Title

Diverse coactive neurons encode stimulus-driven and stimulus-independent variables

Journal

Journal of Neurophysiology, 124(5)

ISSN

0022-3077

Authors

Xia, Ji
Marks, Tyler D
Goard, Michael J
et al.

Publication Date

2020-11-01

DOI

10.1152/jn.00431.2020

Peer reviewed

Title: Diverse co-active neurons encode stimulus-driven and stimulus-independent variables.

Authors:

Ji Xia[1], Tyler D. Marks[2], Michael J. Goard[2,3,4*], Ralf Wessel[1*]

1 Department of Physics, Washington University in St. Louis

2 Neuroscience Research Institute, University of California, Santa Barbara

3 Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara

4 Department of Psychological & Brain Sciences, University of California, Santa Barbara

* These authors contributed equally

Correspondence and requests for materials should be addressed to Ji Xia (email: xiaji@wustl.edu)

Abstract

Both experimenter-controlled stimuli and stimulus-independent variables impact cortical neural activity. A major hurdle to understanding neural representation is distinguishing between qualitatively different causes of the fluctuating population activity. We applied an unsupervised low-rank tensor decomposition analysis to the recorded population activity in the visual cortex of awake mice in response to repeated presentations of naturalistic visual stimuli. We found that neurons co-varied largely independently of individual neuron stimulus response reliability and thus encoded both stimulus-driven and stimulus-independent variables. Importantly, a neuron's response reliability and the neuronal

coactivation patterns substantially reorganized for different external visual inputs. Analysis of recurrent balanced neural network models revealed that both the stimulus specificity and the mixed encoding of qualitatively different variables can arise from clustered external inputs. These results establish that co-active neurons with diverse response reliability mediate a mixed representation of stimulus-driven and stimulus-independent variables in the visual cortex.

Introduction

Neural variability is a key feature of neocortical neuronal responses. During repeated sensory stimulation, most neurons exhibit high trial-to-trial variability, while only a small number of neurons display reliable responses across trials (Softky and Koch 1993; Stringer et al. 2019a). The abundance of unreliable neurons in the cerebral cortex raises the question to what extent these neurons contribute to the representation of stimulus-driven and stimulus-independent variables (Olshausen and Field 2006). Possible answers to this question arise from multiple sources. First, neural variability is correlated across neurons (Cohen and Kohn 2011) such that untuned/unreliable neurons enhance sensory information coding (Leavitt et al. 2017; Safaai et al. 2013). Second, sensory cortex not only encodes stimuli, but also encodes behavioral variables (Dipoppa et al. 2018; McGinley et al. 2015; Niell and Stryker 2010; Stringer et al. 2019b; Vinck et al. 2015) or internal state variables (Allen et al. 2019; Vinck et al. 2015). Thus, neural response variability to sensory stimuli can be partially explained by experimentally observed stimulus-independent variables (Stringer et al. 2019b). These observations suggest that unreliable neurons may play a role in encoding both stimulus-driven and stimulus-

independent unobserved variables.

There is a growing consensus in neuroscience that co-active ensembles of neurons, as opposed to single neurons, are the underpinning of cognition and behavior (Buzsáki 2010; Saxena and Cunningham 2019; Yuste 2015). How then do neurons in sensory cortex co-vary and encode stimulus-driven or stimulus-independent variables? We assume that single-trial neuronal responses consist of additive modulations of distinct latent factors. Furthermore, each latent factor is modulated by the gain specific to the neuron and the trial (Fig. 1A). The question of encoding stimulus-driven and stimulus-independent variables is usefully illustrated by considering the extreme ends of a spectrum of possibilities (Fig. 1B). At one extreme, reliable neurons co-vary and encode stimulus-driven variables, while unreliable neurons co-vary and encode stimulus-independent variables. At the other extreme,neurons covary and encode both stimulus-driven and -independent variables regardless of their reliability. Identifying where along this spectrum cortical encoding operates is fundamentally challenging because the stimulus-driven and the stimulus-independent variables are unobserved (Keemink & Machens, 2019). These unobserved variables must be inferred from observed neuronal population activity, which, however, is highly variable across trials of repeated stimulus presentation. Supervised methods, such as demixed principal component analysis (Kobak et al. 2016) and targeted dimensionality reduction (Mante et al. 2013) can only partially solve this problem by inferring unobserved variables that are correlated to observed behavioral or task-related variables. A promising direction is to solve the problem using unsupervised methods, as shown by recent works in visual cortex (Stringer et al. 2019b) and frontal cortex (Hirokawa et al. 2019). Here we employed an unsupervised method, tensor component analysis

(TCA) (Williams et al. 2018), which allowed us to identify stimulus-driven and stimulus-independent unobserved variables in an unbiased fashion from observed neuronal population activity in response to repeated stimulus presentations.

We performed two-photon calcium imaging of excitatory neurons in the primary visual cortex of awake, head-fixed mice during visual stimulation with repeated identical naturalistic movie clips (Nat Mov) or periodic drifting gratings (PDG). We identified unobserved variables, or "latent factors", representing either stimulus-driven variables or stimulus-independent variables. Our results show that neurons with a range of reliability co-vary and encode both stimulus-driven and stimulus-independent variables. Moreover, we found that the neuronal coactivation pattern is randomly redistributed across different stimuli. This suggests that feedforward inputs to neurons in visual cortex have a significant influence on neuronal coactivation patterns. Finally, simulation of a neural network model revealed possible input structures underlying the observed encoding paradigm in visual cortex.

Materials and Methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Ji Xia (xiaji@wustl.edu). We used tools for fitting TCA in https://github.com/ahwillia/tensortools.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

For imaging visual cortical responses, a Emx1-Cre (Jax Stock #005628) x ROSA-LNL-tTA (Jax Stock #011008) x TITL-GCaMP6s (Jax Stock #024104) triple transgenic mouse line (*n* = 7) was bred to express GCaMP6s in cortical excitatory neurons (Madisen et al. 2015). Mice ranging in age from 6 - 20 weeks of both sexes (3 males and 4 females) were implanted with a head plate and cranial window and imaged starting >2 weeks after recovery from surgical procedures and up to 10 months after window implantation. The animals were housed on a 12 hr light/dark cycle in cages of up to 5 animals before the implants, and individually after the implants. All animal procedures were approved by the Institutional Animal Care and Use Committee at University of California, Santa Barbara.

Surgical procedures

All surgeries were conducted under isoflurane anesthesia (3.5% induction, 1.5 - 2.5% maintenance). Prior to incision, the scalp was infiltrated with lidocaine (5 mg/kg, subcutaneous) for analgesia and meloxicam (1 mg/kg, subcutaneous) was administered preoperatively to reduce inflammation. Once anesthetized, the scalp overlying the dorsal skull was sanitized and removed. The periosteum was removed with a scalpel and the skull was abraded with a drill burr to improve adhesion of dental acrylic. A 4 mm craniotomy was made over the visual cortex (centered at 4.0 mm posterior, 2.5 mm lateral to Bregma), leaving the dura intact. A cranial window was implanted over the craniotomy and sealed first with silicon elastomer (Kwik-Sil, World Precision Instruments) then with dental acrylic (C&B-Metabond, Parkell) mixed with black ink to reduce light transmission. The cranial windows were made of two rounded pieces of coverglass (Warner Instruments) bonded with a UV-cured optical adhesive (Norland, NOA61). The bottom coverglass (4 mm) fit

tightly inside the craniotomy while the top coverglass (5 mm) was bonded to the skull using dental acrylic. A custom-designed stainless steel head plate (eMachineShop.com) was then affixed using dental acrylic. After surgery, mice were administered carprofen (5-10 mg/kg, oral) every 24 hr for 3 days to reduce inflammation. The full specifications and designs for head fixation hardware can be found on the Goard lab website (https://goard.mcdb.ucsb.edu/resources).

Two-photon imaging

After >2 weeks' recovery from surgery, GCaMP6s fluorescence was imaged using a Prairie Investigator 2-photon microscopy system with a resonant galvo scanning module (Bruker). For fluorescence excitation, we used a Ti:Sapphire laser (Mai-Tai eHP, Newport) with dispersion compensation (Deep See, Newport) tuned to λ = 920 nm. For collection, we used GaAsP photomultiplier tubes (Hamamatsu). We used a 16x/0.8 NA microscope objective (Nikon) at 1x or 2x magnification, obtaining a square field of view with width ranging from 414 to 828 μm. Laser power ranged from 40–75 mW at the sample depending on GCaMP6s expression levels. Photobleaching was minimal (<1%/min) for all laser powers used. A custom stainless-steel light blocker (https://goard.mcdb.ucsb.edu/resources) was mounted to the head plate and interlocked with a tube around the objective to prevent light from the visual stimulus monitor from reaching the PMTs. During imaging experiments, the polypropylene tube supporting the mouse was suspended from the behavior platform with high tension springs to reduce movement artifacts.

2-Photon Post-processing

Images were acquired using PrairieView acquisition software and converted into TIF files. All subsequent analyses were performed in MATLAB (Mathworks) using custom code (https://goard.mcdb.ucsb.edu/resources). First, images were corrected for X-Y movement by registration to a reference image (the pixel-wise mean of all frames) using 2-dimensional cross correlation.

To identify responsive neural somata, a pixel-wise activity map was calculated using a modified kurtosis measure. Neuron cell bodies were identified using local adaptive threshold and iterative segmentation. Automatically defined ROIs were then manually checked for proper segmentation in a graphical user interface (allowing comparison to raw fluorescence and activity map images). To ensure that the response of individual neurons was not due to local neuropil contamination of somatic signals, a corrected fluorescence measure was estimated according to:

$$F_{corrected}(n) = F_{soma}(n) - \alpha * F_{neuropil}(n)$$

where $F_{neuropil}$ was defined as the fluorescence in the region <30 μm from the ROI border (excluding other ROIs) for frame n and α was chosen from [0 1] to minimize the Pearson's correlation coefficient between $F_{corrected}$ and $F_{neuropil}$. The $\Delta F/F$ for each neuron was then calculated as:

$$\Delta F/F = (F_n - F_0) / F_0$$

Where F_n is the corrected fluorescence ($F_{corrected}$) for frame n and F_0 defined as the mode of the corrected fluorescence density distribution across the entire time series.

Visual stimuli

All visual stimuli were generated with a Windows PC using MATLAB and the Psychophysics toolbox (Brainard 1997). Stimuli used for two-photon imaging were presented on an LCD monitor (17.5 x 13 cm, 800 x 600 pixels, 60 Hz refresh rate) positioned 5 cm from the eye at a horizontal tilt of 30 deg to the right of the midline and vertical tilt of 18 deg downward, spanning 120 deg (azimuth) by 100 deg (elevation) of visual space in the right eye.

For drifting grating visual stimulation, 12 full-contrast sine wave gratings (spatial frequency: 0.05 cycles/deg; temporal frequency: 2 Hz) were presented full-field, ranging from 0 to 330 deg in 30 deg increments. We presented 8 repeats of the drifting grating stimulus; a single repeat of stimulus consisted of all 12 grating directions presented in order for 2 sec with a 4 sec inter-stimulus interval (gray screen).

For natural movie visual stimulation, we displayed a grayscale 30 sec clip from *Touch of Evil* (Orson Wells, Universal Pictures, 1958) containing a continuous visual scene with no cuts (https://observatory.brain-map.org/visualcoding/stimulus/natural_movies). The clip was contrast-normalized and presented at 30 frames per second. We presented 30 repeats of the natural movie stimulus; each repeat started with 5 sec of gray screen, followed by the 30 sec of movie.

When we compared neural responses across stimuli, we did analyses on part of the responses so that their trial structure matches. For Nat Mov, we took the first 240 time

points after movie onset and the first 8 trials of the responses. For PDG, we took concatenated neural responses during PDG without the gray screen periods to get 240 time points (20 time points × 12 orientations). Thus, two types of neural responses would have the same trial structure (240 time points × 8 trials).

Nonnegative Tensor Decomposition with missing data

We organized our data into a 3-way tensor χ ($N \times T \times K$) and let x_{ntk} represent the activity of neuron n at time t and trial k. Nonnegative tensor component analysis (TCA) decomposes χ into a sum of R rank-one tensors, where each rank-one tensor can be written as an outer product of 3 nonnegative vectors:

$$x_{ntk} \approx \sum_{r=1}^{R} w_b^r b_t^r a_k^t = \hat{x}_{ntk}$$

Nonnegative TCA with missing values were fit to minimize the squared reconstruction error:

$$|| M \star (\chi - \hat{\chi}) ||_F^2 \text{ while } W \geq 0, B \geq 0, A \geq 0$$

Here, $\hat{\chi}$ denotes the reconstructed data. $|| \cdot ||_F^2$ denotes the squared Frobenius norm of a tensor:

$$|| \chi ||_F^2 = \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} x_{ntk}^2$$

M denotes a masking tensor with the same shape as χ , and $\star$ denotes entrywise

multiplication of two tensors. For fitting nonnegative TCA on $\Delta F/F$ data, we set $m_{ntk} = 0$ if $x_{ntk} < 0$, otherwise we set $m_{ntk} = 1$. Normalized reconstruction error is the squared reconstruction error normalized by $|| M \star \chi ||_F^2$.

Different from matrix decompositions, tensor decompositions are often unique (Kruskal 1977). However, when R is large or W, B, A have low rank, it could be difficult to optimize. To monitor this possibility, we calculated similarity between different TCA fitting results on the same dataset as described in (Williams et al. 2018). We found that the similarity between fitting results is close to 1 for all the nonnegative TCA models reported in this work.

Preprocessing of $\Delta F/F$ data

$\Delta F/F$ data were normalized such that the averaged squared sum of $\Delta F/F$ traces over time equals to 1 for every neuron:

$$\sqrt{(\sum_{tk} x_{ntk}^2)/TK} = 1$$

This normalization step is crucial for ensuring TCA fitting is not biased by high firing rate neurons, since TCA is optimized to minimize the squared reconstruction error.

Choice of the number of components in TCA

We picked the number of TCA components such that they captured a significant amount of neural responses without over-fitting, checked with cross-validation as previously reported (Williams et al. 2018). To perform cross-validation, we randomly masked out 50%

of tensor entries in χ. The remaining data was training set and the masked-out data was test set. We trained nonnegative TCA with missing values to fit the training set. And then we used the trained TCA model to fit the test set. As we increase the number of components in TCA, if the normalized reconstruction error of the test set went up, the TCA model would overfit the training set. As previously reported (Williams et al. 2018), TCA is unlikely to overfit, even with up to 60 components. For this paper, we chose 20 components for TCA, given that 20 component TCA captured a significant amount of neural responses without over-fitting. Note that all the results in this paper were robust to changes in the number of TCA components (data not shown; we tested TCA with 10 to 40 components).

Balance network model

Neurons were modeled as binary units. We simulated 1600 excitatory neurons and 400 inhibitory neurons. The spiking s_i^x of neuron i in population x ∈ {E,I} was given by

$$s_i^x(t) = \Theta\left(\sum_{j=1}^{2000} J_{ij}s_j + \mu^x + \sum_{m=1}^{20} g_m \times K_{im} \times \eta_m + \sum_{n=1}^{20} L_{in} \times \xi_n - \theta^x\right)$$

Θ is the Heaviside step function. J_{ij} is the connectivity weight from neuron j to neuron i. Each neuron received on average 200 excitatory and 200 inhibitory recurrent inputs, thus most matrix elements J_{ij} were zero. For the non-zero matrix elements J_{ij}, the synaptic weights were $J^{EE} = J^{IE} = 0.07$; $J^{EI} = -0.14$; $J^{II} = -0.13$. Bias current was given by $\mu^E = 1.13$; $\mu^I = 0.91$. Spiking threshold was given by $\theta^E = 1$; $\theta^I = 0.7$. Choices of parameters are motivated by previous work in the balanced network (Litwin-Kumar and Doiron 2012; van Vreeswijk and Sompolinsky 1998b).

Frozen input pulse trains η consisted of 20 pulse trains repeated over trials, thus imitating the stimulus-driven variables (Supplemental Fig. S4D (https://figshare.com/s/576fe4fa7850f8cecde5)). On each trial, each frozen pulse train contained one burst of 3 pulses during a random located time window of 200 ms. (Supplemental Fig. S4D (https://figshare.com/s/576fe4fa7850f8cecde5)). Another set of 20 different input pulse trains ξ varied across trials, thus imitating stimulus-independent variables (Supplemental Fig. S4E (https://figshare.com/s/576fe4fa7850f8cecde5)). Since stimulus-independent variables are not locked to the trial structure, we generated trial-varied input pulse trains as Poisson pulse trains with a rate of 0.005 Hz during 500 sec, i.e., the duration of the simulation.

K is a 2000x20 matrix, describing synaptic weights between frozen input pulse trains and individual neurons. Each neuron only received one frozen pulse train, and each frozen pulse train innervated 100 neurons. The nonzero entries of K followed a lognormal distribution with mean = 2 (Supplemental Fig. S4B (https://figshare.com/s/576fe4fa7850f8cecde5)). g is a constant gain factor varying from trial to trial, randomly selected from a uniform distribution $U(0.3, 0.8)$. L is a 2000x20 matrix, describing synaptic weights between trial-varied input pulse trains and individual neurons. Each neuron only received one trial-varied pulse train and each trial-varied pulse train innervated 100 neurons. Similar to K, the nonzero entries of L followed a lognormal distribution (Supplemental Fig. S4B (https://figshare.com/s/576fe4fa7850f8cecde5)). Both, (i) the burst-like temporal structure of the input pulse trains and (ii) the clusters of neurons with identical input pulse trains were chosen to impose a level of coordinated spiking within the otherwise unstructured recurrent model neural network.

To simulate neural responses to two different stimuli, we generated two sets of frozen input pulse trains and trial-varied input pulse trains as well as the corresponding input synaptic weights independently with the same statistics as described above.

Simulations were performed with a discrete time step of 10 ms and neurons are updated asynchronously with a fixed order. At the beginning of each trial, 20% of neurons were randomly selected to be active, with the rest of neurons being silent. We simulated 20 trials for each stimulus. Each trial was simulated for 25 sec. We convolved the simulated spike train with a kernel $e^{-t/\tau_2} - e^{-t/\tau_1}$ similar to GCaMP6s kernel to generate simulated $\Delta F/F$ traces (rise time τ_1 = 100 ms, decay time τ_2 = 2 s). TCA was fitted on subsampled simulated $\Delta F/F$ traces with a time resolution of 100 ms.

QUANTIFICATION AND STATISTICAL ANALYSIS

Correlation between reliability of co-active neuron pairs

To investigate dependency on reliability for neuronal coactivation, we calculated the Pearson correlation between reliability of significantly positively correlated neuron pairs in all recorded imaging fields. To select those neuron pairs, we calculated the Pearson correlation between pairs of neuronal responses and picked neuron pairs with positive and significant ($p < 0.001$) correlations.

Ordering of TCA components

For analysis on responses during Nat Mov, TCA components were ordered by their consistency over trials. The consistency of TCA components was quantified as coefficient of variation (CV) of their trial factors.

For analysis on concatenated responses, TCA components were first separated into two groups based on whether the sum of trial factors during first 8 trials (during PDG stimulation) was higher than sum of trial factors during second 8 trials (during Nat Mov stimulation). Then, within each group, TCA components were ordered by their consistency.

Sorting neurons by dominant components

Neurons were reordered by their dominant components. There were two steps for this sorting method. First, we grouped neurons by their dominant component. Dominant component was defined as the component with the highest neuron factor value for a given neuron. Second, within each group of neurons with the same dominant component, we sorted neurons by their neuron factor values of the dominant component in descending order.

Fitting performance

We used the coefficient of determination (R^2) to quantify the fitting performance of reconstructed responses by TCA components. Before we calculated R^2 between normalized $\Delta F/F$ traces and reconstructed $\Delta F/F$ traces, we set the negative part of normalized $\Delta F/F$ and corresponding part of reconstructed $\Delta F/F$ traces to zero.

Response reliability

Response reliability was defined as the correlation coefficient of neural responses between pairs of trials averaged over all trial pairs for a given neuron:

$$Reliability_i = \frac{2}{K(K-1)} \sum_{k_1=1}^{K} \sum_{k_2=k_1+1}^{K} \frac{Cov\,(r_{k_1}, r_{k_2})}{\sqrt{Var(r_{k_1})\,Var(r_{k_2})}}$$

Results

Response reliability has a skewed distribution.

We recorded from layer 2/3 pyramidal neurons in V1 of awake, head-fixed mice using two-photon calcium imaging of transgenic mice expressing the calcium indicator GCaMP6s in excitatory neurons (see Methods) (Fig. 1C, D). Mice watched a repeated clip of a 30 sec naturalistic movie for 30 trials while being constrained within a tube (see Methods). We recorded from 10 imaging fields in 7 mice and extracted calcium responses ($\Delta F/F$) from a total of 4077 well-isolated somatic regions of interest (ROIs). Neuronal responses varied across trials. Using previously described methods (Goard and Dan 2009; Rikhye and Sur 2015), we quantified this response variation in terms of the "response reliability", defined as the correlation coefficient of neural responses between pairs of trials averaged over all trial pairs for a given neuron (Fig. 1E; see Methods). Response reliability distributions were skewed, with most neurons exhibiting low response reliability (Fig. 1F; Supplemental Fig. S1A (https://figshare.com/s/59b11baca3948f34db87)). Note that the skewed distribution was not a result from the slow dynamics of calcium transients (Supplemental Fig. S1B (https://figshare.com/s/59b11baca3948f34db87)). Because of the unimodal distribution, a distinction between “reliable” and “unreliable” neurons is not useful.

Neurons co-vary significantly with each other during stimulus presentation.

To quantify the level of coordination among the neurons in the population activity, we applied nonnegative TCA (see Methods) to the normalized $\Delta F/F$ data from recordings organized into a three-dimensional tensor (Fig. 1G-J), as previously described ((Williams et al. 2018)). We found that with 20 components, the nonnegative TCA decomposition captured a significant amount of neural responses (545 neurons x 350 time points x 30 trials) for neurons with diverse reliability without overfitting (Fig. 1H; Fig. 2A). We quantified the fitting performance of individual neurons by the coefficient of determination (R^2), and found that in general, fitting performances on neurons with high reliability were higher than that of neurons with low reliability (Fig. 2B). Given that TCA is built to capture responses that are shared across dimensions (across neurons, time or trials), it is not surprising to see that neurons with high reliability, whose responses are shared across trials, were better fit. However, for some neurons with low reliability, fitting performances were also surprisingly high (Fig. 2B), which suggests that their responses are shared across neurons. To quantify the extent to which neuronal responses are shared across neurons, we fitted TCA on neural responses with randomly shuffled trials for each neuron independently. Note that the reliability of each neuron after shuffling is still the same as that in the original data. Fitting performances on the original data were significantly better than fitting performances on data with shuffled trials (Fig. 2C), especially for neurons with low reliability. Furthermore, neurons were co-active largely independent of their reliability, supported by weak correlation between reliability of co-active neuron pairs (Supplemental Fig. S2D (https://figshare.com/s/e9f4a464bbde7f717bbe), see Methods). In conclusion, this comparison indicates that neuronal coactivation pattern significantly contributes to population activity during stimulus presentation from single (Fig. 2C) and combined

experiments (Supplemental Fig. S2B, C (https://figshare.com/s/e9f4a464bbde7f717bbe)).

Neurons with a range of reliability are co-active and encode stimulus-driven and -independent variables.

To reveal the encoding paradigm of the neurons, we visualized all 20 TCA components in matrix form, sorted by their consistency across trials (Fig. 3A). Here neuron factors directly reflect the coactivation pattern of the neurons (Supplemental Fig. S2E, F (https://figshare.com/s/e9f4a464bbde7f717bbe)), and trial factors indicate whether the latent variables or TCA components are driven by the stimulus. We quantified the consistency of components by the coefficient of variation (*CV*) of their trial factors. Consistent components with low CV represent stimulus-driven variables, while inconsistent components with high CV represent stimulus-independent variables. In addition, we sorted the neurons based on their response reliability when visualizing neuron factors. The sorting by consistency and reliability revealed two key observations. First, there is a continuous distribution of consistency of components. Second, neurons with diverse reliability co-vary and encode different components, as indicated by 10 neurons with the highest neuron factor values for each component spanning a range of reliability (Fig. 3B; Supplemental Fig. S3 (https://figshare.com/s/2789918249d3fd0b7af1); Fig. 4). In other words, a single neuron's response reliability imposes only a weak constraint on its encoding capabilities. This spread of coactivation pattern across reliability leads to a seemingly paradoxical conclusion that neurons with low reliability can encode stimulus-driven variables and neurons with high reliability can encode stimulus-independent variables (Supplemental Fig. S2G, H (https://figshare.com/s/e9f4a464bbde7f717bbe)).

This apparent paradox is illustrated by responses from two example neurons (Fig. 3C, D). The neuron with low reliability in Fig. 3C displayed highly variable responses from trial to trial, however, whenever it fired, it fired at the same time point in the trial. Thus, the neuron with low reliability had a high neuron factor value (higher than one s.d. above mean) for the consistent component shown in Fig. 3C. By contrast, the neuron with high reliability in Fig. 3D had a high neuron factor value for the corresponding inconsistent component. This resulted from the fact that the neuron with high reliability not only encoded stimulus-driven variables, but also encoded stimulus-independent variables. The findings indicate that one neuron can co-vary with different groups of neurons and encode distinct variables. The two key observations largely hold for neural responses to drifting gratings, however there were fewer neurons with high reliability encoding stimulus-independent variables (Supplemental Fig. S3 (https://figshare.com/s/2789918249d3fd0b7af1)).

Neuronal coactivation pattern randomly redistributes across different stimuli.

Cortical neurons are deeply embedded in a recurrent neural circuit (Douglas et al. 1995). The recurrent nature of cortical circuits raises the question of how the observed single-neuron reliability and the population coactivation patterns are modulated by feedforward visual input. To investigate the impact of feedforward and recurrent input, we analyzed neural responses to a naturalistic movie clip (Nat Mov) and periodic drifting gratings (PDG) stimuli from neurons in the same imaging field. In order to make a direct comparison across stimuli, we matched their trial structure for all analyses (see Methods).

First, we compared how single neuron activity changes across stimuli. The activity level (averaged $\Delta F/F$ over time) of cortical neurons followed a skewed distribution during

both Nat Mov and PDG stimulation (Fig. 5A). In addition, neurons' activity level substantially redistributed across stimuli (Fig. 5B). The response reliability to both stimuli also followed skewed distributions (Fig. 5C) and extensively redistributed across stimuli (Fig. 5D).

Second, we compared how neuronal ensembles change across stimuli. Are neurons co-active in the same way during Nat Mov stimulation and PDG stimulation? To answer this question, we fitted TCA with 20 components on concatenated neural responses (Fig. 5E). Note that TCA is ignorant to which stimulus is on during each trial. Despite this lack of information about the trial structure, TCA successfully identified two groups of components corresponding to the two stimuli (Fig. 5E). As expected, the consistent components during PDG stimulation reflect the tuning curves of orientation selective neurons, with two peaks for their temporal factors corresponding to responses to orientations separated by 180 degrees. To quantify similarities between neural ensembles, we calculated the correlation coefficient (CC) between neuron factors of different components (Fig. 5F). Note that TCA factors are not necessarily orthogonal to each other, in contrast to principal component analysis (Kruskal 1977; Williams et al. 2018). Thus, the CC between neuron factors is not expected to be zero or negative. We found that inter-component CCs *within* stimuli were predominantly negative while inter-component CCs *across* stimuli centered around zero (Fig. 5G). A negative CC between two components indicates that if one neuron is recruited by one component, it is unlikely to be recruited by the other component. Consequently, different TCA components within stimuli, i.e., Nat Mov or PDG, tend to be encoded by largely non-overlapped ensembles of neurons, while different TCA components across stimuli, i.e., Nat Mov vs PDG, tend to be encoded by

random ensembles of neurons. Importantly, the fact that neuronal ensembles are randomly reorganized for different external visual inputs, raises the question whether neural ensembles are formed mainly due to feed-forward external inputs instead of cortical recurrent connections.

A balanced model network with random connectivity and correlated external inputs reproduces key features of the observed cortical activity.

To identify a potential mechanism behind the observed cortical dynamics, we simulated a balanced model network (van Vreeswijk and Sompolinsky 1996, 1998a) with random connectivity and clustered external inputs (clustered as defined by grouping of neuron inputs; note that the model has no spatial organization, see Methods and Fig. 6A). In brief, the recurrent model network consisted of 1600 excitatory and 400 inhibitory binary point neurons with uniform random connectivity for each neuron type (see Supplemental Fig S4A (https://figshare.com/s/576fe4fa7850f8cecde5) and Methods). To mimic stimulus-driven and -independent variables in the model, we constructed two qualitatively different sets of external input pulse trains (Supplemental Fig. S4D, E (https://figshare.com/s/576fe4fa7850f8cecde5)). One set of 20 different input pulse trains was identical ("frozen") across trials, thus imitating stimulus-driven variables. Another set of 20 different input pulse trains varied in a trial-independent manner, thus imitating stimulus-independent variables. To mimic coactivation patterns among neurons, we randomly partitioned the 2000 model neurons into 20 clusters of 100 model neurons each. All neurons within a cluster received the same "frozen" input pulse train. For another random partitioning of the model neurons into 20 clusters of 100 neurons, all neurons

within a cluster received the same input pulse train, that, however, varied in a trial-independent manner. To match with the temporal structure of experimental data, we mimicked $\Delta F/F$ responses by convolving simulated spike trains with alpha functions (see Methods). All of the following analyses were performed on the simulated $\Delta F/F$ responses.

With a choice of appropriate set of parameters, key features of the observed cortical activity were reproduced by the model network (Fig. 6). Even though model neurons received highly correlated external inputs, they operated in an asynchronous state (Fig. 6B) due to balanced excitatory and inhibitory recurrent inputs (Renart et al. 2010). In addition, with lognormal distributed synaptic weights of external inputs (Supplemental Fig. S4B (https://figshare.com/s/576fe4fa7850f8cecde5)), the model exhibited a skewed distribution of response reliability (Fig. 6C). Furthermore, consistent with experimental results, simulated activities of model neurons were well fitted by TCA (Fig. 6D) and they co-varied more than expected by chance (Fig. 6E). Moreover, both consistent and inconsistent components recruited neurons with a range of reliability (Fig. 6F, G). Importantly, when the model network was presented with two different stimuli (see Methods), inter-component CCs *within* stimuli were predominantly negative while inter-component CCs *across* stimuli centered around zero (Fig. 6H, I).

By reproducing the observed cortical dynamics, the model revealed several essential insights. First, the clustered structure in external inputs, instead of the clustered structure in recurrent connections (Supplemental Fig. S5A-C (https://figshare.com/s/c95cdcf1477941ee7875)), is more likely to support the observed coactivation pattern in neuronal responses. Clustered recurrent connections would lead to spontaneous slow dynamics during which neurons within the cluster transiently increased

or decreased their firing rate (Litwin-Kumar and Doiron 2012). This spontaneous slow dynamics results in multiple inconsistent components with different temporal factors but the same neuron factors (Supplemental Fig. S5C (https://figshare.com/s/c95cdcf1477941ee7875)), which is contradictory to the experimental results. In contrast, TCA components of the model with clustered external inputs and random connectivity qualitatively resembled TCA components from experimental data (Fig. 6). Second, to impose coactivation patterns on neurons with a range of reliability, each neuron needs to receive two kinds of inputs: (i) frozen input pulse train imitating a stimulus-driven variable, and (ii) trial-varied input pulse train imitating a stimulus-independent variable. If each neuron received either the frozen or the trial-varied input pulse train (Supplemental Fig. S5D (https://figshare.com/s/c95cdcf1477941ee7875)), then neurons' coactivation pattern would be determined by neuron's reliability (Supplemental Fig. S5E, F (https://figshare.com/s/c95cdcf1477941ee7875), Fig. 1B, the constrained case).

In conclusion, this analysis of recurrent balanced neural network models revealed that both the stimulus specificity and the mixed encoding of qualitatively different variables can arise from clustered external inputs.

Discussion

Neural variability is widely studied as a single-neuron feature (Faisal et al. 2008; Mainen and Sejnowski 1995) and a population-wide feature (Cohen and Kohn 2011; Doiron et al. 2016). Here, we related single-neuron variability to population-wide variability by asking how neurons with different levels of reliability encode unobserved variables. Our work

demonstrated that neurons spanning a range of reliability are co-active and encode a mixture of stimulus-driven and stimulus-independent unobserved variables. We found that a neuron's response reliability and the neuronal coactivation patterns substantially reorganized for different external visual inputs. Furthermore, our model suggested clustered external inputs underpin the observed coactivation pattern of neurons. More broadly, this study has made the following contributions to our understanding of connectivity-mediated variability in visual cortex.

First, we found that neural variability is well captured by additive and multiplicative modulation shared across neuron ensembles, as shown by the applicability of the linear TCA analysis (Fig. 2A, B). Neural variability can be modeled as an additive modulation (Scholvinck et al. 2015) by summing the trial-averaged evoked response and some stochastic activity such as spontaneous activity (Arieli et al. 1996). Alternatively, neural variability can be modeled as a multiplicative modulation (Ecker et al. 2014; Goris et al. 2014) by multiplying the trial-averaged evoked response with a gain factor. Both additive and multiplicative modulations are necessary to reproduce neural variability observed in experimental data (Arandia-Romero et al. 2016; Lin et al. 2015). Here, we modeled trial-to-trial variability as a sum of gain-changed temporal factors, where the gain is governed by the corresponding neuron factor and trial factor. Note that the temporal factors represent the shared neural activity across neurons and trials, which might serve as a better neural basis than the trial-averaged evoked responses (Williams et al. 2018).

Second, we found that individual neuron’s response reliability imposes only a weak constraint on its encoding capabilities. One explanation given for the presence of neurons exhibiting weak responses to sensory stimuli is that even poorly-driven neurons may

contribute to sensory coding (Leavitt et al. 2017; Safaai et al. 2013). Indeed, we show that neurons with low reliability often make strong contributions to consistent stimulus-driven factors, despite the fact that the responses of individual neurons can be highly variable across trials (Fig. 3). In contrast, researchers have proposed that variable activity across trials is due to coding of non-sensory information, such as motor or behavioral variables (Niell and Stryker 2010; Vinck et al. 2015). A recent paper using shared variance component analysis identified stimulus-independent latent factors that were linked to facial movements and drove visual cortical neurons independently of sensory input (Stringer et al., 2019b). Our results are also in agreement with this finding, as we show that neurons from a range of reliability contribute to stimulus-independent latent factors (Fig. 3). Taken together, these results show that the encoding of distinct variables are not mutually exclusive, and that both phenomena are evident in visual cortical networks.

Third, our experiment and model results support the possibility that clustered external inputs underpin the neuronal coactivation pattern. Alternatively, co-active neuronal ensembles could result from structured recurrent connectivity, based on the fact that the connectivity probability between co-active neurons is higher than neurons with decorrelated evoked responses (Ko et al. 2011). Additional evidence in support of this alternative mechanism is the similarity between co-active neuronal ensembles during spontaneous and stimulus-modulated activity (MacLean et al. 2005; Miller et al. 2014). However, the evidence might not be sufficient: a neural network with random connectivity can also generate similar neuronal coactivation patterns during spontaneous and evoked activity (Okun et al. 2012). Moreover, consistent with previous work (Hofer et al. 2011), we found that neuronal coactivation pattern is highly dependent on stimulus (Fig. 5), which

548 demonstrated that external inputs, instead of recurrent connection, may be the dominant
549 factor in the formation of neuronal ensembles. The mechanism underlying these
550 coactivation patterns is still unclear. Searching for further evidence for our proposed
551 mechanism might require analyses on simultaneous recordings from external inputs and
552 cortical neurons (Sun et al. 2016).

553 Fourth, the coactivation pattern of neurons with diverse reliability provides insights
554 on the connectivity of external inputs to visual cortex. Neuroanatomy data showed that V1
555 in mice is highly interconnected with other regions of neocortex (Froudarakis et al. 2019).
556 For instance, V1 receives inputs carrying sensorimotor information (Petreanu et al. 2012).
557 However, the structure of inputs at the neuronal population level remains elusive. In Figure
558 1, we described a spectrum of how neurons encode stimulus-driven and -independent
559 variables. Based on model investigations (Supplemental Fig. S5D-F
560 (https://figshare.com/s/c95cdcf1477941ee7875) & Fig. 6), the two extremes of the
561 spectrum correspond to different external input structures. Our experimental and model
562 results suggested that a neuron's reliability imposes only a weak constraint on its encoding
563 capability, indicating that neurons receive both frozen and trial-varied inputs. This input
564 paradigm has a potential functional advantage such that fewer neurons are required to
565 encode the same number of variables, compared to distinct external inputs projecting to
566 separate groups of neurons. Furthermore, different variables are encoded by largely non-
567 overlapped groups of neurons within a stimulus set (Fig. 5). This non-overlapping encoding
568 strategy indicates that each input tends to innervate different groups of neurons. Such a
569 mutually exclusive representation may enable simple linear readout for downstream
570 neurons. This tradeoff between efficient coding and high readout efficiency informed the

choice of the input structure in our model. However, the chosen input structure in our model may not be the only possible solution to reproduce the key features of neuronal coactivation patterns. Another limitation of our model is that we assumed random connectivity between model neurons, which is not true for cortical neurons. Models with spatial dependence in connectivity resembling cortical networks (Huang et al. 2019) are good candidates to be investigated in the future.

An important next step is to identify what stimulus-driven and -independent variables are encoded by neural responses. Earlier work suggests two possible ways to identify the stimulus-independent variables. First, we can look for behavioral or internal variables which have the highest correlation with the trial factors of inconsistent components (Hirokawa et al. 2019; Stringer et al. 2019b). Second, we can use photostimulation to activate the neuronal ensemble corresponding to the stimulus-independent component and observe the changes of behavioral variables (Carrillo-Reid et al. 2019). However, it is much less straightforward to identify the stimulus-driven variables or visual features in this case. One promising idea is using a generative closed-loop system to evolve synthetic images to maximize the corresponding neuronal ensemble's coactivation (Bashivan et al. 2019; Ponce et al. 2019). Such evolved images might provide insight on the visual features encoded by the particular neuron ensemble.

Acknowledgments: We would like to thank Gerald Pho and Barani Raman for comments on the project. This work was supported by the following grants: Whitehall Foundation #20121221 (RW), NSF CRCNS #1308159 (RW), NIH R00 MH104259 (MJG), Whitehall Foundation #20181228 (MJG), NSF NeuroNex #1707287 (MJG).

Author Contributions: Conceptualization, JX, RW and MJG; Methodology, JX, RW and MJG; Software: JX, TDM, and MJG; Formal Analysis: JX; Investigation: TDM and MJG; Resources: MJG; Data Curation: TDM and MJG; Writing - Original Draft: JX; Writing - Review & Editing: RW, TDM and MJG; Visualization: JX; Supervision: RW and MJG; Funding Acquisition: RW and MJG.

Competing Interests The authors declare no competing interests.

References

Allen WE, **Chen MZ**, **Pichamoorthy N**, **Tien RH**, **Pachitariu M**, **Luo L**, **Deisseroth K**. Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* 364: 253, 2019.

Arandia-Romero I, **Tanabe S**, **Drugowitsch J**, **Kohn A**, **Moreno-Bote R**. Multiplicative and Additive Modulation of Neuronal Tuning with Population Activity Affects Encoded Information. *Neuron* 89: 1305–1316, 2016.

Arieli A, **Sterkin A**, **Grinvald A**, **Aertsen A**. Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses. *Science* 273: 1868–1871, 1996.

Bashivan P, **Kar K**, **DiCarlo JJ**. Neural population control via deep image synthesis. *Science* 364, 2019.

Brainard DH. The Psychophysics Toolbox. *Spatial Vision* 10: 433–436, 1997.

Buzsáki G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68: 362–385, 2010.

Carrillo-Reid L, **Han S**, **Yang W**, **Akrouh A**, **Yuste R**. Controlling Visually Guided Behavior by Holographic Recalling of Cortical Ensembles. *Cell* 178: 447–457.e5, 2019.

Cohen MR, **Kohn A**. Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819, 2011.

Dipoppa M, **Ranson A**, **Krumin M**, **Pachitariu M**, **Carandini M**, **Harris KD**. Vision and Locomotion Shape the Interactions between Neuron Types in Mouse Visual Cortex.

Neuron 98: 602–615.e8, 2018.

Doiron B, **Litwin-Kumar A**, **Rosenbaum R**, **Ocker GK**, **Josić K**. The mechanics of state-dependent neural correlations. *Nature Neuroscience* 19: 383–393, 2016.

Douglas R, **Koch C**, **Mahowald M**, **Martin K**, **Suarez H**. Recurrent excitation in neocortical circuits. *Science* 269: 981–985, 1995.

Ecker AS, **Berens P**, **Cotton RJ**, **Subramaniyan M**, **Denfield GH**, **Cadwell CR**, **Smirnakis SM**, **Bethge M**, **Tolias AS**. State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82: 235–248, 2014.

Faisal AA, **Aldo Faisal A**, **Selen LPJ**, **Wolpert DM**. Noise in the nervous system. *Nature Reviews Neuroscience* 9: 292–303, 2008.

Froudarakis E, **Fahey PG**, **Reimer J**, **Smirnakis SM**, **Tehovnik EJ**, **Tolias AS**. The Visual Cortex in Context. *Annu Rev Vis Sci* 5: 317–339, 2019.

Goard M, **Dan Y**. Basal forebrain activation enhances cortical coding of natural scenes. *Nat Neurosci* 12: 1444–1449, 2009.

Goris RLT, **Movshon JA**, **Simoncelli EP**. Partitioning neuronal variability. *Nat Neurosci* 17: 858–865, 2014.

Hirokawa J, **Vaughan A**, **Masset P**, **Ott T**, **Kepecs A**. Frontal cortex neuron types categorically encode single decision variables. *Nature* 576: 446–451, 2019.

Hofer SB, **Ko H**, **Pichler B**, **Vogelstein J**, **Ros H**, **Zeng H**, **Lein E**, **Lesica NA**, **Mrsic-Flogel TD**. Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat Neurosci* 14: 1045–1052, 2011.

Huang C, **Ruff DA**, **Pyle R**, **Rosenbaum R**, **Cohen MR**, **Doiron B**. Circuit Models of Low-Dimensional Shared Variability in Cortical Networks. *Neuron* 101: 337–348.e4, 2019.

Kobak D, **Brendel W**, **Constantinidis C**, **Feierstein CE**, **Kepecs A**, **Mainen ZF**, **Qi X-L**, **Romo R**, **Uchida N**, **Machens CK**. Demixed principal component analysis of neural population data. *Elife* 5, 2016.

Ko H, **Hofer SB**, **Pichler B**, **Buchanan KA**, **Sjöström PJ**, **Mrsic-Flogel TD**. Functional specificity of local synaptic connections in neocortical networks. *Nature* 473: 87–91, 2011.

Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications* 18: 95–138, 1977.

Leavitt ML, **Pieper F**, **Sachs AJ**, **Martinez-Trujillo JC**. Correlated variability modifies

working memory fidelity in primate prefrontal neuronal ensembles. *Proc Natl Acad Sci U S A* 114: E2494–E2503, 2017.

Lin I-C, **Okun M**, **Carandini M**, **Harris KD**. The Nature of Shared Cortical Variability. *Neuron* 87: 644–656, 2015.

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

MacLean JN, **Watson BO**, **Aaron GB**, **Yuste R**. Internal dynamics determine the cortical response to thalamic stimulation. *Neuron* 48: 811–823, 2005.

Madisen L, **Garner AR**, **Shimaoka D**, **Chuong AS**, **Klapoetke NC**, **Li L**, **van der Bourg A**, **Niino Y**, **Egolf L**, **Monetti C**, **Gu H**, **Mills M**, **Cheng A**, **Tasic B**, **Nguyen TN**, **Sunkin SM**, **Benucci A**, **Nagy A**, **Miyawaki A**, **Helmchen F**, **Empson RM**, **Knöpfel T**, **Boyden ES**, **Clay Reid R**, **Carandini M**, **Zeng H**. Transgenic Mice for Intersectional Targeting of Neural Sensors and Effectors with High Specificity and Performance. *Neuron* 85: 942–958, 2015.

Mainen Z, **Sejnowski T**. Reliability of spike timing in neocortical neurons. *Science* 268: 1503–1506, 1995.

Mante V, **Sussillo D**, **Shenoy KV**, **Newsome WT**. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503: 78–84, 2013.

McGinley MJ, **David SV**, **McCormick DA**. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* 87: 179–192, 2015.

Miller J-EK, **Ayzenshtat I**, **Carrillo-Reid L**, **Yuste R**. Visual stimuli recruit intrinsically generated cortical ensembles. *Proc Natl Acad Sci U S A* 111: E4053–61, 2014.

Niell CM, **Stryker MP**. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65: 472–479, 2010.

Okun M, **Yger P**, **Marguet SL**, **Gerard-Mercier F**, **Benucci A**, **Katzner S**, **Busse L**, **Carandini M**, **Harris KD**. Population Rate Dynamics and Multineuron Firing Patterns in Sensory Cortex. *Journal of Neuroscience* 32: 17108–17119, 2012.

Olshausen BA, **Field DJ**. What Is the Other 85 Percent of V1 Doing? *23 Problems in Systems Neuroscience* : 182–212, 2006.

Petreanu L, **Gutnisky DA**, **Huber D**, **Xu N-L**, **O'Connor DH**, **Tian L**, **Looger L**, **Svoboda K**. Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature* 489: 299–303, 2012.

Ponce CR, **Xiao W**, **Schade PF**, **Hartmann TS**, **Kreiman G**, **Livingstone MS**. Evolving Images for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences. *Cell* 177: 999–1009.e10, 2019.

Renart A, **de la Rocha J**, **Bartho P**, **Hollender L**, **Parga N**, **Reyes A**, **Harris KD**. The asynchronous state in cortical circuits. *Science* 327: 587–590, 2010.

Rikhye RV, **Sur M**. Spatial Correlations in Natural Scenes Modulate Response Reliability in Mouse Visual Cortex. *Journal of Neuroscience* 35: 14661–14680, 2015.

Safaai H, **von Heimendahl M**, **Sorando JM**, **Diamond ME**, **Maravall M**. Coordinated population activity underlying texture discrimination in rat barrel cortex. *J Neurosci* 33: 5843–5855, 2013.

Saxena S, **Cunningham JP**. Towards the neural population doctrine. *Current Opinion in Neurobiology* 55: 103–111, 2019.

Scholvinck ML, **Saleem AB**, **Benucci A**, **Harris KD**, **Carandini M**. Cortical State Determines Global Variability and Correlations in Visual Cortex. *Journal of Neuroscience* 35: 170–178, 2015.

Softky WR, **Koch C**. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* 13: 334–350, 1993.

Stringer C, **Pachitariu M**, **Steinmetz N**, **Carandini M**, **Harris KD**. High-dimensional geometry of population responses in visual cortex. *Nature* 571: 361–365, 2019a.

Stringer C, **Pachitariu M**, **Steinmetz N**, **Reddy CB**, **Carandini M**, **Harris KD**. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364: 255, 2019b.

Sun W, **Tan Z**, **Mensh BD**, **Ji N**. Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs. *Nat Neurosci* 19: 308–315, 2016.

Vinck M, **Batista-Brito R**, **Knoblich U**, **Cardin JA**. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron* 86: 740–754, 2015.

van Vreeswijk C, **Sompolinsky H**. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274: 1724–1726, 1996.

van Vreeswijk C, **Sompolinsky H**. Chaotic balanced state in a model of cortical circuits. *Neural Comput* 10: 1321–1371, 1998a.

van Vreeswijk C, **Sompolinsky H**. Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation* 10: 1321–1371, 1998b.

Williams AH, **Kim TH**, **Wang F**, **Vyas S**, **Ryu SI**, **Shenoy KV**, **Schnitzer M**, **Kolda TG**, **Ganguli S**. Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron* 98: 1099–1115.e8, 2018.

726 **Yuste R**. From the neuron doctrine to neural networks. *Nat Rev Neurosci* 16: 487–497,
727 2015.

728

Figure 1. Response reliability has a skewed distribution.

A. We decompose single trial neural responses into stimulus-driven latent factors (green) that are consistent across trials and stimulus-independent latent factors (black) that are inconsistent across trials.

B. Schematic shows two extremes of a spectrum of possibilities for the encoding of stimulus-driven and stimulus-independent variables: left, reliable neurons co-vary and encode stimulus-driven variables, while unreliable neurons co-vary and encode stimulus-independent variables; right, neuron's reliability does not constrain its encoding capability, thus, neurons covary and encode both kinds of variables regardless of their reliability.

C. Experimental setup. We performed two-photon calcium imaging of excitatory neurons in the primary visual cortex of awake, head-fixed mice during visual stimulation with periodic drifting gratings and repeated identical naturalistic movie clips.

D. Visual cortex (contralateral to visual stimulus delivery) is retinotopically mapped in Emx1-Cre::TITL-GCaMP6s mice. V1 fields are chosen from the region selective for the center of the presentation screen. Widefield scale bar = 1 mm; 2-photon scale bar = 100 μm.

E. $\Delta F/F$ responses of one example neuron with high reliability (top) and one example neuron with low reliability (bottom) during the same naturalistic movie clip for 30 trials (movie starts at 5 sec and lasts for 30 sec duration).

F. Distribution of response reliability for 545 recorded neurons in one example imaging field.

G. Schematic of Tensor Component Analysis (TCA). Neural data is organized into a third-order tensor with dimensions N x T x K. TCA approximates the data as a sum of outer products of three vectors from R components: neuron factors describe the weights of each neuron, temporal factors describe the temporal dynamics of each latent factor, and trial factors describe the modulation across trials.

H. Cross validation of TCA (Williams et al., 2018) on one example dataset (545 neurons x 350 frames x 30 trials). Normalized reconstruction error (see Methods) plotted against the number of

components of TCA for training set (blue) and test set (orange). Dashed line denotes the TCA model with 20 components.

I. One example component is displayed in the form of three vectors: neuron factor, temporal factor and trial factor.

J. All the components are displayed in the form of three heatmaps. Each row corresponds to one component (in this example R = 5).

Figure 2. Neurons co-vary significantly with each other during stimulus presentation.

A. $\Delta F/F$ traces (top) and reconstructed $\Delta F/F$ traces (bottom) based on 20 TCA components for one example neuron with low reliability (left) and one example neuron with high reliability (right) across trials in one example imaging field.

B. Fitting performance R^2 plotted against response reliability. Each dot represents one neuron in the example imaging field.

C. Fitting performance R^2 for original data plotted against R^2 for data with shuffled trials. Each dot represents one neuron. Color indicates response reliability. Fitting performance for the original data is significantly better than for data with shuffled trials (Mann-Whitney rank test, $p < 0.001$) in the same imaging field as in A and B.

Figure 3. Neurons with a range of reliability are co-active and encode stimulus-driven and -independent variables.

A. Neuron, temporal and trial factors of nonnegative TCA with 20 components. For all three factors, components are ordered by coefficient of variance (CV) of trial factors. In addition, within the neuron factors, neurons are ordered by their response reliability. Two example components are highlighted by horizontal rectangles: (yellow) A "consistent" component with a low CV value of trial factors. (red) An "inconsistent" component with a high CV value of trial factors.

B. Reliability (abscissa) and R^2 values (color and dot diameter) for the top 10 neurons with the largest neuron factor values within a component, shown for all 20 components (ordinate). Components are in the same order as in A.

C. One example component (same as yellow rectangle in D) that is consistent across trials (trial factor has low CV value). For clarity, here we used the display format of factors as described in Fig. 1I. For one neuron (red dot), the normalized responses and the reconstructed responses are shown below. As seen from the reconstructed response using this component alone (bottom), this neuron with low reliability has a large contribution from the consistent component.

D. One example component (same as red rectangle in D) that is inconsistent across trials (trial factor has high CV value). For one neuron (red dot), the normalized responses and the reconstructed responses are shown below. As seen from the reconstructed response using this component alone (bottom), this neuron with high reliability has a large contribution from the inconsistent component.

Figure 4. A single neuron's response reliability imposes only a weak constraint on its encoding capabilities.

A. Schematic shows reliability distribution of neurons encoding stimulus-driven and -independent variables for extremes of a spectrum of possibilities illustrated in Fig. 1B. Reliability of neurons with large neuron factor values are shown for each TCA component. Consistent components are assumed to represent stimulus-driven variables, while inconsistent components are assumed to represent stimulus-independent variables. Left: neuron's encoding capability is constrained by its reliability; right: neuron's encoding capability is not constrained by its reliability.

B. Reliability (abscissa) averaged over top 10 neurons with the largest neuron factor values within a component, shown for all 20 components (ordinate). Components are ordered by consistency, similar to Fig 3B. Shaded area denotes standard deviation of reliability over the top 10 neurons.

Here different colors denote different imaging fields. Reliability is positively correlated with the consistency of components. (10 imaging fields; Spearman correlation $r = 0.37 \pm 0.093$, $p < 0.05$) Dashed line denotes the expected relation between reliability and the consistency of components under constrained extreme, while solid line denotes the expected relation under unconstrained extreme.

Figure 5. Neuronal coactivation pattern randomly redistributes across different stimuli.

A. Distribution of averaged $\Delta F/F(\%)$ over time and trials during Nat Mov and PDG for 1 imaging field (opaque color) and the other 9 imaging fields (transparent color).

B. Averaged $\Delta F/F(\%)$ during Nat Mov plotted against averaged $\Delta F/F(\%)$ during PDG for neurons in 1 imaging field (black dots) and neurons in the other 9 imaging fields (gray dots). Averaged $\Delta F/F(\%)$ during Nat Mov is weakly correlated with averaged $\Delta F/F(\%)$ during PDG (4077 neurons, Pearson correlation $r = 0.07$, $p < 0.001$).

C. D. Same as A,B, but for response reliability. Reliability during Nat Mov is weakly correlated with reliability during PDG (4077 neurons, Pearson correlation $r = 0.09$, $p < 0.001$).

E. Twenty TCA components for concatenated neural responses to visual stimulation with PDG and Nat Mov. Ordering of components is determined by their trial factors (see Methods). Neuron factors are plotted with neurons ordered by their dominant components (see Methods).

F. The correlation coefficient (CC) between neuron factors are displayed with the same component order as in E (diagonal entries are set to zero).

G. Distribution of CC between neuron factors. Orange is for CC between neuron factors during Nat Mov; blue is for CC between neuron factors during PDG; green is for CC between neuron factors across stimuli; black is for CC between random vectors with the same dimension as neuron factors, representing the chance level. Opaque color is for 1 imaging field; transparent color is for the other 9 imaging fields. Both CC during Nat Mov and during PDG is significantly negative (one-sample t-

test, for all 10 imaging fields, $p < 0.001$). CC across stimuli is centered around zeros (one-sample t-test, for all 10 imaging fields, $p > 0.1$).

Figure 6. A balanced network model with random connectivity and clustered external inputs reproduces key features of observed cortical activity.

A. Illustration for the input structure to the model network. We simulated a balanced network with uniform random connectivity. There are two types of external inputs: frozen input pulse trains and trial-varied input pulse trains. Both inputs have a clustered input structure but with different neuron partitions. Model network consists of 1600 excitatory neurons and 400 inhibitory neurons. 25 sec x 20 trials are simulated.

B. Raster plot of 500 randomly subsampled neurons during 1 trial. Blue dashed line separates inhibitory neurons from excitatory neurons.

C. Response reliability histogram for subsampled excitatory neurons. Response reliability is calculated based on simulated $\Delta F/F$ traces.

D. Fitting performance R^2 (20 TCA components) plotted against response reliability for subsampled excitatory neurons.

E. Fitting performance R^2 for original data plotted against R^2 for data with shuffled trials for subsampled excitatory neurons (Mann-Whitney rank test, $p < 0.005$). Color indicates response reliability.

F. Twenty TCA components. Components are ordered by CV of trial factors. In neuron factor, neurons are ordered by their response reliability.

G. Reliability of 10 neurons with the largest neuron factor values for different components. Components are in the same order as in F.

H. Twenty TCA components for concatenated neural responses to two stimuli. Neuron factors are plotted with neurons ordered by their dominant components (see Methods).

I. Distribution of CC between neuron factors. Color code is the same as Fig. 5G: orange is for CC between neuron factors during stimulus one; blue is for CC between neuron factors during stimulus two; green is for CC between neuron factors across stimuli; black is for CC between random vectors with the same dimension as neuron factors, representing the chance level. Both CC during stimulus one and CC during stimulus two is significantly negative (one-sample t-test, $p < 0.001$), CC across stimuli is centered around zero (one-sample t-test, $p = 0.23$).

Supplemental figure legends

https://figshare.com/s/7058a75fb2cd75dd22f4

Figure 1

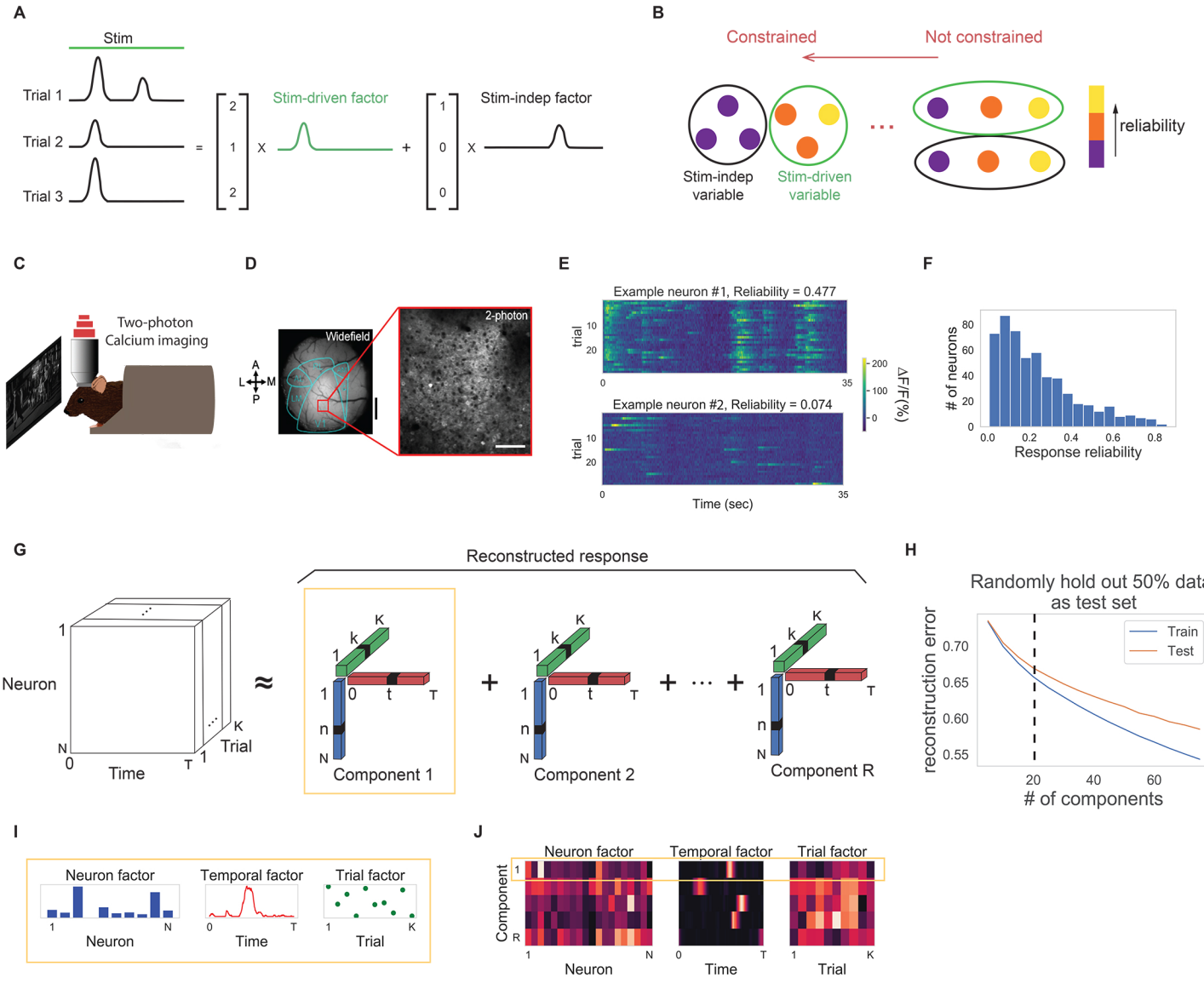

A
Stim
Trial 1
Trial 2
Trial 3
Stim-driven factor
Stim-indep factor
B
Constrained
Not constrained
Stim-indep variable
Stim-driven variable
reliability
C
Two-photon Calcium imaging
D
Widefield
2-photon
E
Example neuron #1, Reliability = 0.477
Example neuron #2, Reliability = 0.074
trial
Time (sec)
ΔF/F(%)
F
# of neurons
Response reliability
G
Neuron
Time
Trial
Reconstructed response
Component 1
Component 2
Component R
H
Randomly hold out 50% data as test set
reconstruction error
# of components
Train
Test
I
Neuron factor
Temporal factor
Trial factor
Neuron
Time
Trial
J
Neuron factor
Temporal factor
Trial factor
Component
Neuron
Time
Trial

Figure 2

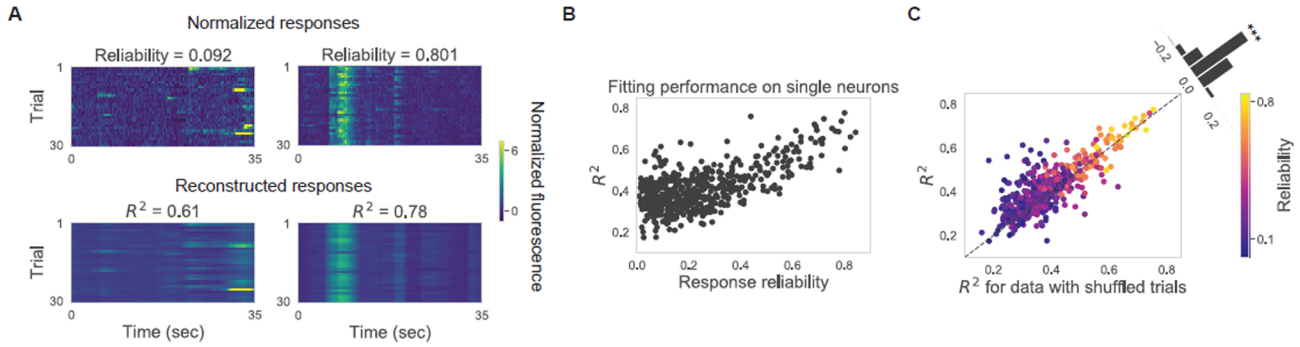

A
Normalized responses
Reliability = 0.092
Reliability = 0.801
Trial
Reconstructed responses
R^2 = 0.61
R^2 = 0.78
Time (sec)
Normalized fluorescence
B
Fitting performance on single neurons
R^2
Response reliability
C
R^2 for data with shuffled trials
Reliability

Figure 3

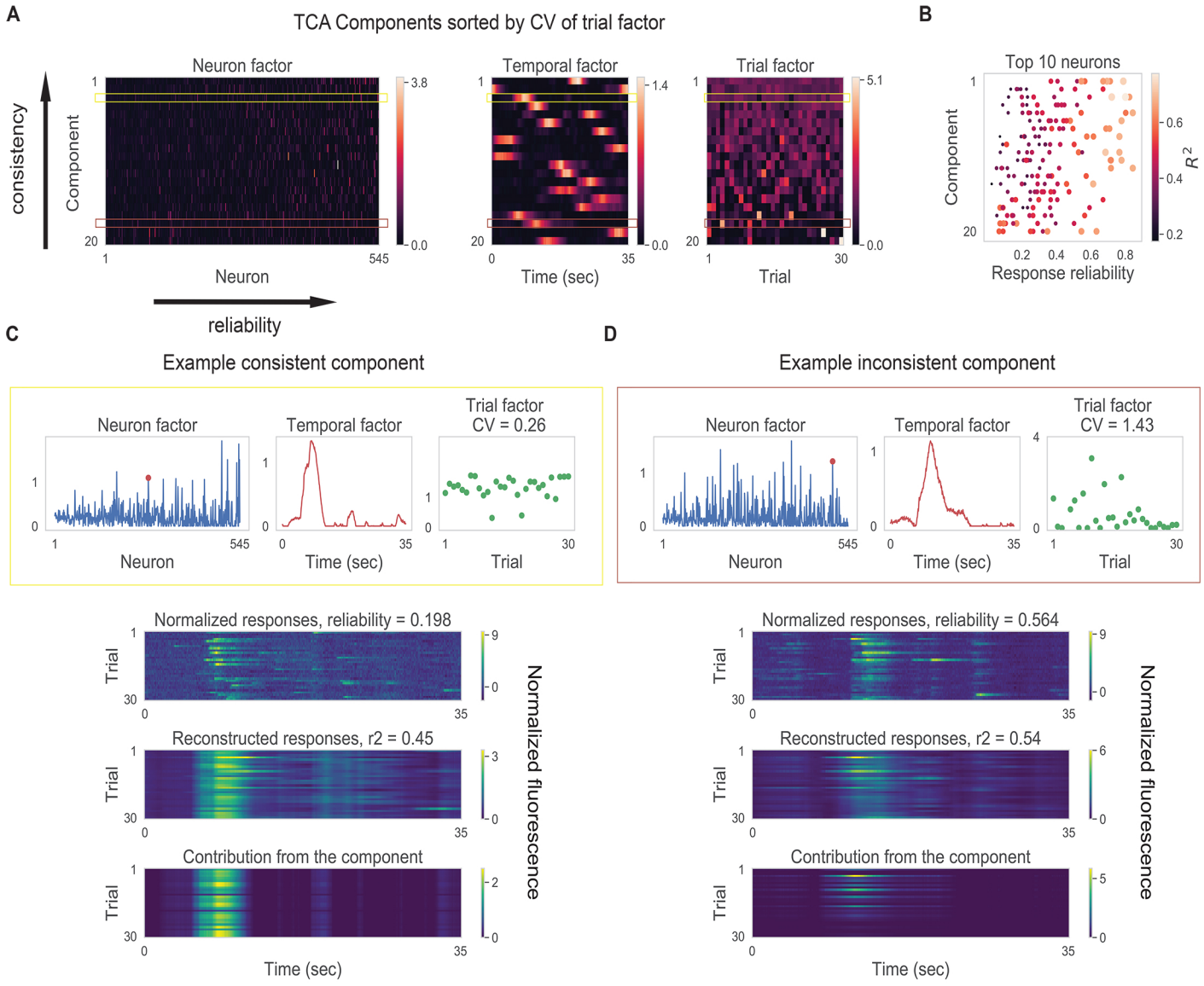

A
TCA Components sorted by CV of trial factor
Neuron factor
Temporal factor
Trial factor
consistency
Component
Neuron
Time (sec)
Trial
reliability
B
Top 10 neurons
Component
Response reliability
R^2
C
Example consistent component
Neuron factor
Temporal factor
Trial factor
CV = 0.26
Neuron
Time (sec)
Trial
D
Example inconsistent component
Neuron factor
Temporal factor
Trial factor
CV = 1.43
Neuron
Time (sec)
Trial
Normalized responses, reliability = 0.198
Reconstructed responses, r2 = 0.45
Contribution from the component
Trial
Time (sec)
Normalized fluorescence
Normalized responses, reliability = 0.564
Reconstructed responses, r2 = 0.54
Contribution from the component
Trial
Time (sec)
Normalized fluorescence

Figure 4

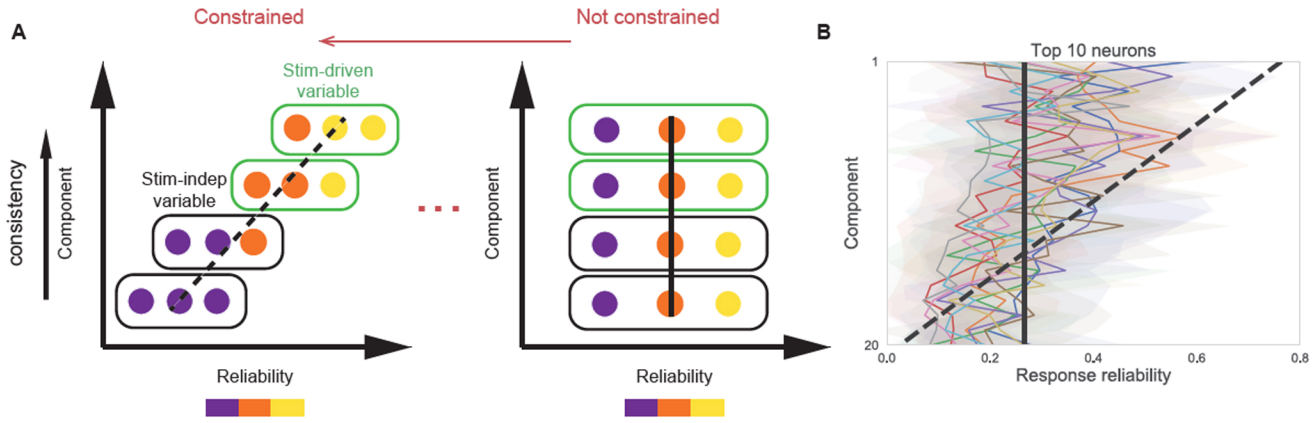

A
Constrained
Not constrained
Stim-driven variable
Stim-indep variable
consistency
Component
Reliability
Component
Reliability
B
Top 10 neurons
1
20
Component
0.0
0.2
0.4
0.6
0.8
Response reliability

Figure 5

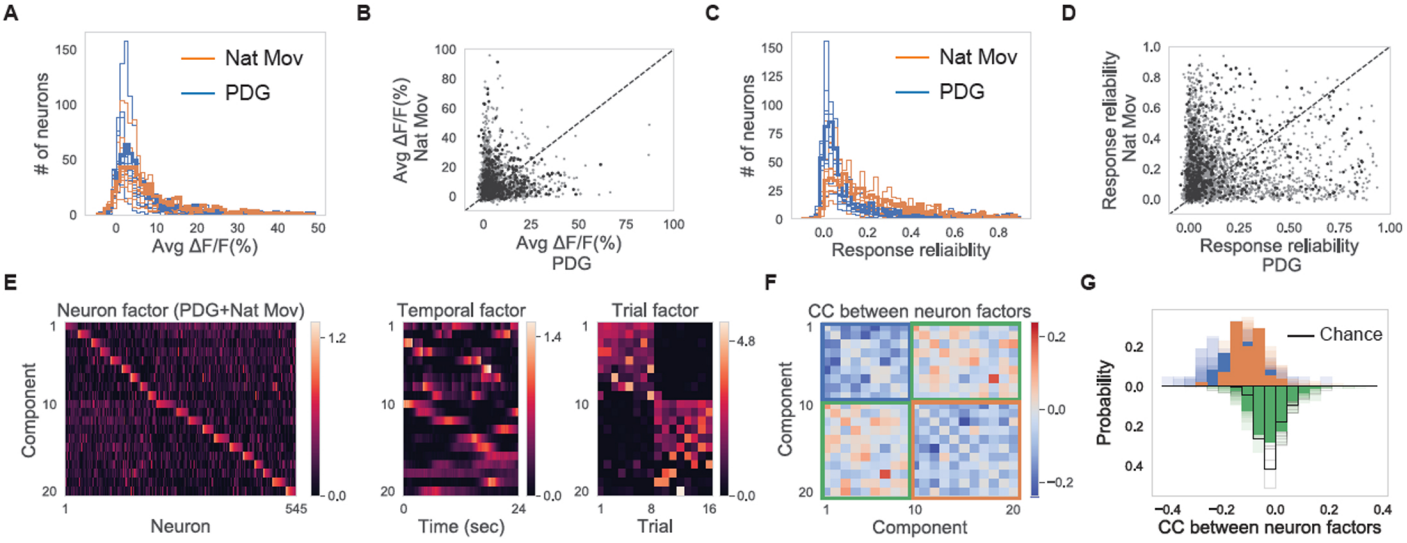
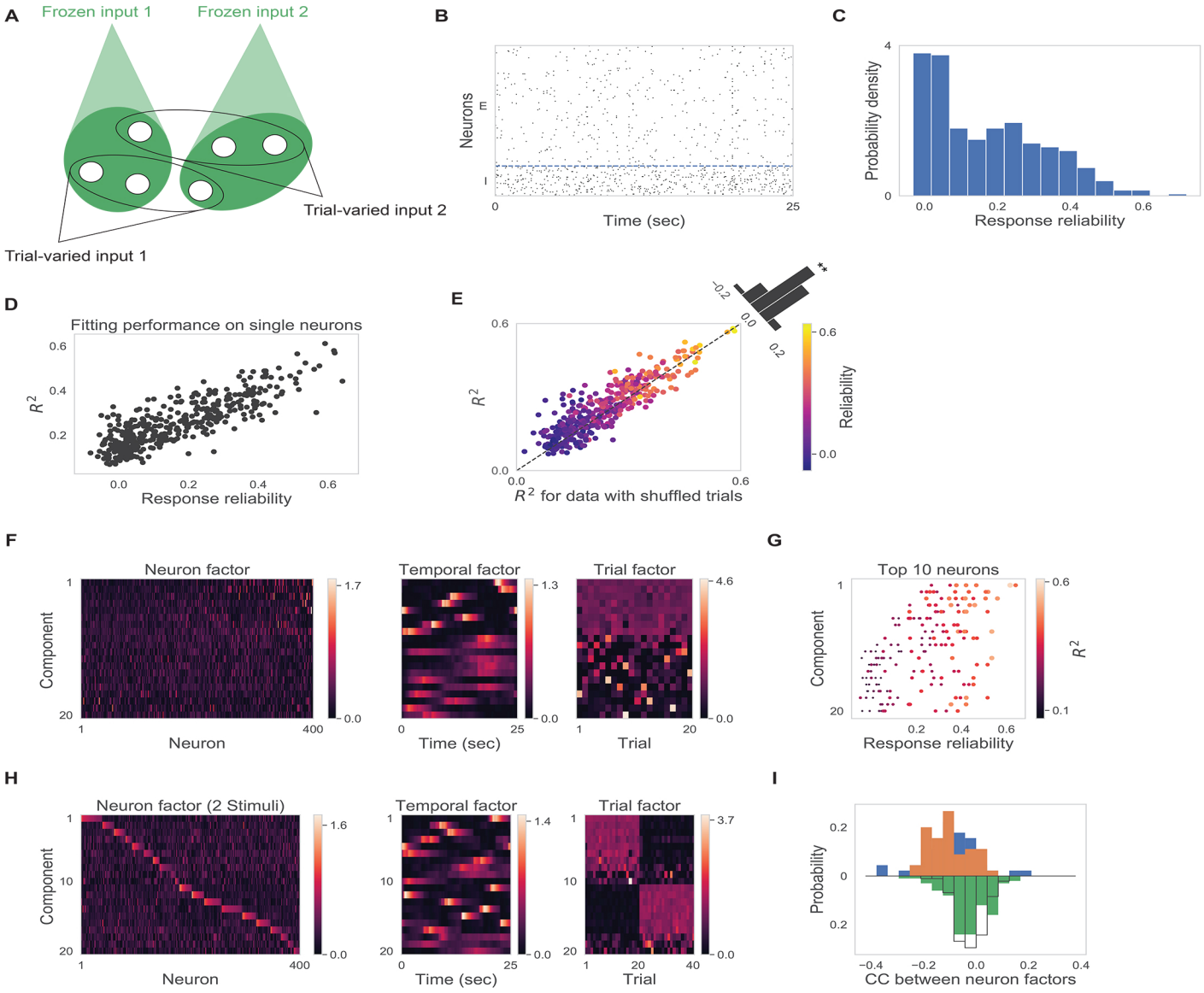

A
Nat Mov
PDG
# of neurons
Avg ΔF/F(%)
B
Avg ΔF/F(%)
Nat Mov
Avg ΔF/F(%)
PDG
C
Nat Mov
PDG
# of neurons
Response reliaiblity
D
Response reliability
Nat Mov
Response reliability
PDG
E
Neuron factor (PDG+Nat Mov)
Component
Neuron
Temporal factor
Time (sec)
Trial factor
Trial
F
CC between neuron factors
Component
Component
G
Chance
Probability
CC between neuron factors

Figure 6


A
Frozen input 1
Frozen input 2
Trial-varied input 2
Trial-varied input 1
B
Neurons
E
I
0
25
Time (sec)
C
Probability density
4
0
0.0
0.2
0.4
0.6
Response reliability
D
Fitting performance on single neurons
R^2
0.6
0.4
0.2
0.0
0.2
0.4
0.6
Response reliability
E
R^2
0.6
0.0
0.0
0.6
R^2 for data with shuffled trials
-0.2
0.0
0.2
*
0.6
0.0
Reliability
F
Neuron factor
Component
1
20
1
400
Neuron
1.7
0.0
Temporal factor
0
25
Time (sec)
1.3
0.0
Trial factor
1
20
Trial
4.6
0.0
G
Top 10 neurons
Component
1
20
0.2
0.4
0.6
Response reliability
0.6
0.1
R^2
H
Neuron factor (2 Stimuli)
Component
1
10
20
1
400
Neuron
1.6
0.0
Temporal factor
1
10
20
0
25
Time (sec)
1.4
0.0
Trial factor
1
10
20
1
20
40
Trial
3.7
0.0
I
Probability
0.2
0
0.2
−0.4
−0.2
0.0
0.2
0.4
CC between neuron factors