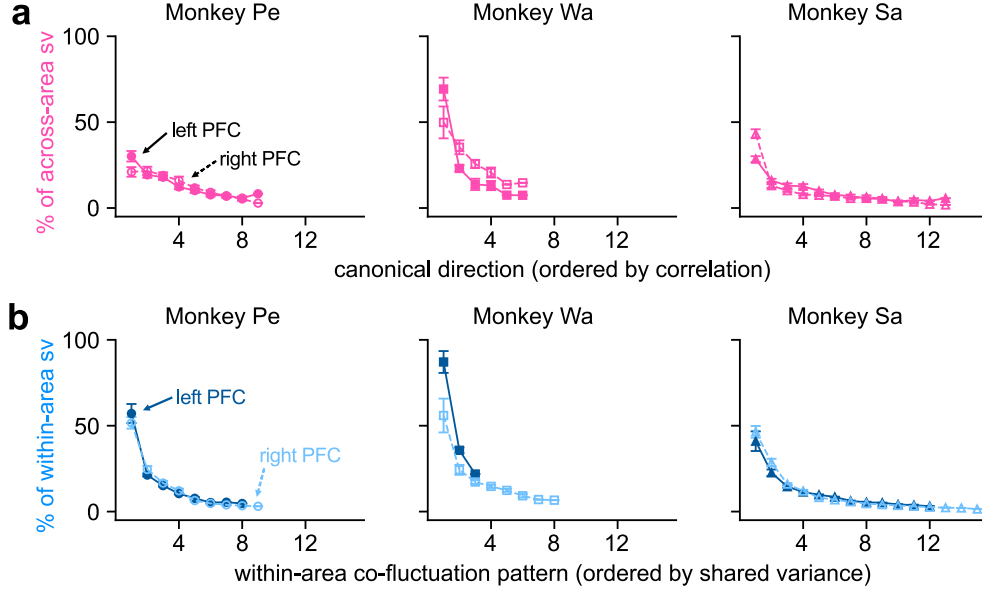


Supplementary Fig. 5: Shuffling area labels destroys within-area shared variance.

In Fig. 4b, we compared the amount of shared variance (%sv) in the across-area component versus the within-area component. Here, we validated that we identified meaningful shared variance in our PFC recordings by randomly permuting the area labels (i.e., right or left hemisphere) of the neurons. In other words, we randomly assigned each neuron to area 1 or area 2. When we permute the area labels of the neurons, each shuffled population contains some true area 1 neurons and some true area 2 neurons. We can think of each shuffled population as comprising neurons drawn from the same overall population of neurons. Thus, it is unlikely for there to be activity variance that is shared exclusively among neurons within a shuffled population. We would expect nearly all of the shared variance among neurons to be assigned to the across-area component, and little shared variance to be assigned to the within-area component.

To perform this shuffle control, we took the same spike counts that were used to fit pCCA-FA in our main analyses (Figs. 4-6, see Methods) and permuted the area labels of the neurons. Then, we fit pCCA-FA to the spike counts of the shuffled neurons in a cross-validated manner, identical to the analyses in the main figures (Figs. 4-6, see Methods). We computed the across- and within-area %sv using the parameters of the resulting pCCA-FA model in each session. Compared to Fig. 4b (which is the same analysis using the pCCA-FA models fit using true area labels, same plotting conventions used here), the within-area %sv (horizontal axis) in the shuffled models was reduced to near zero. The across-area %sv (vertical axis) was similar or slightly greater for the shuffled neurons compared to the neurons with true area labels. This indicates that the true area labels of the neurons and the differences we found in the across-area versus within-area interactions in Fig. 4 are meaningful.

* Source data are provided as a Source Data file.



Supplementary Fig. 6: Strength of across- and within-area co-fluctuation patterns.

In Fig. 5 we compared the similarity of the across- and within-area co-fluctuation patterns identified by pCCA-FA. Here, we analyze the relative strength of the co-fluctuation patterns.

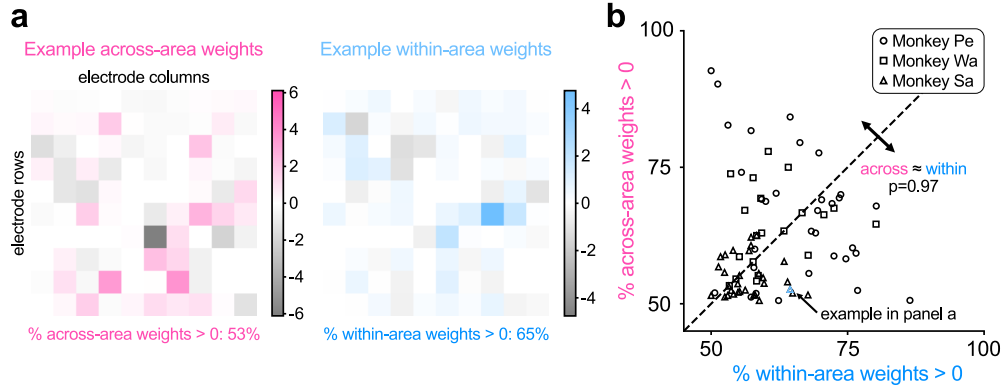
a, We wondered how dominant the first co-fluctuation pattern of each type was relative to the subsequent co-fluctuation patterns. The across-area latent variables in pCCA-FA are defined by the co-fluctuation patterns in each area that explain the most across-area shared variance. Here we transformed the across-area co-fluctuation patterns to order them by across-area correlation (as in CCA and pCCA). These co-fluctuation patterns correspond to the canonical directions identified by CCA or pCCA (with equivalence in the regime of sufficient data; see Methods). In all other analyses, we ordered the across-area co-fluctuation patterns by the amount of across-area shared variance. The transformation into the canonical directions allowed us to understand whether they also explained the most shared variance across areas, which is not guaranteed by the transformation. We computed the proportion of across-area shared variance explained by each of the most correlated co-fluctuation patterns across areas (i.e., canonical directions). We found that the co-fluctuation patterns that described the greatest correlation across areas indeed also tended to explain the most shared variance across areas.

Specifically, we computed the percentage of across-area shared variance explained by the i^{th} canonical direction $\mathbf{u}_{mi} \in \mathbb{R}^{n_m}$ in area m as: $\frac{\mathbf{u}_{mi}^T (W_m W_m^T) \mathbf{u}_{mi}}{\text{tr}(U_m^T (W_m W_m^T) U_m)}$, where $\text{tr}(\cdot)$ is the trace, and $U_m \in \mathbb{R}^{n_m \times d}$ is the matrix whose columns contain the d canonical directions $\mathbf{u}_{mi}, i = 1, 2, \dots, d$ (see Methods). Note that even though two curves are shown (one for neurons in right PFC, one for neurons in left PFC), the across-area

variance is described by a single set of latent variables that involves both areas. Error bars indicate the mean $\pm$ standard error across sessions in each animal. Because we averaged across sessions, the plotted values need not sum to 100% (though they do for each session individually).

b, We also assessed the relative strength of the within-area co-fluctuation patterns. By definition, the within-area latent variables identified by pCCA-FA explain the greatest shared variance within an area that is not explained by activity in the other area. As a post-processing step, we reordered the within-area co-fluctuation patterns by the amount of shared variance explained, such that the first co-fluctuation pattern explains the greatest shared variance, the second co-fluctuation pattern explains the second-most shared variance, etc. (see Methods). The percentage of within-area shared variance explained by the i^{th} within-area co-fluctuation pattern in area m is computed as: $\frac{\mathbf{a}_{mi}^T L_m L_m^T \mathbf{a}_{mi}}{\text{tr}(L_m L_m^T)}$, where $\mathbf{a}_{mi}$ is the i^{th} eigenvector of $L_m L_m^T$. The numerator of this expression is equivalent to the i^{th} eigenvalue of $L_m L_m^T$. Note that the two curves shown here now represent two different sets of latent variables identified by the pCCA-FA model: one involving neurons in right PFC (light blue), and one involving neurons in left PFC (dark blue). Across sessions, we found that the top co-fluctuation pattern explained a large proportion of shared variance relative to the subsequent co-fluctuation patterns (sharp drop off from first to second co-fluctuation patterns). This indicated that much of the within-area shared variance could be attributed to a single mode of variance, a finding that is sometimes termed low-rank variability [6]. Error bars indicate the mean $\pm$ standard error across sessions in each animal.

* Source data are provided as a Source Data file.



Supplementary Fig. 7: Characterizing weights of across- and within-area co-fluctuation patterns.

We examined the diversity of weights onto the neural populations of the top across- and within-area co-fluctuation patterns, which indicates the extent to which a co-fluctuation pattern involved all neurons increasing or decreasing their activity together. For example, if the across-area variance involved a global modulatory process, all neurons might increase their activity together or decrease their activity together. If this were the case we would observe many weights of the same sign, or a low diversity of weights.

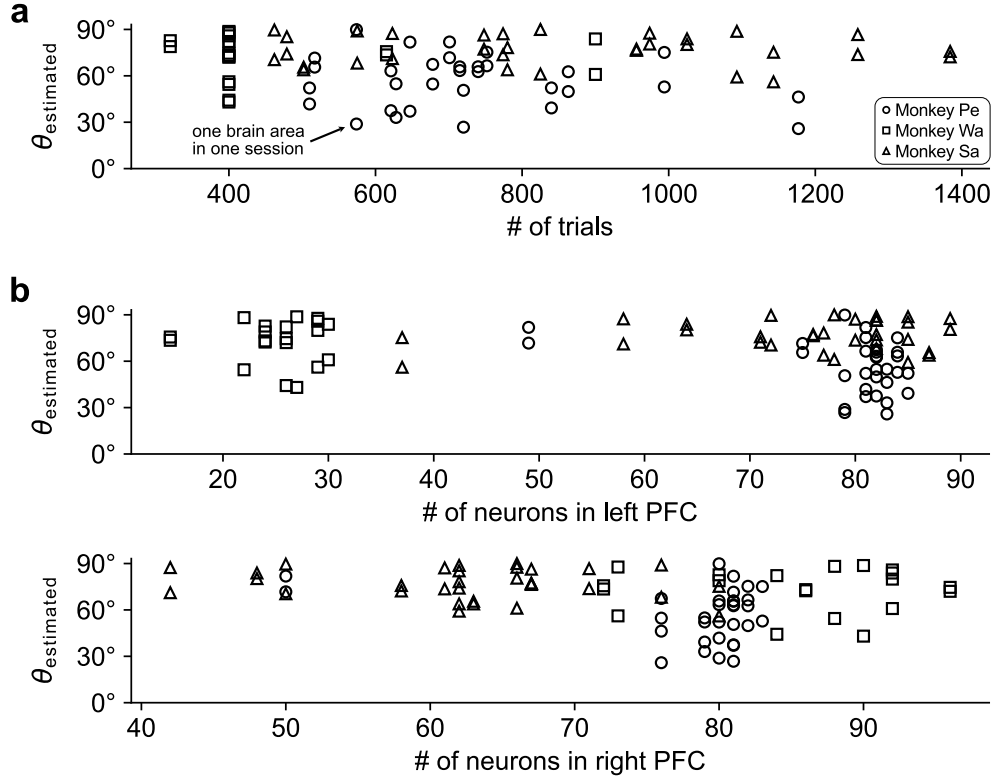
a, In an example session (Monkey Sa, right PFC), we investigated whether the weights on the top across- (left, pink) or within-area (right, blue) co-fluctuation patterns shared the same sign. The color maps indicate the signed value of the weight onto an individual neuron, arranged according to the spatial layout of the implanted electrode array. We found that positive (colored; pink or blue for weights on the across- and within-area co-fluctuation pattern, respectively) and negative (gray) weights were interspersed spatially and showed no apparent organization. Further, we found a mix of positive and negative weights on both the across- and within-area co-fluctuation patterns. In other words, the diversity of weights for this example session was high, with 53% of across-area and 65% of within-area weights greater than zero.

b, Across sessions, we quantified the diversity of weights by computing the percentage of same-signed weights for the top across- and within-area co-fluctuation pattern in each area (a total of four values per session). The overall signed orientation of a co-fluctuation pattern is arbitrary, as flipping the sign of every weight along a co-fluctuation pattern yields an equivalent model. We oriented the sign of each co-fluctuation pattern such that the majority of weights were positive (i.e., the minimum percentage of same-signed weights is 50%, which we notate as positive by convention). We performed the orientation change on co-fluctuation patterns from each area individually to ensure a fair comparison between across- and within-area co-fluctuation patterns. This sign change was only performed for this analysis. Each point corresponds to the top across- and within-area co-fluctuation pattern from one area, such

that each session contributes two points (one point for each area). The example session from panel **a** (only one area shown, right PFC) is marked in light blue.

We found that neither across- nor within-area co-fluctuation patterns had weights that were close to 100% positive. Rather than all neurons increasing and decreasing their activity together, the across- and within-area co-fluctuation patterns had diverse weights onto the neural populations. In fact, the across- and within-area co-fluctuation exhibited similar diversity in their weights. There was no significant difference between the percentage of positive weights in the top across- and within-area co-fluctuation patterns ($p = 0.97$, pooled across animals and brain areas, $n = 84$ brain areas; Monkey Pe: $p = 0.80$, $n = 32$; Monkey Wa: $p = 0.19$, $n = 20$; Monkey Sa: $p = 0.035$, $n = 32$; paired two-tailed t-tests). Notably this analysis does not include slow-timescale fluctuations in neural activity (which were removed in our preprocessing, see Methods), but instead focuses on fast-timescale trial-to-trial variability. Our previous work, across several brain areas and many animals [1, 7–9], has investigated the diversity of weights identified from neural population activity including slow-timescale fluctuations and found varied effects.

* Source data are provided as a Source Data file.



Supplementary Fig. 8: Neither trial count nor neuron count explained variation in $\theta_{\text{estimated}}$.

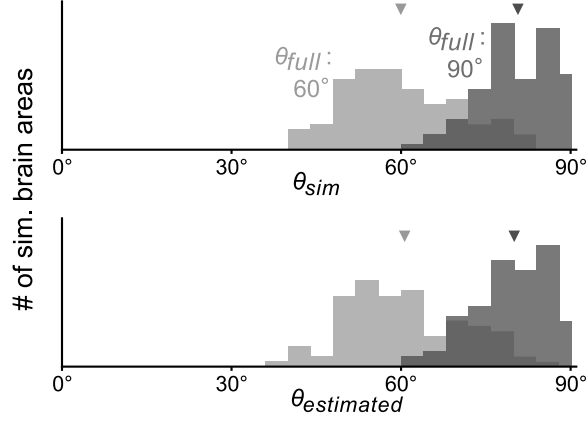
In Fig. 5c, we observed a distribution of angles between the top across- and within-area co-fluctuation patterns ($\theta_{\text{estimated}}$) across sessions (two angles per session, one per brain area). One possibility is that there existed a true θ in PFC, and estimation error due to lack of neurons or trials created the session-to-session variability of $\theta_{\text{estimated}}$. In this analysis, we confirmed that neither (a) trial count nor (b) neuron count explained the variation in $\theta_{\text{estimated}}$.

In all panels, the vertical axis values are the same distribution shown in Fig. 5c ($\theta_{\text{estimated}}$). Along the horizontal axis, we show either (a) the trial count or (b) neuron count (top and bottom panels for left and right PFC, respectively) from the corresponding session. Each point indicates the value from one brain area in one session, such that each session contributes two points to each panel (one per brain area).

We found no obvious relationship between the session-to-session variation of $\theta_{\text{estimated}}$ and trial count or neuron count. We split sessions into two groups based on the values of the horizontal axis in each panel (i.e., by number of trials or number of neurons). We then compared the values of $\theta_{\text{estimated}}$ in each of the two groups, which yielded a distribution of $\theta_{\text{estimated}}$ for low trial or neuron counts and a distribution of

$\theta_{\text{estimated}}$ for high trial or neuron counts. We found that the session-to-session variance of $\theta_{\text{estimated}}$ was not different between sessions with low trial counts versus high trial counts ($p = 0.63$; pooled across animals and brain areas, $n = 84$ brain areas, 2 per session; Levene’s test for equality of variance), or between sessions with low neuron counts versus high neuron counts (right PFC: $p = 0.36$, left PFC: $p = 0.30$; pooled across animals and brain areas, $n = 84$ brain areas, 2 per session; Levene’s test for equality of variance). We also found that the mean value of $\theta_{\text{estimated}}$ was unchanged by trial count ($p = 0.93$; pooled across animals and brain areas, $n = 84$ brain areas, 2 per session; two-tailed t-test) or neuron count (right PFC: $p = 0.055$, left PFC: $p = 0.041$; pooled across animals and brain areas, $n = 84$ brain areas, 2 per session; two-tailed t-test). Together these results indicate that the variation in $\theta_{\text{estimated}}$ across sessions in Fig. 5c was not due to estimation using different numbers of trials or neurons per session.

* Source data are provided as a Source Data file.



Supplementary Fig. 9: Sampling different subsets of neurons did not fully explain variation in $\theta_{\text{estimated}}$.

In Fig. 5c, we observed a distribution of angles between the top across- and within-area co-fluctuation patterns ($\theta_{\text{estimated}}$). We considered two possible explanations for this variation across sessions. First, in Fig. 5d, we considered whether the variation of $\theta_{\text{estimated}}$ was due to fitting pCCA-FA using a finite number of noisy simulated data samples. Second, we consider here the possibility that the variation arose from sampling different subsets of neurons on each recording session from a larger fixed population.

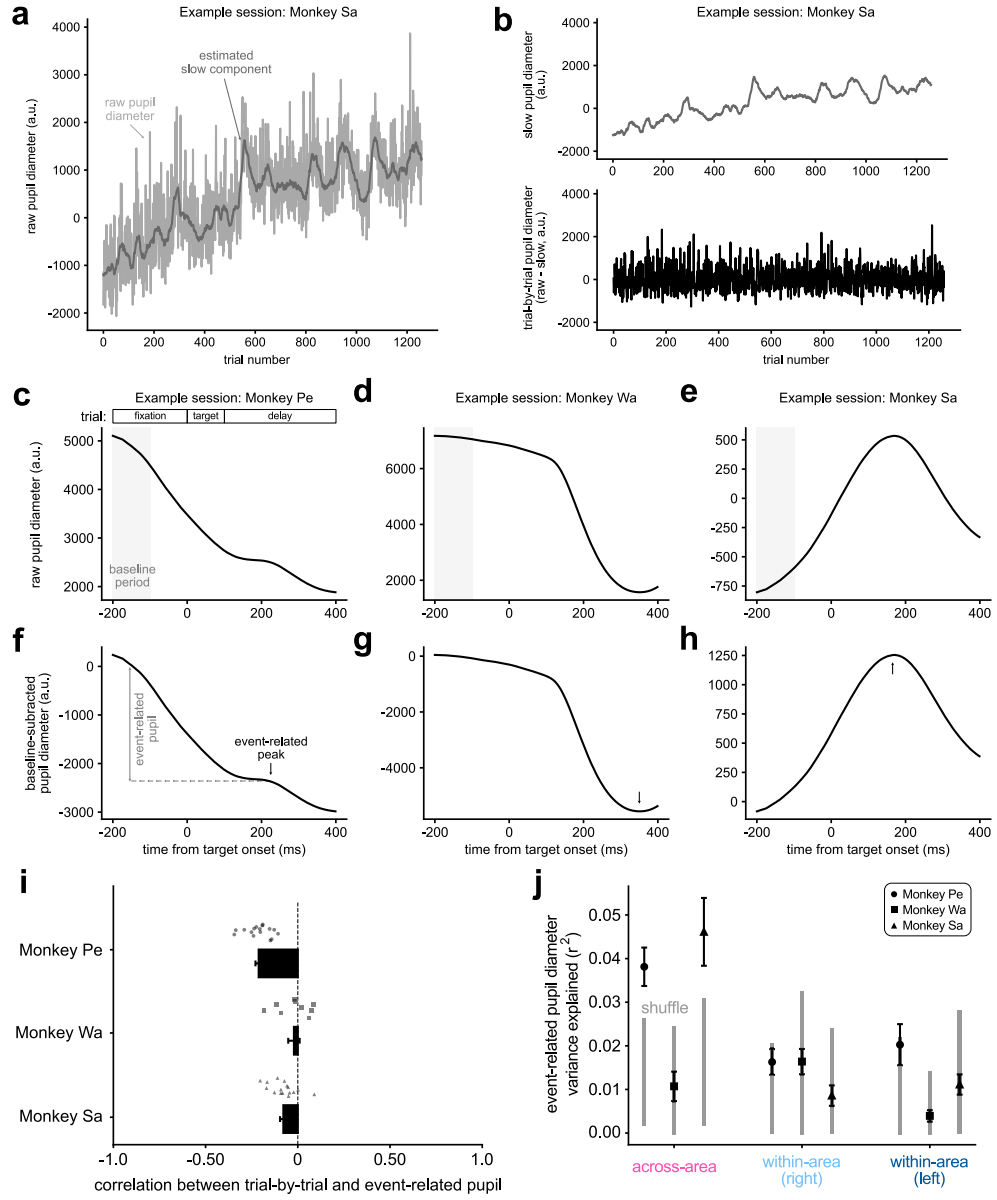
To simulate the second scenario, we defined a large number of neurons (10,000) representing the entirety of each simulated brain area (e.g., one hemisphere of PFC). For the 10,000 simulated neurons in each brain area, we generated random across- and within-area loading matrices W_m and L_m for each area m using independent draws from a standard normal distribution, and scaled their columns ($d = 5$ across-area latent variables, and $d_m = 3$ within-area latent variables for each area m) to ensure that the ground truth value of across-area %sv was $\approx 15\%$ and within-area %sv was $\approx 10\%$, as in Supplementary Fig. 2 and Fig. 3e-f. We then fixed the angle between the top across- and within-area co-fluctuation pattern, θ_{full} , to be either 60° or 90° (as in Methods section for Simulated data generation for θ_{sim}). Recordings with a Utah array cover only a fraction of the brain area and yield a partially overlapping subset of neurons on each recording session. To represent the pool of neurons from which we sample on each recording session, we randomly selected a subset of neurons in each simulated brain area ($n_{\text{sub1}} = 100$; $n_{\text{sub2}} = 100$). Then, for each of 100 independent simulations, we then randomly subsampled $n_1 = 30$ and $n_2 = 30$ neurons from the fixed pool of n_{sub1} and n_{sub2} neurons as the neurons recorded in that simulated session. To quantify the variation in angle that arises from sampling a different subset of neurons on each recording session, for each simulation, we computed the angle between the top across- and within-area co-fluctuation patterns (θ_{sim}) using only the n_1 and n_2 neurons. Looking across the 100 simulations yields a distribution of θ_{sim} for each value of θ_{full} (top panel, mean of each distribution marked with a triangle). For visual

clarity, we plotted the θ_{sim} values from only one simulated brain area, as both areas yielded similar results. The width of each of these distributions (100 values in each distribution) represents the variation in angle that arises from sampling a different subset of neurons on each recording session.

For each simulation (corresponding to one value of θ_{sim} above), we generated 1,000 trials of simulated data and fit a pCCA-FA model using the ground truth values of across- and within-area dimensionality, as in Fig. 5d. Using the across- and within-area loading matrices estimated by pCCA-FA, we computed one value of $\theta_{\text{estimated}}$ per simulated brain area. Looking across the 100 simulations yields a distribution of $\theta_{\text{estimated}}$ for each value of θ_{full} (bottom panel, mean of each distribution marked with a triangle). As in the top panel, we plotted the $\theta_{\text{estimated}}$ values from one simulated brain area, as both areas yielded similar results. Note that each element of the histogram of $\theta_{\text{estimated}}$ here corresponds to one element of the histogram of θ_{sim} in the top panel. The width of each of these distributions (100 values in each distribution) represents the variation in estimated angle arising both from sampling a different subset of neurons on each recording session and sampling a finite number of noisy data samples.

Interestingly, the distributions of $\theta_{\text{estimated}}$ in the lower panel are wider than those reported in Fig. 5d for the same values of θ_{sim} (for $\theta_{\text{full}} = 60^\circ$, $p = 6.7\text{e-}5$; for $\theta_{\text{full}} = 90^\circ$, $p = 7.4\text{e-}6$; Levene’s tests comparing equality of variance for the distribution of $\theta_{\text{estimated}}$ in the lower panel to the corresponding distribution in Fig. 5d). This indicates that some of the breadth of the distribution of $\theta_{\text{estimated}}$ from our neural recordings in Fig. 5c may be attributable to sampling a different subset of neurons on each recording session. These results, together with those in Fig. 5d, indicate that sampling a different subset of neurons on each recording session and fitting pCCA-FA to a finite number of noisy samples cannot fully account for the width of the distribution of $\theta_{\text{estimated}}$ that we observed in the neural recordings. This supports the conclusion that the across- and within-area components involved co-fluctuation patterns that varied in similarity from session to session.

* Source data are provided as a Source Data file.



Supplementary Fig. 10: Across-area latent variables were related to event-related pupil diameter.

In Fig. 6, we investigated whether the across- or within-area components of neural activity were related to the trial-by-trial pupil diameter measured during the delay period (relative to the ongoing 25-trial average, see Methods). This assessed trial-to-trial fluctuations in pupil diameter relative to the ongoing mean (with a baseline

subtraction that was matched to our neural data processing), and did not explicitly assess event-related changes in the pupil locked to saccade target onset. Measures of tonic pupil size have been linked to arousal and other cognitive states [10, 11]. However, cognitive modulation has also been observed in evoked or event-related pupillary dynamics [12, 13]. Here, we investigated whether the across- and within-area components of neural activity were related to event-related pupil diameter and compared that to the delay period pupil measurements presented in Fig. 6.

a, While animals performed the task, we monitored the raw pupil size during each trial’s delay period over the course of the session (light gray). The size of the pupil slowly fluctuated over the course of the session [1], in this example session drifting upwards. As with our preprocessing of neural activity, we estimated the slowly varying component of pupil diameter by computing a moving average of 25 trials (dark gray, see Methods).

b, We removed the estimated slow component of pupil diameter (top) from the raw pupil diameter. This yielded what we termed the trial-by-trial pupil diameter, one value per trial (bottom). In Fig. 6, we found that the trial-by-trial pupil diameter was related to across-area latent variables.

c-e, We next examined the changes in pupil diameter that occur just after saccade target onset (evoked or event-related pupil diameter). Here, we plotted the average within-trial raw pupil diameter trace from an example session from each animal (Monkey Pe: panel **c**, Monkey Wa: panel **d**, Monkey Sa: panel **e**). After target onset (time = 0), the diameter of the pupil fluctuated rapidly over the course of a few hundred milliseconds. To compensate for slow changes in pupil diameter from moment to moment, we identified a baseline period from 100-200ms before target onset (gray shading). The average baseline pupil diameter during this window was subtracted from the pupil diameter trace on a trial-by-trial basis. This subtraction equalized the size of the pupil at the start of each trial to effectively measure the event-related pupil diameter magnitude as a deviation from this baseline value. Note that this differs from the trial-by-trial pupil diameter measure, for which the baseline is identified on a moving average of 25 trials rather than on every trial.

f-h, The baseline-subtracted pupil diameter traces averaged across trials are shown for an example session from each animal (Monkey Pe: panel **f**, Monkey Wa: panel **g**, Monkey Sa: panel **h**). We defined the event-related pupil diameter as the change in pupil diameter from the baseline period (-200 to -100ms before target onset, now at zero (a.u.)) to the manually identified event-related peak in pupil diameter (marked by arrows). This peak was manually adjusted based on the trial-averaged pupil trace in each session to account for shifts in observed pupil latencies across sessions (Monkey Pe: 221-297ms, Monkey Wa: 276-356ms, Monkey Sa: 127-200ms after target onset). We measured the mean pupil diameter in a 100ms window centered on the identified peak, which is relative to the baseline window on each trial. To compare across sessions, we z-scored the event-related pupil diameter within each session.

i, We found that the magnitude of the event-related pupil diameter was negatively correlated with the trial-by-trial pupil diameter across sessions (mean Pearson’s correlation, Monkey Pe: $\rho = -0.21$, Monkey Wa: $\rho = -0.021$, Monkey Sa: $\rho = -0.078$) in two of three animals ($p = 1.0e-8$; pooled across animals, $n = 41$ sessions; Monkey Pe,

$p = 1.9\text{e-}9$, $n = 16$; Monkey Wa, $p = 0.26$, $n = 9$; Monkey Sa, $p = 0.00026$, $n = 16$; one-tailed t-tests against zero). The bar plots indicate the means across sessions for each animal, and the error bars indicate the standard error of the mean. While the negative relationship we observed might appear counterintuitive, it in fact aligns with previous studies [8, 14–16] and may reflect physical limits imposed by the musculature of the iris [17]. That is, during a time period that the trial-by-trial pupil diameter is especially large the biomechanics of the iris may constrain any event-related pupillary response to be smaller than usual (and vice versa).

j, Although the trial-by-trial pupil diameter and event-related pupil might be correlated merely due to physical constraints from the iris musculature, it is also possible that they are influenced by different arousal processes and timecourses. As such, they might relate independently to the across-area component of neural activity. To this end, we performed an analogous regression analysis to that of Fig. 6 to relate the latent variables identified by pCCA-FA to the event-related pupil size (see Methods). Briefly, we regressed the full set of either the across- or within-area latent variables onto the event-related pupil diameter for each trial. We then assessed r^2 , the proportion of variance in event-related pupil diameter explained by the regression. As in Fig. 6c (same plotting conventions used here), we found that across-area latent variables explained more variance in the event-related pupil diameter than the within-area latent variables. We also report a 95% confidence interval over a shuffle distribution (gray bars, see Methods).

Aggregated across sessions, r^2 values for explaining event-related pupil diameter aggregated across sessions were greater for across-area latent variables (mean $r^2 = 0.035$) than for within-area latent variables (right PFC, mean $r^2 = 0.013$; left PFC, mean $r^2 = 0.013$) for two of three animals ($p = 8.1\text{e-}10$; pooled across animals and brain areas, $n = 82$ brain areas, 2 per session; Monkey Pe: $p = 1.8\text{e-}5$, $n = 32$; Monkey Wa: $p = 0.43$, $n = 18$; Monkey Sa: $p = 4.2\text{e-}7$, $n = 32$; one-tailed t-tests on the differences in r^2 for each session). When controlling for the number of latent variables (see Methods), we found similar results ($p = 0.00052$; pooled across animals and brain areas, $n = 82$ brain areas, 2 per session; one-tailed t-tests on the differences in r^2 for each session).

Further, we determined that the across-area latent variables explained variance in the event-related pupil beyond what was explained by its correlation with the trial-by-trial pupil diameter (mean residual $r^2 = 0.028$). To do this, we regressed out the contribution of the trial-by-trial pupil diameter to the event-related pupil, yielding a residual event-related pupil measure. Then, we assessed the variance explained by the across- and within-area latent variables in the residual event-related pupil diameter using the same linear regression approach as above (across-area residual $r^2 >$ shuffle r^2 , $p = 1.1\text{e-}5$; pooled across animals, $n = 41$ sessions; Monkey Pe, $p = 4.9\text{e-}5$, $n = 16$; Monkey Wa, $p = 0.13$, $n = 9$; Monkey Sa, $p = 0.0025$, $n = 16$; one-tailed t-tests against the mean of the shuffle distributions). Together with Fig. 6, we found that the across-area latent variables were related to multiple timescales of pupil activity, suggesting a link to one or more brainwide arousal processes.

* Source data are provided as a Source Data file.

References

- [1] 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 (2020).
- [2] Harris, K. D. Nonsense correlations in neuroscience. Preprint at <https://doi.org/10.1101/2020.11.29.402719> (2021).
- [3] Gokcen, E., *et al.* Disentangling the flow of signals between populations of neurons. *Nature Computational Science* **2**, 512–525 (2022).
- [4] Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes For Machine Learning* (The MIT Press, 2006).
- [5] Bach, F. R. & Jordan, M. I. A Probabilistic Interpretation of Canonical Correlation Analysis. Technical Report 688, University of California, Berkeley, Department of Statistics (2005).
- [6] Huang, C., *et al.* Circuit models of low-dimensional shared variability in cortical networks. *Neuron* **101**, 337–348 (2019).
- [7] Hennig, J. A., *et al.* Learning is shaped by abrupt changes in neural engagement. *Nature Neuroscience* **24**, 727–736 (2021).
- [8] Johnston, R., *et al.* The eyes reflect an internal cognitive state hidden in the population activity of cortical neurons. *Cerebral Cortex* **32**, 3331–3346 (2021).
- [9] Johnston, R. & Smith, M. A. Brain-wide arousal signals are segregated from movement planning in the superior colliculus. *eLife* **13**, RP99278 (2024).
- [10] Strauch, C., *et al.* Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences* **45**, 635–647 (2022).
- [11] Burlingham, C. S., Mirbagheri, S. & Heeger, D. J. A unified model of the task-evoked pupil response. *Science Advances* **8**, eabi9979 (2022).
- [12] Mathôt, S. & Van der Stigchel, S. New light on the mind’s eye: The pupillary light response as active vision. *Current Directions in Psychological Science* **24**, 374–378 (2015).
- [13] Ebitz, R. B. & Moore, T. Both a gauge and a filter: cognitive modulations of pupil size. *Frontiers in Neurology* **9 - 2018** (2019).
- [14] Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M. & Cohen, J. D. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience* **10**, 252–269 (2010).

- [15] Murphy, P. R., *et al.* Pupillometry and P3 index the locus coeruleus – Noradrenergic arousal function in humans. *Psychophysiology* **48**, 1532–1543 (2011).
- [16] Joshi, S., *et al.* Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* **89**, 221–234 (2016).
- [17] Loewenfeld, I. E. & Newsome, D. A. Iris mechanics I. Influence of pupil size on dynamics of pupillary movements. *American Journal of Ophthalmology* **71**, 347–362 (1971).