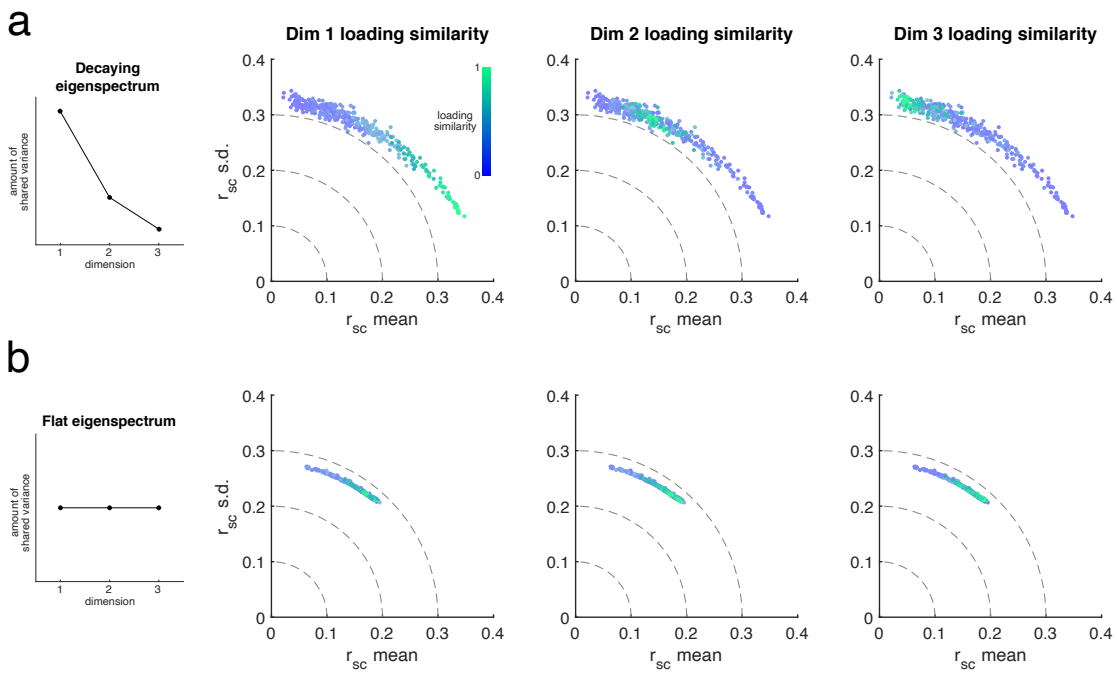
Neuron, Volume *109*

Supplemental information

Bridging neuronal correlations and dimensionality reduction

Akash Umakantha, Rudina Morina, Benjamin R. Cowley, Adam C. Snyder, Matthew A. Smith, and Byron M. Yu

Supplementary Information



Supplementary Figure 1: Relationship between pairwise metrics, loading similarity of each latent dimension, and the relative strengths of each dimension. Related to Figure 4.

In Fig. 3*e* and Math Note A, we considered the relationship between loading similarity and pairwise metrics when population activity was one dimensional. Here, we asked about the informativeness of loading similarity when population activity varies along multiple dimensions, and the impact of the relative strengths of each dimension (i.e., the shape of the eigenspectrum of Σ_{shared}, which specifies the amount of shared variance explained by each dimension).

We considered two cases. First, we considered an eigenspectrum that decays quickly, as has been widely reported in population recordings (Sadtler et al., 2014; Williamson et al., 2016; Mazzucato et al., 2016; Gallego et al., 2018; Huang et al., 2019; Stringer et al., 2019a; Ruff et al., 2020). In this case, we found that the loading similarity of the strongest dimension (i.e., dimension with largest eigenvalue) was most informative about pairwise metrics, while the loading similarities of the other dimensions were less informative. Second, we considered a flat eigenspectrum. In this case, the loading similarities of each dimension were equally informative.

a. Loading similarity for a decaying eigenspectrum of the shared covariance matrix (Σ_{shared} in Supplementary Fig. 5*a*). We reproduced the simulation in Fig. 3 for a latent dimensionality of 3 and %sv=50%. For each 3-d model, we evaluated the r_{sc} mean and s.d., and then plotted the same point in 3 separate panels colored by loading similarity of each of the 3 different dimensions. The loading similarity of strongest dimension ('Dim 1') is very informative–high loading similarity implies high r_{sc} mean and low r_{sc} s.d. (green dots), whereas low loading similarity implies low r_{sc} mean and high r_{sc} s.d. (blue dots). This is the same relationship as shown in Fig. 3*e* for the case of one dimension. The loading similarities of 'Dim 2' and 'Dim 3' are less informative–in both cases, low loading similarity points (blue dots) are scattered throughout the arc. The only case when the loading similarity of 'Dim 2' or 'Dim 3' is informative is when either of them have a high loading similarity (green dots). This is informative because it implies that 'Dim 1' must have low loading similarity ('Dim 1' is blue for dots where 'Dim 3' is green; see Math Note E), implying low r_{sc} mean and high r_{sc} s.d. (continued on next page...)

Supplementary Figure 1: (continued from previous page...)

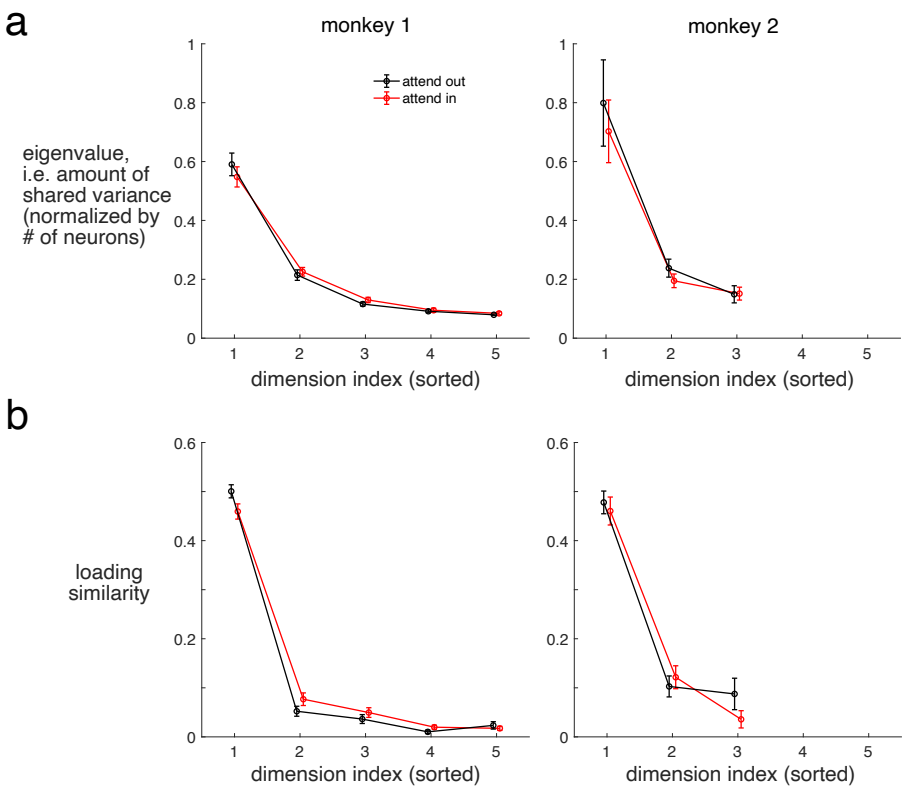
b. Same as panel **a** but for flat eigenspectrum across the three dimensions. In this case, r_{sc} mean will tend to be small and r_{sc} s.d. will tend to be large because: 1) all three dimensions contribute equally, and 2) it is not possible for all three dimensions to have high loading similarity, while multiple dimensions can have low loading similarity (Math Note E). However, knowing whether any of the three dimensions have high loading similarity can provide more specific information about r_{sc} mean and s.d. within this limited range (green dots tend to have high r_{sc} mean and lower r_{sc} s.d. in each panel).

Because most studies of population neuronal recordings have shown quickly decaying eigenspectra as in panel **a** (Sadtler et al., 2014; Williamson et al., 2016; Mazzucato et al., 2016; Gallego et al., 2018; Huang et al., 2019; Stringer et al., 2019a; Ruff et al., 2020), we recommend considering the loading similarity of the strongest dimension for concision and simplicity (as we do in Fig. 6*c*; and see eigenspectra in Supplementary Fig. 2). However, if it happens that the data have an eigenspectrum that decays slowly or has multiple dimensions that are very strong, then one may benefit by considering the loading similarities of additional dimensions as well.

This analysis also highlights how the shape of the eigenspectrum influences pairwise metrics. First, an exponentially-decaying eigenspectrum tended to have a higher r_{sc} mean and s.d. compared to its corresponding flat eigenspectrum (dots in panel **a** are farther from origin here than in panel *a*). This occurs because, for an exponentially-decaying eigenspectrum, an added dimension explains relatively little shared variance. Thus, the added dimension tends to result in only a small decrease in r_{sc} mean and s.d. On the other hand, adding a dimension to the flat eigenspectrum affects r_{sc} mean and s.d. as much as any other dimension, leading to larger changes (i.e., decreases) in r_{sc} mean and s.d. than in the case of an exponentially-decaying eigenspectrum.

Second, we observed a greater radial and angular spread for exponentially-decaying eigenspectra (panel **a**) compared to flat eigenspectra (panel **b**). This occurs because, when the eigenspectra are not flat, there is greater diversity in how the co-fluctuation patterns of different dimensions can contribute to r_{sc}. In other words, permuting the eigenvectors of three dimensions with equal eigenvalues (i.e., both dimensions explain the same amount of shared variance) results in the same model and same covariance matrix—yielding the same values for r_{sc} mean and s.d. However, permuting the eigenvectors of three dimensions with different eigenvalues will likely result in a different covariance matrix and different values of r_{sc} mean and s.d. Thus, for non-flat eigenspectra, the greater diversity by which co-fluctuation patterns can contribute to the shared covariance matrix leads to greater spread in the r_{sc} mean vs s.d. plots. The mathematical details regarding this observation are provided in Math Note D.

An implication of this analysis is that it is important to report the eigenspectrum shape whenever one reports dimensionality. Thus, considering both dimensionality and the eigenspectrum curve, instead of dimensionality alone, will lead to a more complete picture of the structure of population activity. Inspecting the eigenspectrum will also help determine whether assessing loading similarity in the strongest dimension is sufficient (panel **a**), or whether one needs to consider the loading similarities of other dimensions as well (panel **b**).



Supplementary Figure 2: Eigenvalues and loading similarity by dimension for V4 population activity. Related to Figure 6.

Although we observed only a modest change in dimensionality with attention (Fig. 6*c*), our simulations showed that the relative strength of each dimension (i.e., shape of the shared eigenspectrum) could alter the "effective dimensionality" of population activity and have large effects on pairwise metrics (Fig. 4*a*). Here, we asked whether the relative strengths of each dimension changed with attention. We also considered the loading similarities across different dimensions.

a. We found that the shape of the eigenspectra was qualitatively similar for 'attend in' and 'attend out' conditions (red and black curves have similar shape). In both conditions, the eigenvalues of the shared covariance matrix decayed (dot for each subsequent dimension was below dot for the previous dimension), indicating that a small number of dimensions were needed to explain the population-wide covariability.

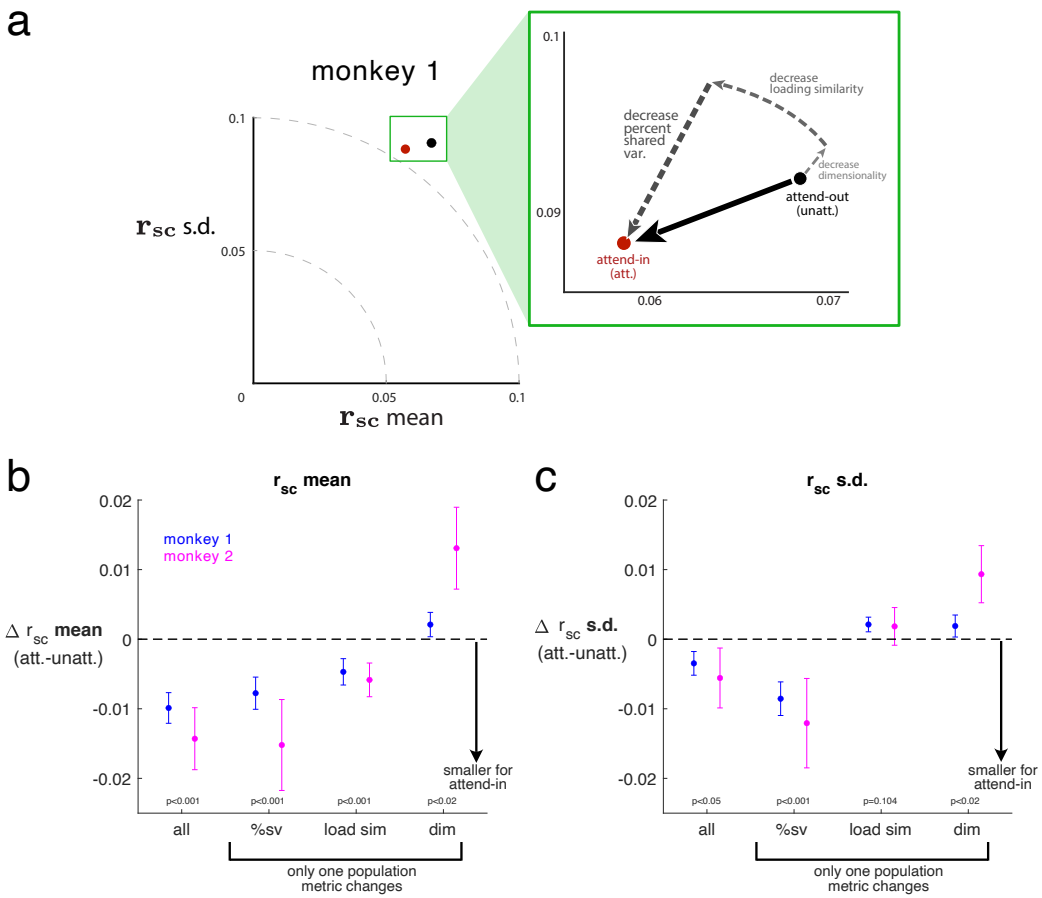
When comparing eigenspectra (i.e., the amount of shared variance explained by each dimension), one also needs to consider the firing rates under each condition. Mean firing rates tend to be higher for attend-in than attend-out trials. Higher firing rates typically correspond to higher spike count variance due to the Poisson-like firing of neurons. All else being equal, the higher mean firing rates imply higher levels of both shared variance and independent variance (Churchland et al., 2010). Thus, a direct comparison of the eigenspectra should be done with caution. Nonetheless, we plotted attend-in and attend-out together to relate our results to previous reports (Huang et al., 2019; Ruff et al., 2020). Consistent with these studies, we found that attention decreased the strength of the strongest dimension (red below black dot for dimension index 1), though the magnitude of the decrease we observed was more consistent with Ruff et al. (2020) than Huang et al. (2019). Had we been able to equalize the mean firing rate across the two conditions, we likely would have observed an even greater difference between attend-in and attend-out. We note that the caveat described here for comparison of eigenspectra (i.e., the amount of shared variance) does not apply to comparisons of %sv (Fig. 6*c*) because %sv is normalized by the overall spike count variance.

(...continued on next page)

Supplementary Figure 2: (...continued from previous page)

The eigenspectra were computed in the following way. We decomposed the V4 spike count covariance matrix into shared and independent components using factor analysis (see Methods). We then computed the eigendecomposition of the shared covariance matrix (Supplementary Fig. 5, $\Sigma_{\text{shared}} = U\Lambda U$). We found that eigenvalues (diagonal of Λ) tended to increase linearly with the number of neurons recorded; therefore, in order to combine across sessions, we normalized the eigenvalues by dividing by the number of neurons recorded in each session. After normalizing, we computed the eigenspectrum averaged across sessions and stimulus orientations. Because the dimensionality identified by cross-validation differed across sessions, there were a different number of sessions that contributed to each average. We did not plot mean eigenvalues when there were fewer than 5 sessions to average (i.e., dimensions ≥ 6 for monkey 1; dimensions ≥ 4 for monkey 2). Error bars indicate standard error. Data points have been jittered horizontally for visual clarity.

b. Loading similarity for 'attend-in' (red) and 'attend-out' (black) by dimension. Pooled across monkeys, the loading similarity for the first (i.e., strongest) dimension was larger for 'attend-out' than 'attend-in' (same result as Fig. 6*c*). We also observed some differences in loading similarity across the other dimensions in both monkeys. These differences could be important for specific scientific questions (see Fig. 7, for example). However, as we show in Fig. 4, Supplementary Fig. 1, and Math Note C, the first dimension plays the largest role in determining the r_{sc} distribution because it explains the greatest amount of shared variance.



Supplementary Figure 3: Quantifying the extent to which each population metric contributes to changes in pairwise metrics. Related to Figure 6.

In Fig. 6*c*, we observed changes in several population metrics with attention in V4 population responses. However, it was unclear to what degree the change we observed in each population metric contributed to the overall changes in pairwise metrics (Fig. 6*b*). In order to quantify this, here we used a population metric matching procedure to assess how much each individual change in a population metric contributed to the changes in a pairwise metric. We found that for these V4 data, %sv contributes the most, followed by loading similarity, and finally dimensionality. We illustrate these results in Fig. 6*d* (also reproduced here as panel **a** for convenience).

a. Reproduction of Fig. 6*d* to aid the interpretation of panels **b** and **c** here. For pairwise metrics, we observed decreases in both r_{sc} mean and s.d. with attention. For population metrics, we observed decreases in %sv, loading similarity, and dimensionality with attention.

b. Contribution of population metrics to changes in r_{sc} mean. For each recording session, we assessed how allowing all population metrics to vary ("all") or only a single population metric to vary between "attend-out" (unatt.) and "attend-in" (att.) influenced r_{sc} mean. The procedure for assessing this is detailed at the end of the caption ("Details of population metric matching procedure"). When only %sv or only loading similarity were allowed to vary, r_{sc} mean decreased with attention; when only dimensionality was allowed to vary, r_{sc} mean increased. When all population metrics were allowed to vary, r_{sc} mean decreased, consistent computations directly from data (Fig. 6*b*). Results for both monkeys were consistent; means and standard errors across sessions are shown.

(continued on next page...)

Supplementary Figure 3:

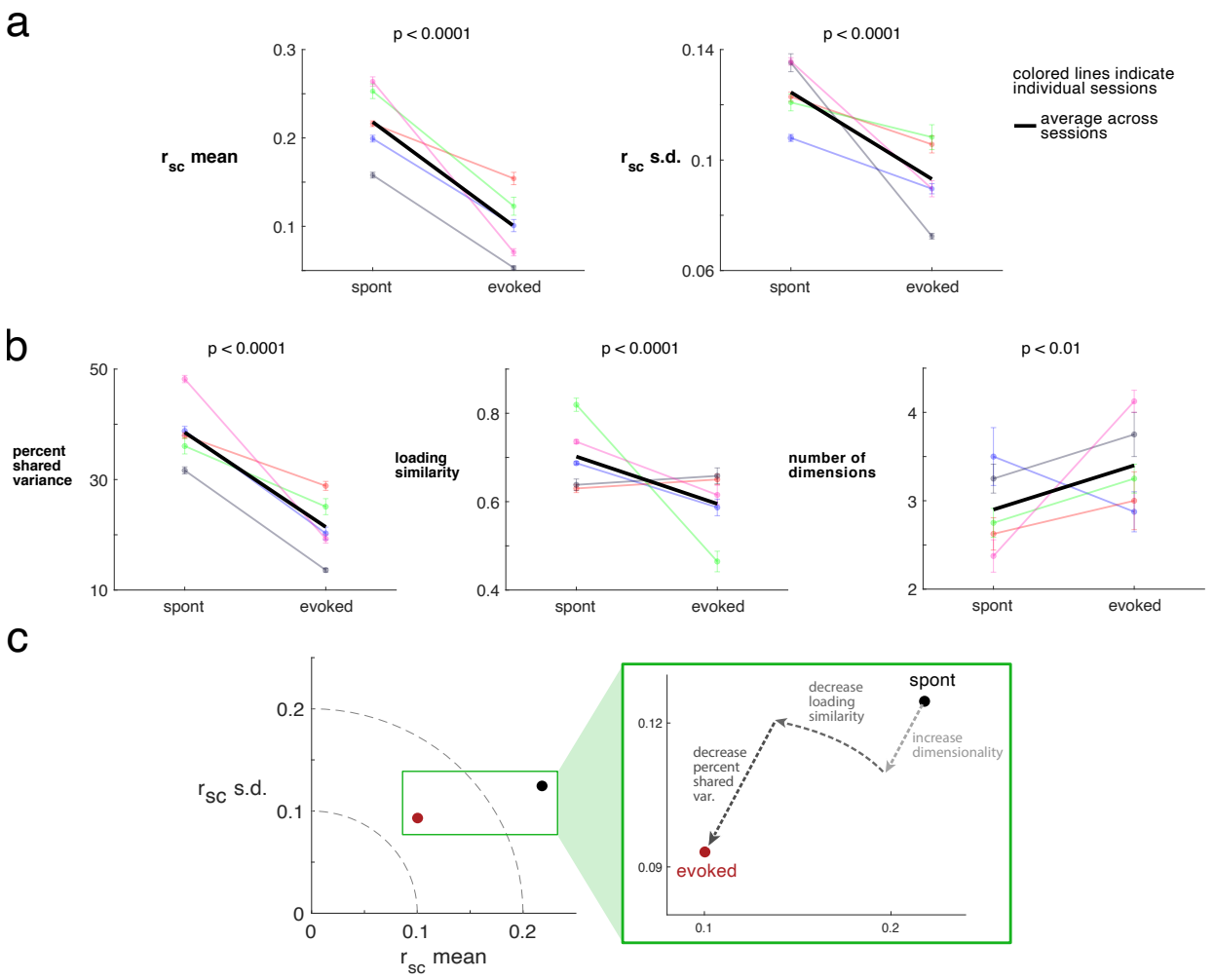
c. Contribution of population metrics to changes in r_{sc} s.d. Same format as **b**. When only %sv was allowed to vary, r_{sc} s.d. decreased with attention. When only loading similarity was allowed to vary, r_{sc} s.d. slightly increased (not significant). Also, when only dimensionality could vary, r_{sc} s.d. increased. When all population metrics were allowed to vary, we found that r_{sc} s.d. decreased with attention, consistent with our computations from data (Fig. 6*b*).

These results provide a systematic quantification of the illustration that relates pairwise and population metrics in V4 (panel **a**). Based on direction and magnitude of contributions, we conclude that for overall changes in pairwise metrics in these data: 1) %sv is most important, 2) followed by loading similarity, 3) followed by dimensionality. More generally, the population metric matching procedure (described below) provides a framework for assessing how changes in population metrics contribute to changes in pairwise metrics in recorded neuronal population activity.

Details of population metric matching procedure. Given two factor analysis (FA) models (e.g., fitted to two different experimental conditions), we first assess the overall change in pairwise metrics by computing r_{sc} mean and s.d. directly from the two fitted models (see Methods). In this case, all three population metrics are allowed to change between the two conditions, and contribute to the overall observed change in pairwise metrics (labeled "all" in the plots above).

Next, we use population metric matching to assess the contribution of each individual population metric change to the overall change in pairwise metrics. To do so, we choose one of the two fitted FA models (e.g., the model fitted to "attend-out") and systematically change the model such that one of its population metrics matches that of the other FA model (e.g., "attend-in"), while the other two population metrics remain the same. We then assessed the change in pairwise metrics between the base FA model (i.e., "attend-out") and the "matched" FA model (i.e., modified "attend-out" model). This allowed us to assess the change in pairwise metrics that would have resulted from a change in a single population metric.

For systematically modifying %sv, we scaled the eigenspectrum (see Methods) of the base FA model in order to match the %sv of the other FA model. For systematically modifying loading similarity, we replaced the co-fluctuation patterns (U in Supplementary Fig. 5*a*) in the base FA model (e.g., "attend-out") with the co-fluctuation patterns from the other FA model (e.g., "attend-in"). In cases where the dimensionality of the two models was different, we swapped the top k co-fluctuation patterns, where k is equal to the smaller dimensionality in the two models. For systematically modifying dimensionality, we removed dimensions from the base FA model if it had higher dimensionality than the other FA model, or added dimensions (after orthogonalization) from the other model to the base FA model if it had lower dimensionality. Because adding or removing dimensions changes the %sv, we then scaled the eigenspectrum to match the original %sv of the base FA model. These procedures allowed us, using two FA models fit to real data, to systematically vary one of the population metrics while keeping the other two the same and assess the contribution to a change in pairwise metrics.



Supplementary Figure 4: Relationship between pairwise and population metrics in V1 population responses. Related to Figure 6.

In Fig. 6, we assessed the relationships between pairwise and population metrics in V4 population recordings where a decrease in r_{sc} mean with spatial attention had been widely reported (Cohen & Maunsell, 2009; Mitchell et al., 2009; Gregoriou et al., 2014; Luo & Maunsell, 2015; Snyder et al., 2018). To demonstrate the applicability of the identified relationships to other brain areas, we applied the same analysis to population recordings in primary visual cortex (V1). Previous studies have shown that the r_{sc} mean is lower after stimulus onset (i.e., evoked activity) than before stimulus onset (i.e., spontaneous activity) in V1 (Smith & Kohn, 2008; Churchland et al., 2010). Here, we analyzed population activity recorded using Utah arrays in V1 (88 to 159 units per session, 112.2 on average) in three macaque monkeys (previously reported in Zandvakili & Kohn, 2015; Semedo et al., 2019, http://dx.doi.org/10.6080/K0B27SHN). Two monkeys had two recording sessions each, while the third monkey had a single recording session, for a total of 5 recording sessions. Animals were presented with 1.28s of oriented gratings (1 of 8 possible orientations) interleaved with 1.5s of a blank screen.

(continued on next page...)

Supplementary Figure 4: (...continued from previous page)

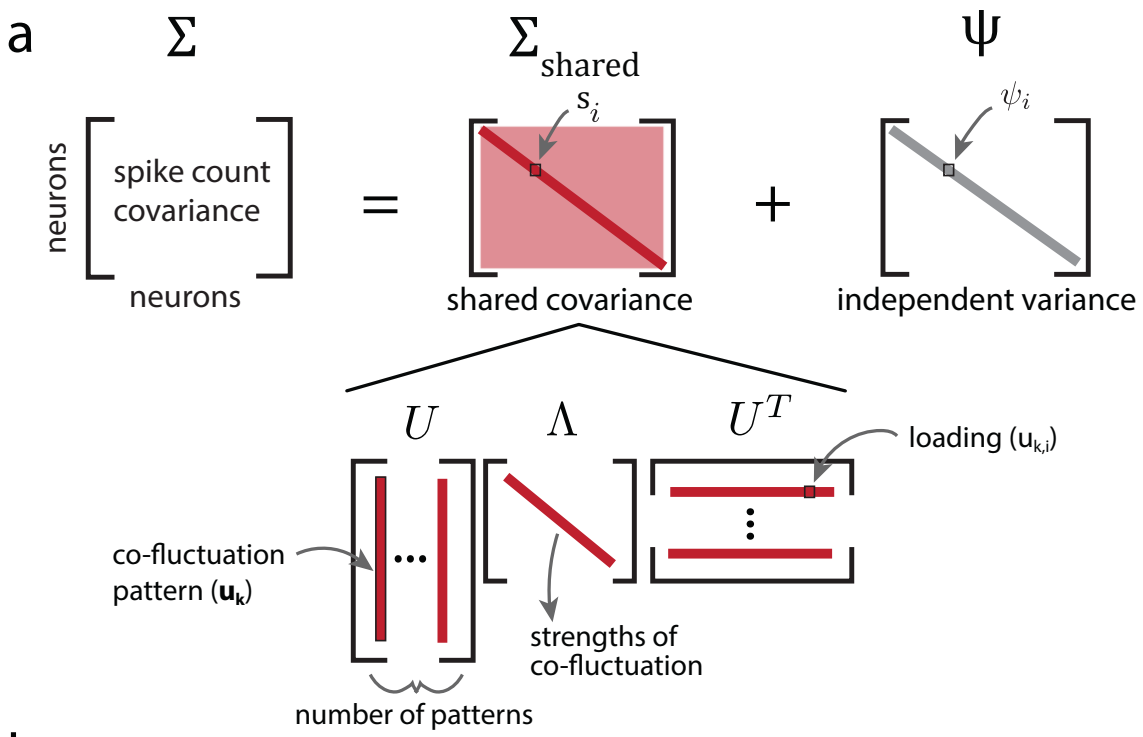
In V1, evoked activity had smaller r_{sc} mean and s.d. than spontaneous activity. We also found that evoked activity had smaller %sv and loading similarity, but larger dimensionality than spontaneous activity. Most changes in pairwise and population metrics between V1 "spontaneous" and "evoked" activity matched the direction of the changes in V4 "attend-out" and "attend-in", *except* for the change in dimensionality (cf. panel **c** and Fig. 6*d*). Taken together, our analyses of V1 and V4 population activity demonstrate that similar changes in pairwise metrics need not correspond to precisely the same changes in population metrics. In this case, measuring population metrics provided insight about the dimensionality of the population-wide variability that would not have been gleaned from changes in pairwise metrics alone.

a. The r_{sc} mean was smaller in evoked activity than in spontaneous activity (left panel; $p < 0.0001$) (Smith & Kohn, 2008; Churchland et al., 2010). We also found that r_{sc} s.d. was smaller in evoked activity than in spontaneous activity (right panel; $p < 0.0001$), which has not been previously reported.

b. Next, we assessed how population metrics changed between evoked and spontaneous V1 activity. Consistent with (Churchland et al., 2010), we found that %sv was smaller for evoked activity than spontaneous activity (left panel; $p < 0.0001$). We also found that loading similarity for the dominant dimension was smaller for evoked activity than spontaneous activity (middle panel; $p < 0.0001$). Finally, we found that dimensionality was higher for evoked activity than spontaneous activity (right panel; $p < 0.01$). This result differed from a previous study in which dimensionality was lower for evoked activity than spontaneous in neural recordings from rat gustatory cortex and in a clustered network model (Mazzucato et al., 2016). This could be explained by a difference in sensory modality or the way in which dimensionality was measured.

c. Using the framework we developed to understand the relationships between pairwise and population metrics (Fig. 5), the decrease in both r_{sc} mean and s.d. with evoked V1 activity corresponds to: 1) a decrease in %sv, 2) a decrease in loading similarity, and 3) an increase in dimensionality. The direction of the changes in pairwise metrics between spontaneous and evoked activity are the same as those we observed between "attend-out" and "attend-in" in V4 (Fig. 6*b*), as are the changes in %sv and loading similarity population metrics (Fig. 6*c*). However, the increase in dimensionality from V1 spontaneous to evoked is in the opposite of what we observed from "attend-out" to "attend-in" in V4 (Fig. 6*c*, right panel).

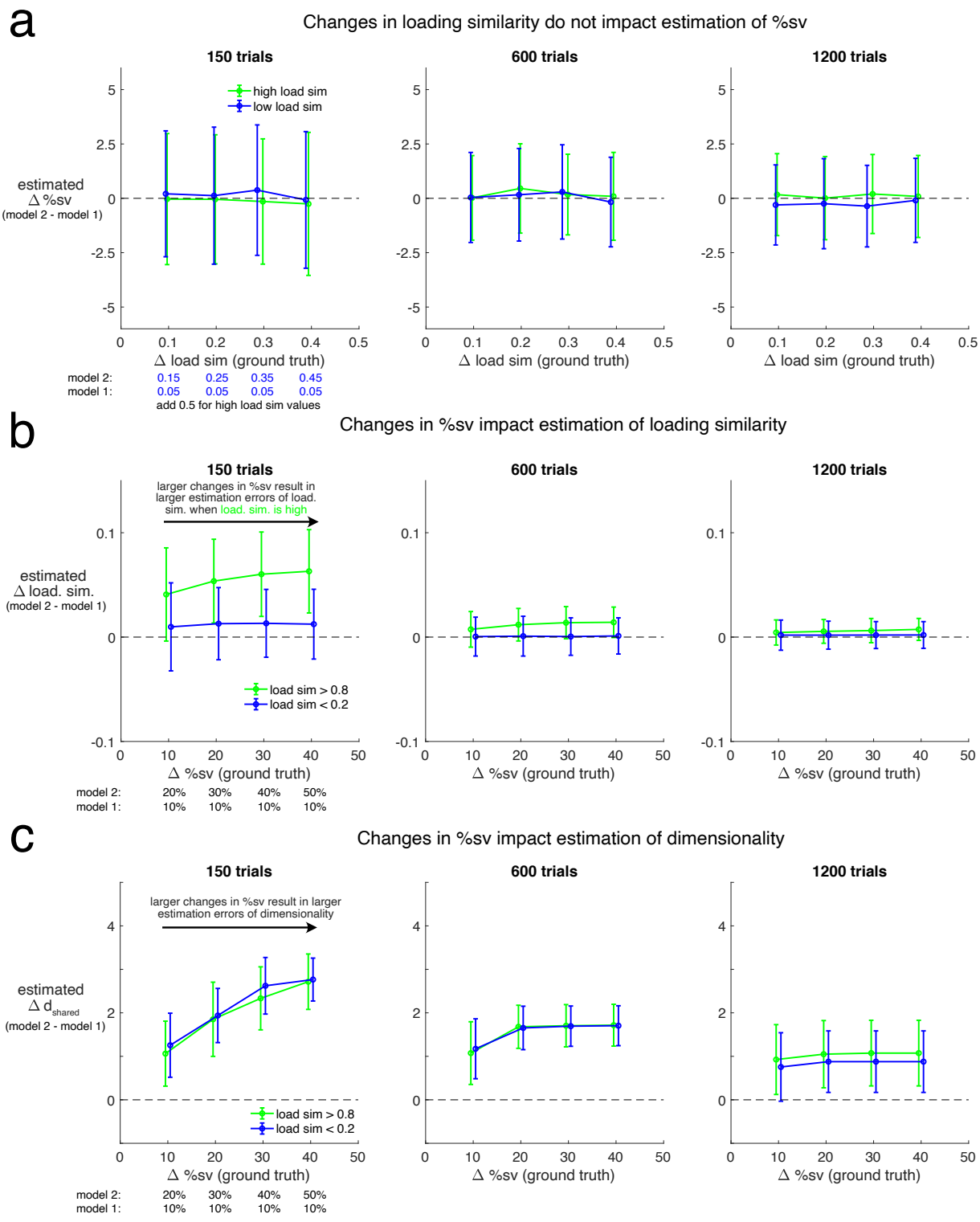
Methods. For evoked activity, we computed spike counts for each trial in the time period 160-260 ms after stimulus onset. For spontaneous activity, we computed spike counts during the blank screen in the 100 ms immediately prior to stimulus onset. We chose to use 100 ms bin sizes to match those used in Semedo et al. (2019). We define spike counts during these two time periods during a trial as a "spont-evoked pair". Each recording session consisted of 400 repeats of a spont-evoked pair for each of the 8 oriented stimuli. For each session, we assessed changes in metrics for each orientation and computed the mean and standard error of the metric across the 8 orientations (transparent colored data points connected by lines). We also plot the average across the 5 sessions (thick black line). To compare metrics for spontaneous and evoked activity, we computed p-values across all 40 datasets (5 sessions, 8 orientations per session) using a paired t-test.



b

population metric	mathematical definition	intuition
loading similarity	$\text{loading sim.} = 1 - \frac{\text{var}(\mathbf{u_k})}{1\,/\,\text{n}}$	similarity in weights in co-fluctuation pattern $\mathbf{u_k}$
percent shared variance	$\%\text{sv} = \frac{1}{\text{n}} \sum_{i=1}^{\text{n}} \frac{s_i}{s_i+\psi_i} \cdot 100\%$	% of neuron's variability explained by other neurons in the population
dimensionality	$\text{dim.} = \text{rank}(U)$	number of co-fluctuation patterns

Supplementary Figure 5: Decomposition of the spike count covariance matrix and defining population metrics. Related to Figures 1 and 2, and STAR Methods. a. We use factor analysis to decompose the spike count covariance matrix Σ into the sum of a low-rank shared covariance matrix Σ_{shared} and a diagonal independent variance matrix Ψ. The ith diagonal entry of Σ_{shared} (s_i) corresponds to the spike count variance that neuron i shares with other neurons in the population (i.e., shared variance), while the ith diagonal entry of Ψ_i corresponds to spike count variance of neuron i that cannot be explained by the other neurons (i.e., independent to neuron i). We can further decompose Σ_{shared} via an eigendecomposition to extract the co-fluctuation patterns (i.e., the eigenvectors) and the strength of each latent co-fluctuation (i.e., the eigenvalues). **b**. The population metrics used in this study are loading similarity, percent shared variance (%sv), and dimensionality.



Supplementary Figure 6: Characterizing how changes in one population metric can impact the estimates of another population metric. Related to Figure 6.

(continued on next page...)

Supplementary Figure 6: (...continued from previous page)

In the main text, we related population metrics to pairwise metrics by systematically changing a single population metric and measuring the resulting changes in r_{sc} mean and r_{sc} s.d. (Figs. 3 and 5). However, in real neuronal data, multiple population metrics could change together between experimental conditions (e.g., see Fig. 6 and Supplementary Fig. 4). When we measure that multiple population metrics changed, it could be the case that a change in one population metric impacted the estimates of the other population metrics (e.g., we could have measured a change in multiple population metrics when only one metric truly changed). This can affect the precision by which we can distinguish population metric changes in real neuronal data.

Here, we assessed this by systematically changing one population metric while keeping the other two population metrics constant. We then simulated data and fit factor analysis (FA) to the data to obtain the population metrics. We examined in turn each of the three population metrics under conditions when they did not actually change (but one of the other metrics did). If there were no dependencies between estimates of population metrics, then all the vertical values in panels **a**-**c** would be 0. We found that this was the case for estimates of %sv were under conditions in which the true loading similarity changed (**a**). However, estimates of loading similarity and dimensionality were affected by true changes in %sv. Increasing the number of simulated trials reduced the estimation error caused by true changes in %sv (**b**, **c**). These findings allow us to better interpret changes in population metrics estimated from neuronal activity.

a. Estimation error in %sv due to changes in loading similarity. "Model 1" and "model 2" had the same dimensionality (1) and %sv (20%). The only difference between the two models was their loading similarity. We varied how different the loading similarity was between the two models (horizontal axis), while assessing how different was the estimated %sv across the two models (vertical axis). We found that %sv estimates remained unaffected in the presence of true changes in loading similarity (all changes in %sv are near 0). This was true for both low loading similarities where "model 1" had loading similarity of 0.05 (blue) and high loading similarities where "model 1" had loading similarity of 0.55 (green). As we simulated more trials, estimates of %sv became more precise (error bars decrease in size going from left to right panels). Error bars show means and standard deviations across simulations.

b. Estimation error in loading similarity due to changes in %sv. "Model 1" and "model 2" had the same dimensionality (1) and loading similarity. The only difference between the two models was their %sv. We varied how different the %sv was between the two models (horizontal axis), while assessing how different was the estimated loading similarity between the two models (vertical axis). We found little changes in estimates of loading similarity when the true loading similarity was low (blue points). However, we found larger changes in estimates of loading similarity when the true loading similarity was high (green points). The the size of the change increased with larger true changes in %sv. To understand this, recall that there are relatively few ways to have high loading similarity (e.g., all loadings must be the same to have loading similarity of 1), while there are many ways to have low loading similarity (Math Note E). Thus, high loading similarities are more likely to be underestimated than low loading similarities. This underestimate tends to be larger when %sv is low than when %sv is high. However, increasing the trial counts reduced the estimation error of loading similarity (vertical values closer to 0 going from left to right panels).

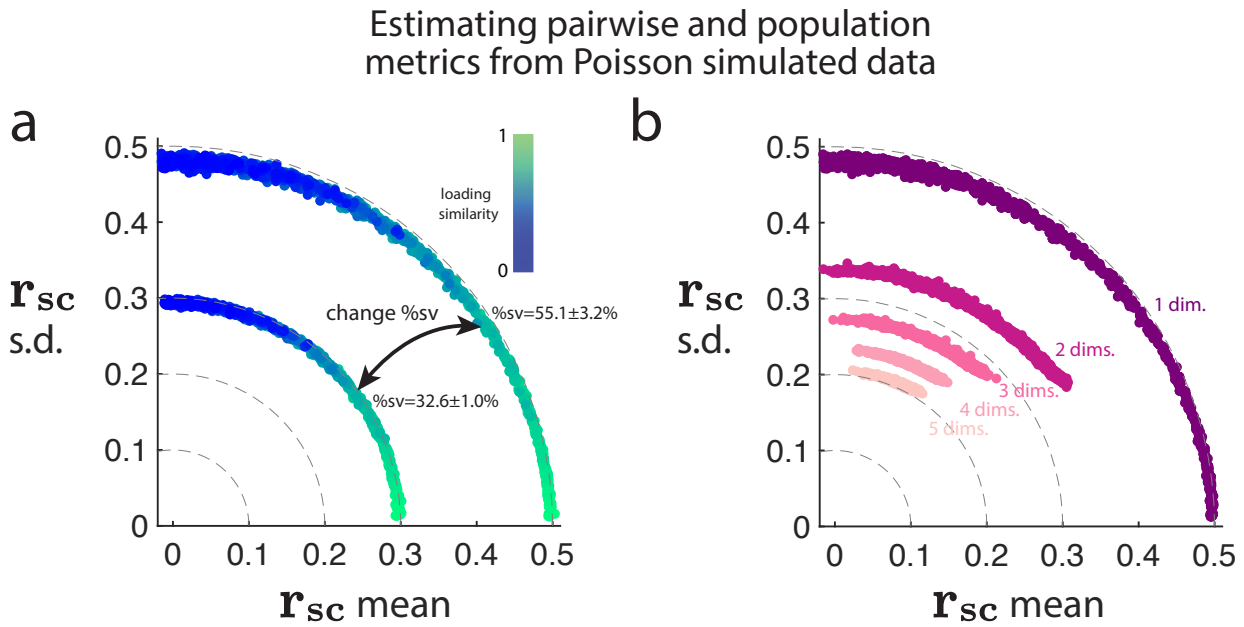
(...continued on next page)

Supplementary Figure 6: (...continued from previous page)

c. Estimation error in dimensionality due to changes in %sv. "Model 1" and "model 2" had the same dimensionality (5; with eigenspectrum defined as $\lambda_k = e^{-0.75k}$) and loading similarity. The only difference between the two models was their %sv. We varied how different the %sv was between the two models (horizontal axis), while assessing how different was the estimated dimensionality (d_shared; see Methods) between the two models (vertical axis). We found changes in the estimates of dimensionality between "model 1" and "model 2", and the size of the change increased with larger true changes in %sv. This can be understood by the fact that dimensions with small eigenvalues can be difficult to recover when fitting FA to data, particularly when %sv is low. However, increasing the trial counts reduced the estimation error of dimensionality (vertical values closer to 0 going from left to right panels).

These results have important implications for interpreting estimated changes in population metrics in real neuronal data. First, because estimation error depends on trial count, one should equalize the number of trials across conditions in order to make fair comparisons across conditions using population metrics. Second, when changes in %sv are large and trial counts are small, one may need to be careful in interpreting estimated changes in loading similarity and dimensionality. For trial count, the key quantity to consider is the ratio of observed trials to the number of recorded neurons (Wainwright, 2019). In the simulations above, we used 30 neurons–the left column represented a ratio of 5x trials to neurons, the middle column represented 10x, and the right column represented 20x.

In our V4 data (Fig. 6), most sessions had 10 times (or more) the number of trials as the number of neurons (ratio of trials to neurons: 9.90 ± 0.66 for monkey 1, 27.60 ± 2.68 for monkey 2). We observed a difference in %sv of $\approx$3% between "attend-out" and "attend-in". Based on the results in panels **a** and **b**, the differences we measured in %sv and loading similarity in our V4 data are unlikely to be due to estimation error. Based on panel **c**, the small difference we measured in dimensionality in our V4 data could potentially be explained by a change in %sv, if the only true change between conditions was in %sv (and not loading similarity or any other aspect of the population activity).



Supplementary Figure 7: Relationships between pairwise and population metrics hold for metrics estimated from Poisson simulated data. Related to Figure 3.

In our simulations and analytical derivations, we created covariance matrices with specified population metrics from which we computed pairwise metrics (Fig. 3). However, when assessing population metrics in neuronal recordings, one needs to fit a factor analysis (FA) model to data. Here, we simulated Poisson data to assess whether the relationships between pairwise and population metrics were impacted by: 1) needing to estimate metrics from data, and 2) the mismatch between the linear-Gaussian assumption of FA and the Poisson-like statistics of neuronal activity. We found that the relationships between pairwise and population metrics were very similar to those shown in Figs. 3 and 5.

a. Estimating loading similarity and %sv. We simulated data from a model with a single co-fluctuation pattern and Poisson observations (see details at the end of the caption). In the ground truth models, we varied loading similarity smoothly between 0 and 1 and chose %sv equal to 30% or 50%. We estimated r_{sc} mean and s.d. from the simulated data. To estimate population metrics, we fit the FA parameters to the same simulated data. We then plotted estimates of pairwise metrics and colored or labeled points according to the *estimated* population metrics (as opposed to the ground truth population metrics used to generate the data). We found that as estimated loading similarity increased, r_{sc} mean increased and r_{sc} s.d. decreased (blue to green). We also found that as estimated %sv increased, r_{sc} mean and s.d. both increased (inner arc with %sv=32.6±1.0% to outer arc with %sv=55.1±3.2%). These results are consistent with Fig. 3*e-f*.

b. Same as **a**, but fixing %sv=50% and varying the dimensionality of the ground truth model, with a flat eigenspectrum (corresponding to Fig. 3*g*). We colored points according to *estimated* dimensionality as opposed to the ground truth dimensionality. We found that as the estimated dimensionality increased, r_{sc} mean and s.d. both tended to decrease (purple outer arc with dim=1 to salmon inner arc with dim=5), consistent with Fig. 3**g**.

(continued on next page...)

Supplementary Figure 7: (...continued from previous page)

The results here, based on estimating factor analysis parameters from Poisson simulated data, are qualitatively the same as those in the main text (Fig. 3*e-g*) and analytical derivations (Appendices). This indicates that the relationships between pairwise and population metrics are robust to: 1) having to estimate these metrics from data and 2) the Poisson-like variability of neuronal activity.

Simulating from a Poisson observation model. According to FA, the observations x (i.e., spike counts) have a linear-Gaussian relationship with latent variables z (which represent shared activity among neurons; see Fig. 2): $\mathbf{z} \sim N(0, I)$ and $\mathbf{x}|\mathbf{z} \sim N(L\mathbf{z}+\mu, \Psi)$). We fit the FA parameters to data simulated from a Poisson observation model. We generated Poisson spike counts for 30 neurons as follows. For neuron i, we sample from from $x_i|\mathbf{z} \sim Poisson(ReLu(L_{i,:}\mathbf{z} + \mu_i))$, where $ReLu$ indicates a rectified linear unit, and $L \in R^{30\times d}$ is the loading matrix with $L_{i,:}$ as the i^{th} row. We set $\mu_i = 10$ for each neuron, a typical average firing rate (10 Hz) for neurons across many areas of macaque cortex (assuming a 1 second time bin). We consider the asymptotic case by simulating many trials for each model (corresponding to a single dot in panels *a* and *b*; see Methods for how model parameters are randomly chosen). We consider estimation from limited data in Supplementary Fig. 6. We drew 6000 samples (i.e., 6000 trials) of $\mathbf{x}$ from the Poisson observation model. Thus, this procedure generated a data matrix $X \in R^{30\times 6000}$ of simulated spike counts, which we then used to estimate pairwise and population metrics.