Spontaneous behaviors drive multidimensional, brain-wide neural activity

Carsen Stringer*[1,2], Marius Pachitariu*[1,3,4], Nicholas Steinmetz[5], Charu Reddy[5], Matteo Carandini†[5] and Kenneth D. Harris†[3,4]

[1] HHMI Janelia Research Campus, Ashburn, Virginia, 20147, USA. [2] Gatsby Computational Neuroscience Unit, UCL, London W1T 4JG, UK. [3] UCL Institute of Neurology, London WC1E 6DE, UK. [4] UCL Department of Neuroscience, Physiology, and Pharmacology, London WC1E 6DE, UK. [5] UCL Institute of Ophthalmology, London EC1V 9EL, UK.

* equal first authors, † equal senior authors. Correspondence to CS (stringerc@janelia.hhmi.org)

Sensory cortices are active in the absence of external sensory stimuli. To understand the nature of this ongoing activity, we used two-photon calcium imaging to record from over 10,000 neurons in the visual cortex of mice awake in darkness while monitoring their behavior videographically. Ongoing population activity was multidimensional, exhibiting at least 100 significant dimensions, some of which were related to the spontaneous behaviors of the mice. The largest single dimension was correlated with the running speed and pupil area, while a 16-dimensional summary of orofacial behaviors could predict ∼45% of the explainable neural variance. Electrophysiological recordings with 8 simultaneous Neuropixels probes revealed a similar encoding of high-dimensional orofacial behaviors across multiple forebrain regions. Representation of motor variables continued uninterrupted during visual stimulus presentation, occupying dimensions nearly orthogonal to the stimulus responses. Our results show that a multidimensional representation of motor state is encoded across the forebrain, and is integrated with visual input by neuronal populations in primary visual cortex.

Introduction

In the absence of sensory inputs, the brain produces structured patterns of activity, which can be as large or larger than sensory driven activity (Ringach, 2009). Ongoing activity exists even in primary sensory cortices, where neuronal activity at any time reflects an interaction of sensory-driven and ongoing components.

Several not mutually exclusive hypotheses exist for the function of ongoing activity and its interaction with stimulus responses. The first is that ongoing activity is simply correlated noise, for example produced by strong recurrent connectivity in neural circuits (Parga and Abbott, 2007; Stringer et al., 2016; van Vreeswijk and Sompolinsky, 1996). According to this view, ongoing activity is not a feature but a bug: it impairs cortical representations of sensory stimuli by corrupting sensory responses with noise of similar correlation structure (Averbeck et al., 2006; Cohen and Kohn, 2011; Shadlen and Newsome, 1998). A second theory holds that spontaneous activity patterns reflect recapitulation of previous sensory experiences, or expectations of possible sensory events. This hypothesis is supported by studies that found similarities between sensory-driven and spontaneous firing events (Berkes et al., 2011; Han et al., 2008; Hoffman and McNaughton, 2002; Kenet et al., 2003; Luczak et al., 2009). A third possibility is that ongoing activity could be related to behavioral and cognitive states. The firing of visual cortical neurons correlates with behavioral variables such as locomotion and pupil diameter (Dipoppa et al., 2016; McGinley et al., 2015; Niell and Stryker, 2010; Pakan et al., 2016; Polack et al., 2013; Reimer et al., 2016; Saleem et al., 2013; Stringer et al., 2016; Vinck et al., 2015), while in barrel cortex, whisking modulates neuronal firing (Gentet et al., 2010, 2012; Peron et al., 2015). These behaviorally-related changes in ongoing activity may reflect a "gain modulation" that boosts the salience of sensory modalities that are behaviorally relevant in that context, while suppressing others (Fu et al., 2014; Schneider and Mooney, 2015; Wimmer et al., 2015).

To try to distinguish between these possibilities and others, we considered the dimensionality of the ongoing neural activity and its relation to sensory responses and behavior. If ongoing activity represents a recapitulation of possible stimulus responses, it should be high dimensional, and these dimensions should be similar to those of sensory responses but unrelated to ongoing motor actions. If ongoing activity represents a single behavioral state such as locomotion or arousal, it should be one dimensional. Alternatively, if a sensory area encodes a high-dimensional representation of multiple motor variables, this suggests the area is performing a more complex integration of sensory input with cognitive and behavioral variables.

Here we investigate the statistical structure of ongoing population activity, first in visual cortex, and then in mul-

tiple areas of the forebrain. Because estimating dimensionality requires recording from many neurons simultaneously (Gao et al., 2017), we made simultaneous recordings of over 10,000 cells in visual cortex using two photon microscopy. We also recorded from over 2,000 neurons simultaneously in multiple cortical areas and deeper structures using high-density electrophysiology. We found that ongoing activity is complex, spanning at least 100 linear dimensions. These dimensions were represented by overlapping sets of neurons that did not have a spatial organization in the imaged tissue, and encoded multiple behavioral variables visible by videography of the whiskers, snout and other facial features. The dimensions occupied by ongoing activity were largely distinct from those encoding responses to visual stimuli.

Results

We monitored the activity of large populations of neurons in awake head-fixed mice that were not engaged in a behavioral task but were free to run on an air-floating ball or rotate a wheel with their front paws. Mice spontaneously performed behaviors such as whisking, sniffing, and other facial movements, which we monitored videographically using an infrared camera. Six out of nine of the recordings were performed in complete darkness. We did not observe differences between recordings in darkness (shown in red on all plots) and recordings with gray screen (yellow on all plots).

We recorded from 11,262 $\pm$ 2,282 (mean $\pm$ s.d.) neurons simultaneously in the visual cortex of six mice over nine sessions, using 2-photon imaging of GCaMP6s, in an 11-plane configuration with 2.5Hz scan rate (Pachitariu et al., 2016b) (Figure 1A). We focused our analysis on excitatory neurons, identified by the expression of GCaMP6s in transgenic mice with an excitatory promoter, or by expression of tdTomato under a Gad1 promoter, which allowed us to identify and exclude inhibitory neurons from further analysis.

The first dimension of ongoing activity is dominated by arousal

Ongoing population activity in visual cortex was highly structured. We computed pairwise correlations in 1.2 second bins (3 neural frames). The mean correlation coefficient over cell pairs was small, but the distribution of correlations was wide, with substantially more positive and negative correlations than would be expected by chance (Ecker et al., 2010; Renart et al., 2010) (Figure 1C,D). Indeed, the correlation coefficients were highly reliable (Figure 1D,E, mean r = 0.66 between correlations in first and second half of recordings, Figure 1F, G).

The largest component of the variance in ongoing activity was related to arousal. We extracted this factor by principal component analysis of ongoing population activity. An approximately equal number of neurons were correlated and anticorrelated with the first principal component (57% $\pm$ 6.7% SE), indicating that two large subpopulations of neurons alternate their activity in a push-pull manner (Figure 1I). The alternation between the activity of these two neuronal populations occurred over a timescale of many seconds (Figure S1), and was coupled to fluctuations in arousal levels, as indicated by standard measures such as running, pupil area, and whisking (correlations of $r = 0.69$, $r = 0.46$, and $r = 0.75$ respectively) (Figure 1J).

The slowness of these fluctuations implies a different underlying phenomenon to the "up and down states" previously studied in electrophysiological recordings (Constantinople and Bruno, 2011; Crochet and Petersen, 2006; Engel et al., 2016; Luczak et al., 2009; Okun and Lampl, 2008; Okun et al., 2015; Stringer et al., 2016; Vyazovskiy et al., 2011). Up/down states alternate at a much faster timescale (100-300 ms instead of 10-20 s), and a majority of neurons correlate with them positively, whereas approximately equal numbers of neurons correlated positively and negatively with the slow fluctuations in arousal described here. Indeed, up/down state changes could not even have been detected in these recordings, which scanned the cortex every 400 ms.

Across neuronal pairs, correlation in activity was only weakly related to distance (Figure 1K). Furthermore, neurons positively and negatively weighted on the first principal component had only a weak tendency to cluster together on the cortical surface (Figure 1L,M; same- and opposite-weighted neurons had a mean difference of 509 μm vs 526 μm, p=0.0025).

Predicting neural activity from one-dimensional measures of behavior

Having found that the top principal component was highly correlated with arousal, we asked whether further variations in the precise ongoing behavior of the mice might differentially relate to additional dimensions of neural activation. Indeed, the behaviors of the animals during the experiment exhibited transitions between multiple mo-

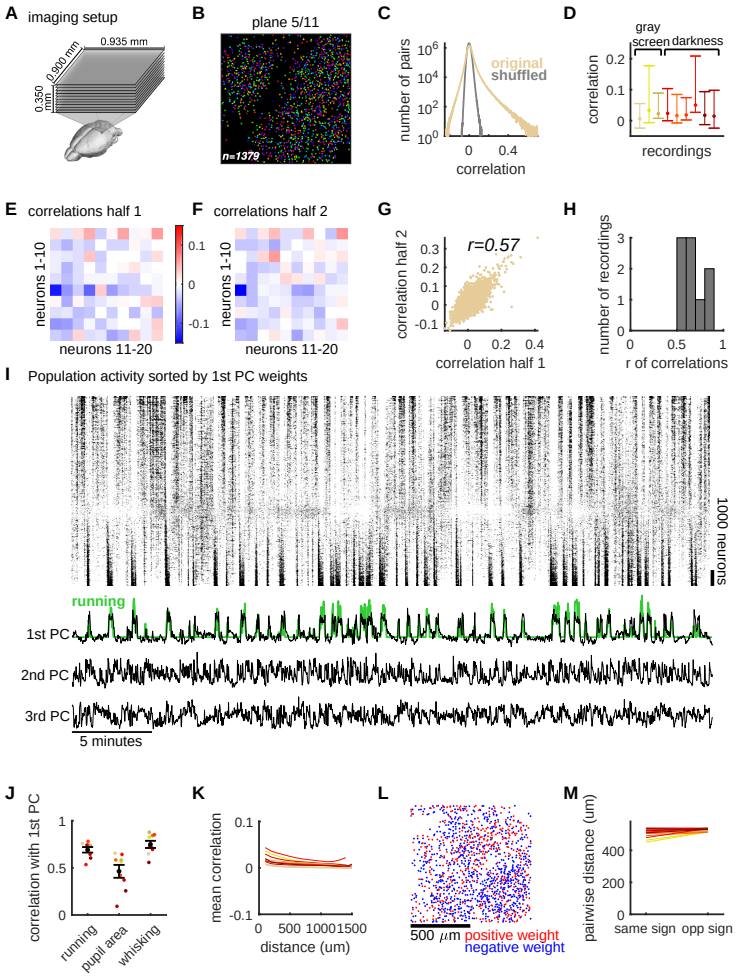


Figure 1. Structured ongoing population activity in visual cortex. (A) Two-photon calcium imaging of ∼10,000 neurons in visual cortex using multi-plane resonant scanning of 11 planes spaced at 35 μm apart. (B) Randomly-pseudocolored cells detected in an example imaging plane. (C) Distribution of pairwise correlations in ongoing activity (yellow). Gray: distribution of correlations after randomly time-shifting each cell's activity. (D) Bars showing of distribution of pairwise correlation coefficients for each recording (dot: mean; bar, range between 5th and 95th percentile). (E,F) Pseudocolor representation of correlation matrices for a subset of cells, computed independently from two halves of recording. (G) Scatter plot of correlations for each cell pair, computed from independent halves. (H) Histogram showing Pearson correlation coefficient of of pairwise correlations (as in G), across recordings. (I) Top: raster representation of ongoing population activity, with neurons arranged vertically sorted by weights onto the 1st principal component. Bottom: Time course of top three principal components of population activity. Green: running speed. (J) Correlation of first principal component with behavioral measures: running speed, pupil area and whisking. The mean and standard error across recordings is plotted. (K) Mean correlation over cell pairs, as a function of distance on the cortical surface. (L) Spatial distribution in tissue of neurons with positive and negative weights, in an example imaging plane. (M) Mean distance in tissue between neurons with same- or opposite-signed weights onto 1st PC.

tor states such as whisking, grooming, sniffing, running, chewing, sitting, or combinations of these.

We focused first on running (measured by the motion of the air floating ball), whisking (measured by summed videographic motion energy within the whisker region), and pupil area. These three behavioral variables were correlated, but not identical (Figure 2A,B). Based on previous studies, we expected that these variables might differentially affect neural activity (McGinley et al., 2015; Reimer et al., 2016; Vinck et al., 2015).

Prediction from all three arousal-related variables explained more neural variance than the prediction from any one alone (Figure 2C; p<0.05, Wilcoxon rank-sum test). Each of the three variables contained unique information about the neural activity (Figure 2D). This improved prediction did not simply reflect the increased accuracy of averaging three noisy measures of a one-dimensional quantity. To show this, we found the single combination of weights that best predicted the activity of all neurons in a population, using reduced-rank regression (see Methods) (Figure 2E). This one-dimensional prediction weighted all three regressors positively (Figure 2F). The one-dimensional model yielded a poorer prediction than allowing each neuron to be predicted from a unique weight combination (Figure 2G). It follows that different neurons were best predicted by differently-weighted combinations of the 3 behavioral variables. This shows that a single linear combination of running, whisking and pupil area cannot account for the full diversity of patterns seen in ongoing activity, suggesting that the influence of behavioral states on neural activity is at least three-dimensional. Also, this multidimensional influence was spread throughout cortical layers: the neurons most correlated with different behavioral variables were not localized in depth (Figure 2H).

Multidimensional neural activity

Ongoing population activity showed a complex and high-dimensional structure, beyond the single arousal-related dimension revealed by the first principal component, and beyond the three-dimensional structure highlighted in the previous analysis. This higher-dimensional activity could be visualized by replotting the raster diagram after vertically sorting neurons along a continuum so that nearby neurons showed strong correlations (Figure 3A, see Methods; see Figure S2 for all recordings). This sorting revealed a more complex and higher-dimensional structure than could be seen when sorting neurons by their weights on the first principal component (c.f. Figure 1H). Activity patterns changed smoothly along this continuum, and there were no hard boundaries identifying separate clusters of neurons. Nonetheless, activity patterns varied markedly between neurons that were distant in this continuum, and to analyze the data further, we divided the neurons into 30 groups along the sorting dimension, with neurons within a group showing stronger correlations than across groups. Most of the resulting groups exhibited periods of high activity, intermingled with periods of near-complete silence, but these switches occurred at different times for different groups. Neurons assigned to the same group were spatially distributed: the average within cluster distance was only 13% smaller than out-of-cluster distance (Figure 3B-D) (453 μm in-cluster distance versus 522 μm out-of-cluster distance, p<0.001).

To quantify the apparent multidimensional structure of population activity, we used a technique of "peer prediction" (Harris et al., 2003; Pillow et al., 2008). This method attempts to predict the activity of a target neuron from its simultaneously-recorded peers. Each neuron took a turn as target, and we summarized the activity of the remaining population (the target cell's "peers") by principal component analysis, excluding cells with somas located physically close to the target cell (within a 3d distance of 70 μm; see Methods). We predicted the target cell's activity from the top N principal components, and determined how prediction quality varied as a function of N using cross-validation (Figure 3H). We observed a peak in predictability at N=213+/-23 components (Figure 3I), confirming that ongoing activity was indeed multidimensional. This analysis represents a lower bound on the true dimensionality of the data: because the analysis is cross-validated, genuine dimensions of smaller variance may not be detected with the quantity of training data we have (see Methods). This method was on average able to explain 18.6% of the neuronal variance using an optimal multidimensional predictor; in contrast, one dimension alone explained 3.7% variance (Figure 3F). The predictability using this method was markedly better than the three-dimensional model of neural activity obtained from whisking, running and pupil area (Figure 2), which explained on average $\sim$4% of the neural variance. The remaining variance that is not predicted by peer neurons may be unpredictable due to Poisson-like neural spiking and the noise added by the two-photon recording method.

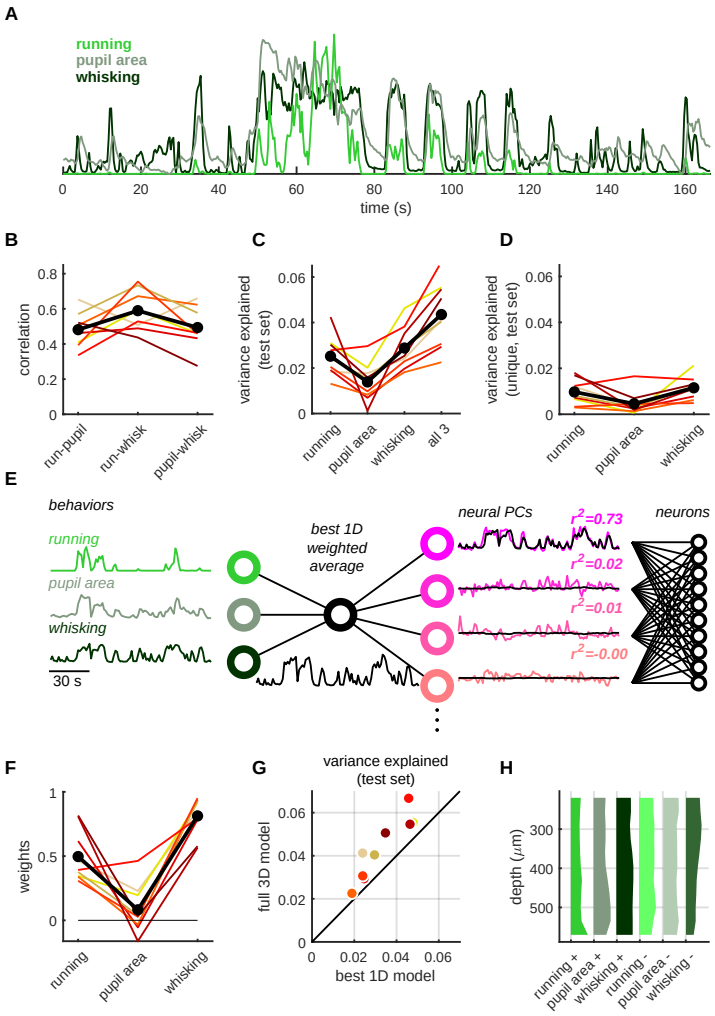


Figure 2. Predicting population activity from running, whisking and pupil area. (A) Time courses of running speed, pupil area, and whisker motion energy, for an example experiment. Y-axis scale: scaled units. (B) Correlation coefficients for each pair of these three behavioral traces. Each line denotes an individual experiment, black denotes average. (C) Mean fraction of single neuron variance predicted by each behavioral variable individually, or combined. (D) Unique contribution of each behavioral variable to prediction of single-neuron activity, defined as difference between full model and model excluding that predictor. (E) Schematic of reduced rank regression, which finds the single weighted average of running, pupil area and whisking that best predicts all neurons in a recording. (F) Weights of each predictor found by reduced rank regression. (G) Prediction quality for best 1D weighted average model, where all neurons are predicted by the same predictor combination, against prediction quality full model where different neurons are predicted by different combinations. Each point represents the mean predictablilty of all neurons recording in an experiment. (H) Distribution across depth in tissue of six non-overlapping groups of neurons with different behavioral correlates. Neurons were grouped according which behavioral variable they best correlated with, and the sign of this correlation.

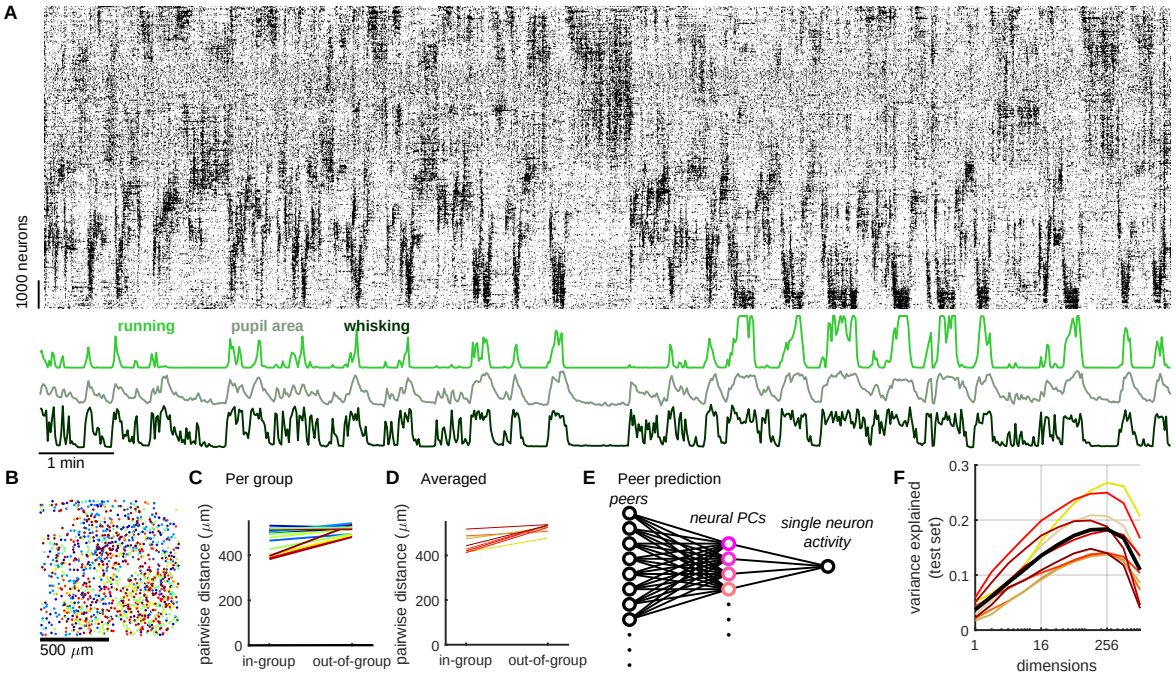


Figure 3. Multi-dimensional ongoing neural activity. (A) Raster representation of ongoing population activity, with neurons arranged vertically on a continuum such that nearby neurons had similar activity patterns. (B) Neurons were split into 30 groups along their position in the continuum. Spatial distribution of the 30 clusters (represented by pseudocolor), in a single example imaging plane. (C) Three-dimensional distances between neurons in the same or different clusters, for one experiment. Each line represents a cluster. (D) Average of (F) over all clusters. Each line represents an experiment. (E) "Peer prediction" method for estimating dimensionality of neural activity. Each neuron takes a turn as target cell, and principal component analysis is applied to the activity of its simultaneously recorded "peers", excluding spatially neighboring neurons. The activity of the target cell is predicted from these principal components, and prediction quality is assessed via cross-validation. (F) Average single neuron variance explained by peer prediction as a function of the number of principal components, for each experiment.

Spontaneous orofacial behaviors predict neural activity

The above analysis demonstrated that ongoing activity in visual cortex is high dimensional and behaviorally-correlated, but that only a small fraction of its structure can be explained by combinations of three readily definable behavioral variables (whisking, running and pupil area). To ask whether population activity in visual cortex encodes more behavioral information, we developed a method of unsupervised video analysis, and applied it to movies of the mouse face taken during the recordings (Figure 4A) (Powell et al., 2015). The method worked by computing the spatial distribution of facial motion energy at each moment in time (Figure 4B), and applying principal component analysis to it (Figure 4C). Using this method, we extracted a 1,000-dimensional summary of the motor actions visible on the mouse's face. Because these dimensions are orthogonal to each other, they need not necessarily correspond to individual motor actions (e.g. twitching of the nose, whisking, grooming, sniffing, chewing, wincing, etc), but are rather an abstract representation of such actions. Nevertheless, examining spatial maps of the PC weightings showed that the first PC corresponds to global facial motion, the second corresponds to motion of the whisker pad (contrasted with the rest of the face) and the third PC corresponds to nose movements, such as during sniffing (Figure 4C). Higher PCs generally did not have simple interpretations.

The video analysis method revealed that ongoing activity in visual cortex encodes at least 16 dimensions of motor-related information. To show this, we predicted neural activity from the motion energy PCs using reduced rank regression (Figure 4D); to avoid overfitting, we first predicted the top 128 neural principal components, and only through them predicted the neural activity (Figure 4D). A one-dimensional predictor extracted from facial motion using this unsupervised method captured the same amount of variance as the best one-dimensional combination of the explicitly-computed predictors considered above (whisking, running, pupil area), suggesting that both methods had independently found a similar 1D predictor of neural activity related to arousal (Figure 4E). However, prediction quality continued to increase up to 16 dimensions of videographic information (and in some recordings beyond), suggesting that at least 16 dimensions of motor information are encoded visual cortex (Figure 4F).

To visualize how the mouse's face predicts ongoing population activity, we again sorted the neurons by their embeddings in the 1D continuum (like in Figure 3A), and compared the raw neural activity with the model's prediction based on unsupervised videography (Figure 4G, see Figure S2 for all recordings). The model captured changes in the neural activity over a range of timescales (subsecond to tens of seconds), and accurately predicted multiple types of patterns in the population activity.

The prediction from orofacial behavior accounted for 45.6%$\pm$ 6% SE of the neural variance that could be explained by peer prediction (Figure 4H), substantially more than the three-dimensional model of neural activity using running, pupil area and whisking (Figure 4I). The information available in the explicitly-computed running, pupil area and whisking signals was redundant with that obtained by unsupervised videography: adding these predictors only increased the explained variance by less than 1% (Figure 4J). The timescale with which neural activity could be predicted from facial motor behavior was of the order 1s (Figure 4K), but even with a time delay of $\pm$ 2.44 seconds it was possible to explain half the variance that could be explained by instantaneous prediction (half-width at half-max). Prediction of neural activity from motor behaviors was optimal with a 1-frame time delay (400 ms), however, this was likely due to the dynamics of the calcium sensor (see electrophysiology results below).

Neurons in a majority of brain areas can be predicted from orofacial behaviors

Ongoing activity was predictable from orofacial movements uniformly throughout visual cortex (Figure S3). The lack of spatial localization raised a question: how far outside of our recording volume can neural activity be predicted from multidimensional orofacial behaviors? There are reports of running, whisking and pupil area modulating other cortical areas (Gentet et al., 2010; Peron et al., 2015; Schneider et al., 2014; Shimaoka et al., 2018), as well as thalamus (Erisken et al., 2014; Williamson et al., 2015) and superior colliculus (Ito and Feldheim, 2018), so these areas are also good candidates to be predicted by orofacial behaviors.

To record neurons in a large number of brain regions, we performed large-scale electrophysiological recordings, using 8 Neuropixels probes simultaneously. Neuropixels electrodes allow recording from 374 sites densely spaced across a depth of $\sim$4 mm (Jun et al., 2017). The probes were positioned to record from areas of neocortex, hippocampus, striatum, thalamus, midbrain, and various other nuclei (Figure 5A,B). We obtained such record-

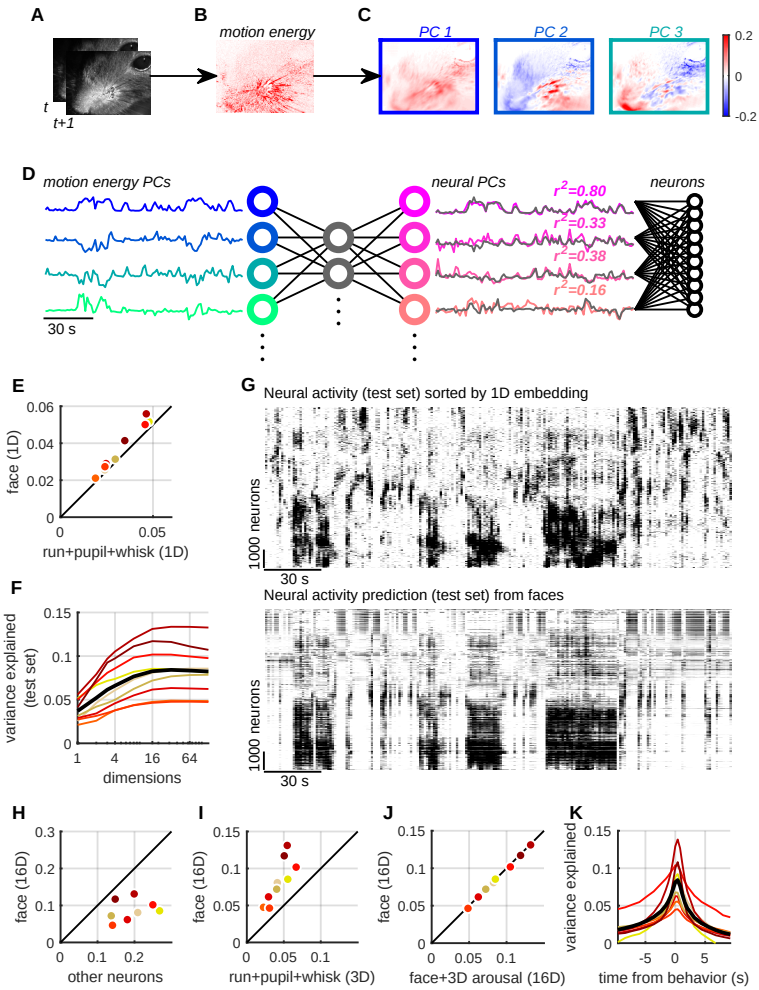


Figure 4. Multi-dimensional behavior predicts neural activity. (A) Consecutive frames from a video recording of a mouse's face. (B) Motion energy, computed as the absolute value of the difference of consecutive frames. (C) Spatial masks corresponding to the top three principal components (PCs) of the motion energy movie. (D) Schematic of reduced rank regression technique used to predict neural activity from the time traces of the motion energy PCs. (E) Prediction quality for the best single 1D combination of face PCs, plotted against best single combination of running, pupil area and whisking (c.f. Figure 2F). (F) Prediction quality as a function of the number of dimensions of facial information used by reduced rank regression. (G) Top: raster representation of ongoing neural activity in an example experiment, with neurons arranged vertically so that correlated cells are close together. Bottom: prediction of this activity from facial videography. (H) Prediction quality from multidimensional facial information, compared to peer prediction from other neurons. (I) Prediction quality of multidimensional facial information, plotted against against prediction from running pupil, and whisker. (J) Adding explicit running pupil and whisker information to facial features provides little improvement in neural prediction quality. (K) Prediction quality as a function of time lag used to predict neural activity from behavioral traces.

ings from three mice, and spike sorted the data using a modified version of Kilosort (Pachitariu et al., 2016a) that tracks changes in neuronal waveforms. In the three recordings, we were able to extract 2998, 2768 and 1958 units that were stable across $\sim$1 hour of ongoing activity, spread out over many different brain areas: cortex (visual: 628, sensorimotor: 475, frontal: 664, retrosplenial: 161), hippocampal formation (1371), striatum (353), thalamus (2882), midbrain (885) (Figure 5C). Neurons were on average more strongly correlated with neurons in the same area, but substantial inter-area correlations also existed suggesting non-localized patterns of neural activity (Figure 5D).

To investigate whether the behavioral modulation we observed in V1 could be seen throughout the brain, we applied the same analysis methods, after binning the spikes in overlapping bins of 1.2 seconds. The top principal component of forebrain-wide population activity again tracked a measure of arousal computed from the facial movements. All areas contained neurons positively and negatively correlated with this measure, however there were tendencies for some brain areas (such as thalamus) to contain a larger fraction of arousal-preferring neurons (Figure 5E).

As in visual cortex, ongoing population activity was high-dimensional. To visualize its structure, we again sorted the neurons along a 1D continuum so that nearby cells were more correlated (Figure 5F, same method as shown in Figure 3A). All brain areas contained a sampling of neurons from the entire continuum, suggesting that a multidimensional structure of ongoing activity is distributed throughout the brain, despite the fact that local correlations were stronger on average (Figure 5D). Applying a peer prediction analysis suggested that spontaneous activity was at least 128 dimensional, with 35% of the variance of individual neurons predictable from peer activity, and with modest variations between brain areas (Figure 5G).

To ask whether neurons across the forebrain encoded motor activity, we again extracted orofacial behavioral traces from video recordings of the mouse face, and used these to predict the neural activity through a low-dimensional bottleneck. As in V1, we found that 16 dimensions of orofacial behaviors predicted on average 43% of the variance explainable by peer prediction in these neurons (15.6% of the total variance) (Figure 5H). Predictability had a slight dependence on brain area: neurons in motor-related areas and thalamus were predicted best (19% of total variance, 49.3% of variance explainable by peer prediction), while neurons in visual and retrosplenial cortex were predicted least well (10.2% of total variance, 30.3% of variance explainable by peer prediction).

Neural activity was best predicted from video 44$\pm$20ms SE after the behavior, but the predicted variance again decayed slowly over time lags of multiple seconds (half width at half max of 2.95 s, Figure 5I). Taking advantage of the high temporal resolution of electrophysiological recordings, we reduced the analysis bin size from 1.2 seconds to 200 ms. The timescale of the prediction became substantially shorter, showing that the model can capture events on the order of 1 s (half width at half max of 1.07 s, Figure 5J). In addition, the prediction at 0 ms time lag was higher dimensional than the prediction at $\pm$5 s. These results suggest that predictability at long timescales (1-10 s) is dominated by the relatively slowly-changing arousal state, while at shorter timescales (<1 s) the prediction can take advantage of higher dimensions of behavioral activity.

Our results in visual cortex therefore largely extend to the rest of the forebrain. Motor-related information was spread across brain areas, including hippocampus, thalamus and striatum. Ongoing population activity was high dimensional, and a considerable fraction of this activity (43% of the variance explainable by peer prediction) could be predicted from the videographic monitoring of the faces of the mice.

Ongoing activity continues uninterrupted throughout sensory stimulation

We next asked how ongoing, behaviorally-related activity interacts with responses to sensory stimuli. For this analysis, we returned to imaging population activity in primary visual cortex, recording blocks of ongoing activity where a gray screen was shown constantly, interspersed with blocks of visual stimulation with flashed natural images, taken from the ImageNet database (Deng et al., 2009) and presented at an average rate of 1 image per second. The $\sim$1mm^2 areas of visual cortex we recorded from were chosen to be at retinotopic locations matched to the locations of the stimulus-display screens.

During stimulus presentation, the mice continued to exhibit the same spontaneous behaviors as in darkness: running, whisking, grooming, etc. The face motion energy traces continued unperturbed throughout the visual stimulus presentations (Figure 6A), with a similar distribution of power across the top videographic principal compo-

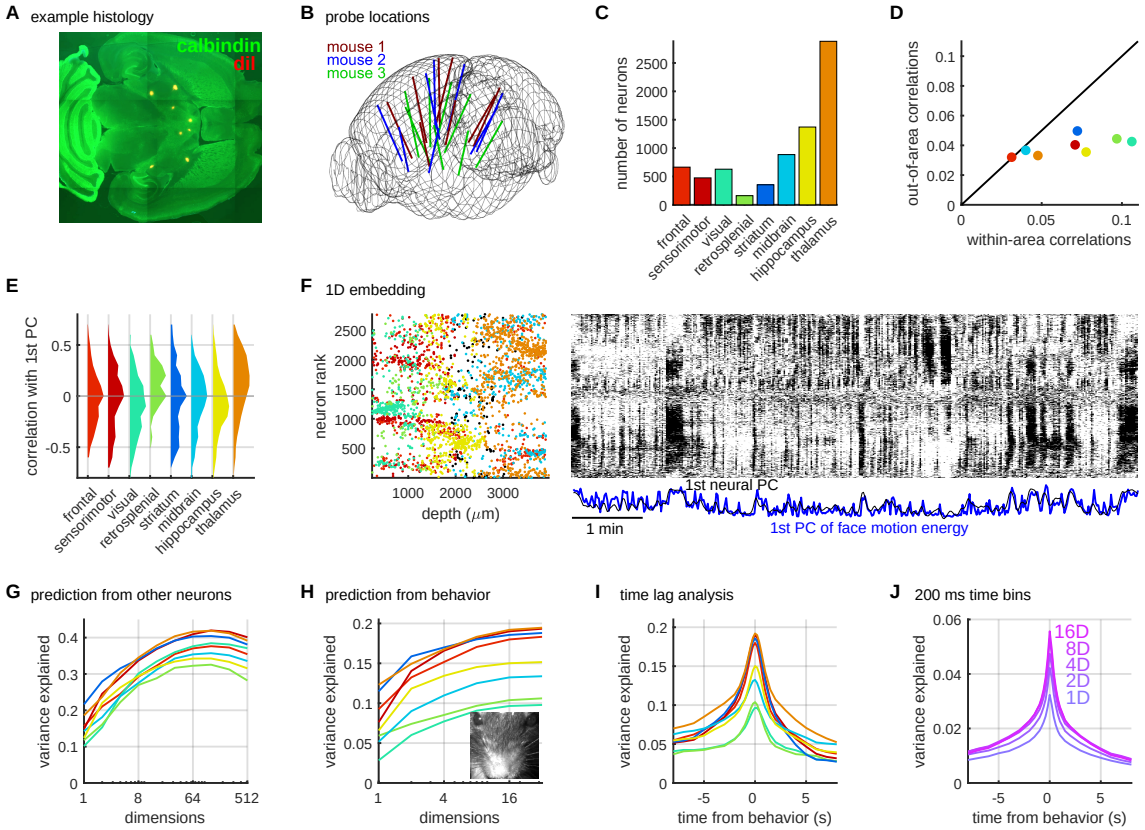


Figure 5. Behaviorally-related activity across the forebrain in recordings with 8 Neuropixels probes. (A) Example histology slice showing orthogonal penetrations of 8 electrode tracks through a calbindin-counterstained horizontal section. (B) Reconstructed probe locations of recordings in three mice. (C) Number of neurons recorded in each area, as determined by alignment to the Allen common coordinate framework. (D) Mean correlation between cells in the same area, or between an area and the rest of the recorded areas. X-axis represents average correlation coefficient over all pairs of cells in a given area, from all recordings; color code as in panel C. Y-axis represents average correlation of pairs with one cell in that area and all the other cells in different areas. (E) Distribution of correlations with first videographic principal component (arousal measure), for each brain region. (F) Neurons were arranged along a vertical continuum such that nearby neurons have correlated ongoing activity. Left panel shows position on continuum (y-axis) vs. recording depth (x-axis), with neurons color coded by brain location as in C. Right panel shows raster representation of ongoing population activity for an example experiment. (G) Peer prediction of single neurons from all other neurons in the recording as a function of the number of principal components (test data). Each curve shows the average prediction quality for neurons in a particular brain area, color coded as in C. (H) Prediction quality of from orofacial behaviors as a function of dimensions of reduced rank regression. Each curve shows average prediction quality for neurons in a particular brain area. (I) Explained variance as a function of time lag between neural activity and behavioral traces. Each curve shows average for a particular brain area. (J) Same as I in 200ms bins, and averaged over all neurons. Each curve shows prediction for a different number of dimensions in the reduced rank regression.

nents during gray screen and stimulus presentation (Figure 6B). This lack of a behavioral response was despite the fact that the stimuli drove strong, robust responses in visual cortex (Figure 6A,C).

Stimuli and behavior were represented together in V1 as a mixed representation: there were not separate sets of neurons encoding stimuli and behavioral variables, but each neuron multiplexed a unique combination of sensory and behavioral information. To show this, we first computed the fraction of each neuron's variance explained by stimuli and by behavior and observed only a slightly negative relationship between the two (Figure 6D; $r = -0.18$, $p<0.01$ Spearman's rank correlation). We then sorted the neurons according to their response patterns, separately for sensory responses and behavioral responses, and found no relation between the ranks of the neurons in the two orderings (Figure 6E,F). There was no correlation between the pairwise distances of neurons in the two orderings (Figure 6E; $r = 0.003$, $p > 0.05$). In other words, neurons with similar stimulus responses did not tend to have more similar responses to behaviors.

Although representations of sensory and behavioral information were mixed together in the same cell population, the coding subspaces corresponding to these two types of information were largely orthogonal. We developed a method to estimate the activity each stimulus caused, over and above ongoing activity, in an unbiased way by analyzing responses to multiple repeats of the same stimuli (see Methods). This method revealed that the amount of stimulus-related variance was 7.7% in single cells on single trials, comparable to the size of the ongoing variance, measured by peer prediction at 11.0%. However, only 11% of this stimulus-related variance lay in the neural subspace predictable from behavior, with the remaining 89% being orthogonal to the behaviorally-related dimensions. Furthermore, the overlap of the stimulus-related and behavior-related subspaces was largely one-dimensional: 93% of the variance of the overlap was contained in a single dimension (Figure 6G). The one-dimensional subspace shared between the stimulus- and behavioral-related dimensions had largely positive weights onto all neurons (85% positive weights, Figure 6H). We obtained similar results for the relation of sensory activity to ongoing activity, defined by the top 32 principal components of the spontaneous firing without regard to its behavioral correlate: the space of ongoing activity contained 18% of the total stimulus-related variance, 92% of which was contained in one-dimension that again was positively weighted on nearly all neurons (89% positive weights). Thus, the similarity between population responses to visual stimuli and ongoing/behavioral activity reflects primarily the fact that both can cause increased activity of all cells together.

To visualize how the V1 population integrated sensory and behavior-related activity, we examined the projection of this activity onto three orthogonal subspaces: a multidimensional subspace modulated by only sensory information; a multidimensional subspace modulated by only behavioral information; and the one-dimensional subspace modulated by both (Figure 6I, J, and zoomed in Supplementary Figure S4). During gray-screen periods there was no activity in the stimulus-only subspace, but when the stimuli appeared this space became very active. Conversely, activity in the behavior-only subspace was present prior to stimulus presentation, and continued unchanged when the stimulus appeared. The one-dimensional subspace modulated by both stimulus and behavioral information showed an intermediate pattern: activity in this subspace was visible prior to stimulus onset, but increased when stimuli were presented.

These analyses indicate that the relation between behavioral tuning and sensory responses is largely restricted to a single dimension corresponding to an overall increase in population rate. Except for this one dimension, the ongoing behavior-related patterns of neural activity continued unperturbed throughout sensory stimulation. Furthermore, neurons were selective to all combinations of stimulus and behavioral tuning.

Discussion

By imaging the ongoing activity of $\sim$10,000 neurons in visual cortex we discovered that this activity is high-dimensional, encompassing at least 100 linear dimensions, and possibly many more. The largest dimension was related to arousal, and it modulated about half of the neurons positively and half negatively. Further dimensions, at least 16 of them, were related to motor information visible by facial videography. Together, these dimensions accounted for approximately 45% of the explainable variance in the ongoing activity of visual cortex. The dimensions encoding motor variables were largely orthogonal to those encoding visual stimuli, overlapping primarily along a single dimension, which coherently increased or decreased the activity of the entire population. Beyond this one dimension, the ongoing behavior-related patterns of neural activity continued unperturbed regardless of sensory stimulation.

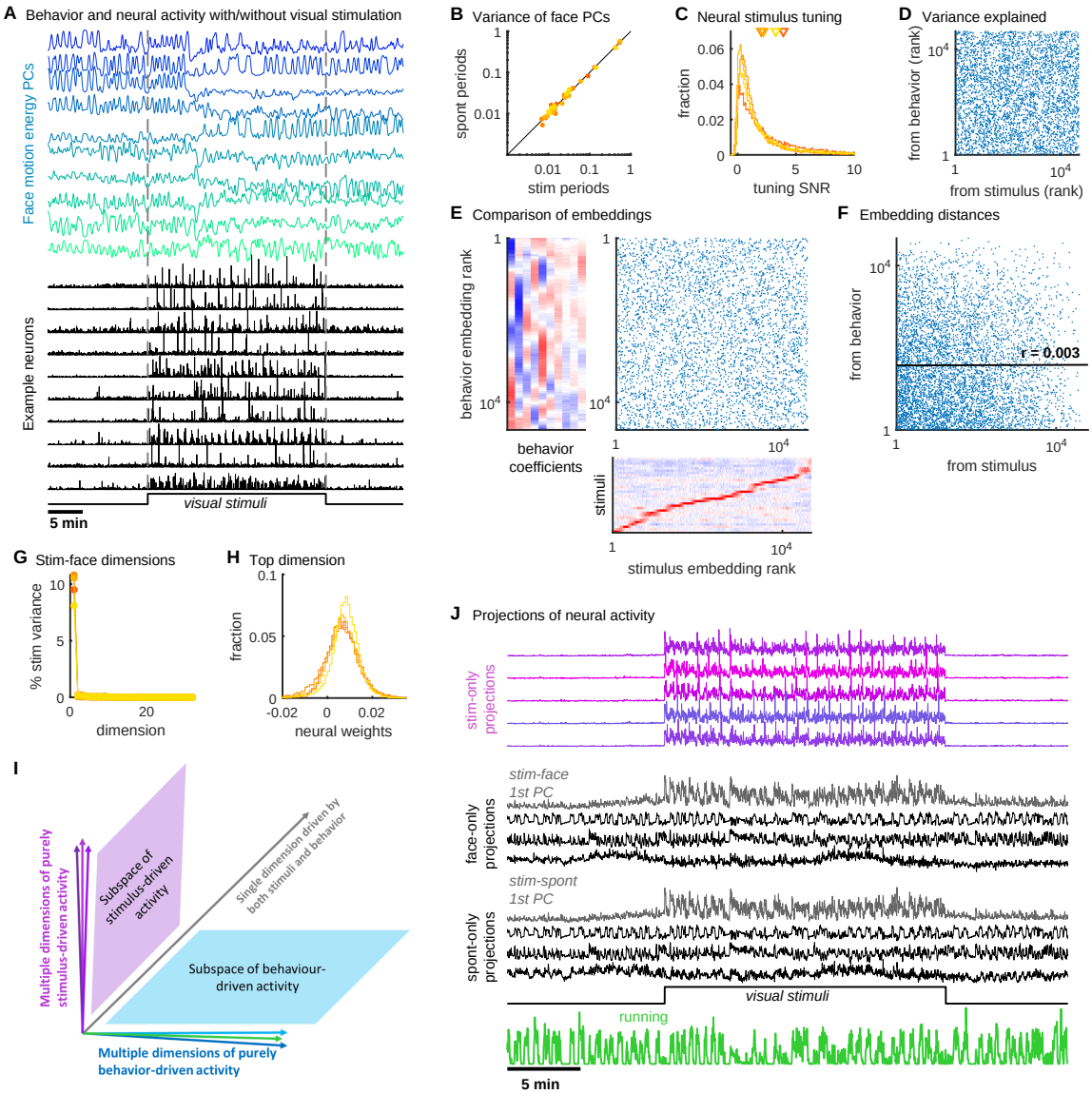


Figure 6. Behaviorally-related activity continues uninterrupted during the presentation of visual stimuli. (A) Principal components of facial motion energy (top) and firing of ten example V1 neurons (bottom), before, during and after a period of visual stimulus presentation. (B) Comparison of the variances of the PCs of motion energy with and without visual stimulation. Each point represents one principal component; data from all experiments shown as separate points of the same color. (C) Distribution of signal-to-noise ratio (SNR) of neural tuning to repeats of the 32 visual stimuli (see Methods). Each curve represents the SNR distribution of all cells in an experiment; arrows represent the mean for each experiment. (D) Comparison of explained variance from stimuli and from behavior videography for single neurons. X- and y-coordinates represent each neuron's rank order from least to most predictable by stimulus and behavior. (E) Comparison of tuning to behavior and stimuli. Each point is a neuron, and the x- and y-coordinates represent its rank in two separate 1D embeddings computed from stimulus responses, and from behavioral coefficients of the face prediction model (16D), respectively. (F) Pairwise comparison of visual-tuning similarity (x- axis) and behavioral-tuning similarity (y-axis). Each point represents a pair of cells; tuning similarities are defined as the pairwise distance in the 1D embeddings from E. (G) The distribution of stimulus variance in the neural space defined by the face prediction PCs from Figure 4. X-axis represents principal component number (specifically for stimulus-related activity within the space defined by face prediction); y-axis represents fraction of stimulus-related variance in this dimension. (H) Distribution of cells' weights on the single dimension of overlap between stimulus-related and face-related subspaces. Each curve represents the distribution of weights over all cells in an experiment. (I) Illustration of three sets of orthogonal dimensions in the space of firing patterns. Activity in multiple dimensions is driven by visual stimuli but not behavior (shades of magenta); multiple other dimensions are driven by behavior but not by stimuli (shades of cyan); a single dimension (gray; characterized in panels G,H) is driven by both. (J) Example of neural population activity projected onto these three sets of dimensions. Top: shades of magenta, projection onto stimulus-related dimensions. Middle: gray, projection onto single dimension related to both stimuli and behavior; black, projection onto dimensions related to behavior alone. Bottom: similar analysis for all ongoing dimensions, even if unrelated to facial behavior.

These results are not specific to visual cortex: we made similar observations in thousands of neurons recorded from multiple brain regions, ranging from visual, sensorimotor, frontal, and retrosplenial cortex to hippocampus, to striatum, thalamus, midbrain and various other nuclei. The resulting population activity showed many of the same features as those that we had observed in visual cortex: modulation by arousal and by many additional higher dimensions, many of them related to motor variables. The encoding of these motor variables in ongoing activity, therefore, was not restricted to visual cortex, but was seen across wide regions of the forebrain.

Our data are consistent with previous reports describing one-dimensional correlates of locomotion and arousal in visual cortex (Dipoppa et al., 2016; McGinley et al., 2015; Niell and Stryker, 2010; Pakan et al., 2016; Polack et al., 2013; Reimer et al., 2016; Saleem et al., 2013; Stringer et al., 2016; Vinck et al., 2015), but suggest these results were just glimpses of a much larger set of motor variables encoded by ongoing activity patterns. In fact, the 16 dimensions of motor activity we report may be a substantial underestimate. Our statistical analyses are conservative: because they are cross-validated, any dimensions below the noise floor would not be detected. Furthermore, our methods would fail to detect motor-related dimensions that are not visible on the face, or visible but not encoded linearly in the spatiotemporal patterns of the face motion energy.

A limitation of our approach is that it rests on an abstract description of motor actions: quantifying behaviors that are easily describable (whisking, running, pupil dilation) led to substantially poorer prediction of neural population activity than unsupervised videographic analysis. It is possible that more advanced methods of video analysis might produce both good predictions and interpretable predictors (Brown et al., 2013; Machado et al., 2015; Robie et al., 2017; Wiltschko et al., 2015). Indeed, an understanding of the function of sensory cortex will likely be impossible without measuring and understanding its relation to ongoing behavior, beyond easily-characterized measures such as running and pupil dilation.

Our finding that the dimensions of face-related and visually-evoked activity are largely orthogonal at first appears to contradict previous reports showing similarity of sensory responses to ongoing activity (Berkes et al., 2011; Han et al., 2008; Hoffman and McNaughton, 2002; Kenet et al., 2003; Luczak et al., 2009). We suggest three non-exclusive explanations for this apparent discrepancy. The first is that the experiments here looked at a slower timescale than most of these previous studies. Specifically, we binned the data into 1.2 s bins, while previous studies binned at 100 ms (Luczak et al., 2009), or even 2 ms (Berkes et al., 2011). The similarities between sensory responses and ongoing activity found by these studies at faster timescales could become very weak in our large time bins. The second explanation concerns the one dimension common to the spaces of ongoing and evoked activity, corresponding to a general increase in population firing. Some previous analyses showing similarity of spontaneous and evoked activity may have simply reflected the existence of this one dimension (Okun et al., 2012). The final explanation regards the approximately additive interaction of non-sensory variables with sensory responses. Because non-sensory variables account for around half the variance of population activity, the firing patterns recorded during stimulation will indeed show substantial similarity to ongoing activity. However, this would not imply that ongoing activity is a recapitulation of sensory events, just that it encodes nonsensory variables constantly throughout stimulation.

We observed that sensory and nonsensory variables are integrated in visual cortex in the form of a mixed representation. Rather than some neurons encoding sensory variables and others encoding nonsensory variables, almost every neuron represents both types of information, in different combinations. The brainwide representation of behavioral variables that we found suggests that information coding nearly anywhere in the forebrain is likely to be combined with behavioral state variables into a similar mixed representation. What benefit could this ubiquitous mixing of sensory and motor information provide? The most appropriate behavior for an animal to perform at any moment depends on the combination of available sensory data, ongoing motor actions, and purely internal variables such as motivational drives. An integration of sensory inputs with motor actions must therefore occur somewhere in the nervous system. Our data suggest that this integration happens as early as primary sensory cortex. While this may challenge the textbook view, it is not incompatible with neuroanatomy: primary sensory cortex receives not only innervation from neuromodulatory systems that carry state information, but also receives substantial inputs from higher-order cortices, which have been shown to encode fine-grained behavioral variables, at least in S1 (Petreanu et al., 2012). Beyond sensory cortex, there are indiscriminate connections between most brain areas (Gămănuţ et al., 2018; Harris et al., 2018), with most reconstructed single neurons extending axons at long range

throughout the brain (Economo et al., 2016), and with a high diversity of single neuron projections from the same area (Chen et al., 2018; Han et al., 2018). This pervasive whole-brain connectivity may form the mechanistic basis for the whole-brain coordinated activity we have reported here.

Acknowledgements

We thank Michael Krumin with assistance with the two-photon microscopes, and Andy Peters for comments on the manuscript.

This research was funded by Wellcome Trust Investigator grants (095668, 095669, 108726, and 205093) and by a grant from the Simons Foundation (SCGB 325512). CS was funded by a four-year Gatsby Foundation PhD studentship. MC holds the GlaxoSmithKline / Fight for Sight Chair in Visual Neuroscience. CS and MP are now funded by HHMI Janelia. NS was supported by postdoctoral fellowships from the Human Frontier Sciences Program and the Marie Curie Action of the EU (656528).

Experimental methods

All experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986). Experiments were performed at University College London under personal and project licenses released by the Home Office following appropriate ethics review.

Preparation for two-photon calcium imaging in visual cortex

The imaging methods were similar to those described elsewhere (Dipoppa et al., 2016). Briefly, surgeries were performed in adult mice (P35–P125) in a stereotaxic frame and under isoflurane anesthesia (5% for induction, 0.5-1% during the surgery). We used mice bred to express GCaMP6s in excitatory neurons (EMX-CRE x Ai94 GCaMP6s, CamKII x tetO GCaMP6s, and Rasgrf-CRE x Ai94 GCaMP6s), or mice bred to express tdTomato in GAD+ inhibitory neurons (GAD-Cre x tdTomato). During the surgery we implanted a head-plate for later head-fixation, made a craniotomy of 3-4 mm in diameter with a cranial window implant for optical access, and, in Gad-Cre x tdTomato transgenics, performed virus injections with a beveled micropipette using a Nanoject II injector (Drummond Scientific Company, Broomall, PA 1) attached to a stereotaxic micromanipulator. We used AAV2/1-hSyn-GCaMP6s, which was acquired from University of Pennsylvania Viral Vector Core. Injections of 50-200 nl virus (1-3 x1012 GC/ml) were targeted to monocular V1, 2.1-3.3 mm laterally and 3.5-4.0mm posteriorly from Bregma. To obtain large fields of view for imaging, we typically performed 4-8 injections at nearby locations, at multiple depths ($\sim$500 μm and $\sim$200 μm).

Data acquisition

We recorded optically the neural activity of head-fixed awake mice implanted with 3-4 mm cranial windows centered over visual cortex. We obtained $\sim$10,000 neurons in all recordings.The recordings were performed using multi-plane acquisition controlled by a resonance scanner, with planes spaced 30-35 μm apart in depth. 10 or 12 planes were acquired simultaneously at a scan rate of 3 or 2.5 Hz. The mice were free to run on an air-floating ball and were surrounded by three computer monitors. We either recorded in full darkness (monitors off), with a gray background or presented visual stimuli on these monitors arranged at 90° angles to the left, front and right of the animal, so that the animal's head was approximately in the geometric center of the setup.

For each mouse imaged, we typically spent the first imaging day finding a suitable recording location, where the following three conditions held:

- the GCaMP signal was strong, in the sense that clear transients could be observed in large numbers of cells
- a large enough field of view could be obtained, to result in 10,000 neuron recordings,
- the receptive fields of the neuropil were clearly spatially localized on our three monitors.

In animals for which there was a choice over multiple valid recording locations, we chose either: 1) a horizontally and vertically central retinotopic location or 2) a lateral retinotopic location, at 90° from the center, but still centered vertically. We did not observe differences related to retinotopic location (central or lateral), and thus pooled data across recording locations. We also did not observe significant differences between recordings obtained from GCaMP transgenic animals, or from virus injections, nor between recordings made in complete darkness or with a gray screen. Thus, we pooled data over all conditions.

Visual stimuli

We presented 96 repetitions of 32 flashed natural images, covering all three screens. The images were manually selected from the ImageNet database, from ethologically-relevant categories: "birds", "cat", "flowers", "hamster", "holes", "insects", "mice", "mushrooms", "nests", "pellets", "snakes", "wildcat". We chose images if the subjects tended to fill out the image (less than 50% of the image was a uniform background), and if the images contained a balanced mixture of low an high spatial frequencies. The images were flashed for 0.5 sec with a randomized inter-stimulus interval between 0.3 sec and 1.1 sec, during which a gray screen was presented.

Calcium imaging processing

The pre-processing of all raw calcium movie data was done using a toolbox we developed called Suite2p, using the default settings (Pachitariu et al., 2016b). The software is available at `www.github.com/cortex-lab/Suite2P`.

Briefly, Suite2p aligns all frames of a calcium movie using 2D rigid registration based on a regularized form of phase correlation, subpixel interpolation and kriging. For all recordings we validated the inferred X and Y offset traces, to monitor any potential outlier frames that may have been incorrectly aligned. In a very small percentage of all recordings, frames that had artifacts were removed and the extracted traces were replaced with interpolated values at those frames. In all recordings, the registered movie appeared well-aligned by visual inspection. Next, Suite2p performs automated cell detection and neuropil correction. To detect cells, Suite2p computes a low-dimensional decomposition of the data, and uses the decomposition to run a clustering algorithm that finds regions of interest (ROIs) based on the correlation of the pixels inside them. The extraction of ROIs stops when the pixel correlations of new potential ROIs drops below a certain value, which is set as a fraction of the correlation in the high SNR ROIs; thus, it does not require the number of clusters to be set a priori. A further step in the Suite2p GUI classifies these ROIs as somatic or not, partially based on user input, which is used to train a classifier. The classifier reaches 95% estimation of somatic/non-somatic signals on this data (Pachitariu et al., 2016b), thus allowing us to skip the manual step altogether for most recordings. We note that the 5% errors might either be attributable to human labelling error, or to dendritic signals, which would nonetheless most likely reflect backpropagating APs, which primarily measures the spiking signal for deeper cells. Thus, there is little risk of ROIs potentially measuring non-somatic signals.

We took great care to compensate cellular fluorescence traces for the surrounding neuropil signal (Chen et al., 2013). This contamination is typically removed by subtracting out from the ROI signal a scaled-down version of the neuropil signal around the ROI; the scaling factor was set to 0.7 for all neurons. Importantly, for computing the neuropil signal, we excluded all pixels that Suite2p attributed to an ROI, whether this was a somatic or dendritic ROI. After neuropil subtraction, we further subtracted a running baseline of the calcium traces with a sliding window of 60 seconds to remove long timescale additive shifts in the signals. Spike deconvolution was then performed on the baseline-subtracted traces. We performed non-negative spike deconvolution with a fixed timescale of calcium indicator decay of 2 seconds (Friedrich et al., 2017; Pachitariu et al., 2017).

All of the processed deconvolved calcium traces are available on figshare (`https://figshare.com/articles/Recordings_of_ten_thousand_neurons_in_visual_cortex_during_spontaneous_behaviors/6163622`), together with the behavioral traces described in computational methods.

Recordings of face

Infrared LEDs (850nm) were pointed at the face of the mouse to enable infrared video acquisition in darkness. The videos were acquired at 30Hz using a camera with a zoom lens with an infrared filter (850nm, 50nm cutoff). The wavelength of 850nm was chosen so that it avoids the 970nm of the laser, while remaining outside the visual detection range of the mice.

8-probe Neuropixels recordings

Neuropixels electrode arrays (Jun et al 2017) were used to record extracellularly from neurons in three mice. The mice were: 73 days old, male, and Drd1a-Cre(-/-) (mouse 1); 113 days old, female, and TetO-GCaMP6s;Camk2a-tTa (mouse 2); 99 days old, male, and Ai32;Pvalb-Cre (mouse 3). In all cases, a brief (<1 hour) surgery to implant a steel headplate and 3D-printed plastic recording chamber (∼12mm diameter) was first performed. Following recovery, mice were acclimated to head-fixation in the recording setup. During head-fixation, mice were seated on a plastic apparatus with forepaws on a rotating

rubber wheel. Three computer screens were positioned around the mouse at right angles. On the day of recording, mice were again briefly anesthetized with isoflurane while eight small craniotomies were made with a dental drill. After several hours of recovery, mice were head-fixed in the setup. Probes had a silver wire soldered onto the reference pad and shorted to ground; these reference wires were connected to a Ag/AgCl wire positioned on the skull. The craniotomies as well as the wire were covered with saline-based agar. The agar was covered with silicone oil to prevent drying. Probes were each mounted on a rod held by an electronically position-able micromanipulator (uMP-4, Sensapex Inc.) and were then advanced through the agar and through the dura. Once electrodes punctured dura, they were advanced slowly ($\sim$10 μm/sec) to their final depth (4 or 5 mm deep). Electrodes were allowed to settle for approximately 15 minutes before starting recording. Recordings were made in external reference mode with LFP gain=250 and AP gain=500, using SpikeGLX software. The mice were in a light-isolated enclosure and, during the part of the recording considered here, the computer screens were black. Data were preprocessed by re-referencing to the common median across all channels (Ludwig et al., 2009).

Spike sorting the Neuropixels data

We spike sorted the data using a modification of Kilosort that tracks drifting clusters (Pachitariu et al., 2016a), which we will refer to as Kilosort2. This modification was necessary to obtain an automated algorithm, and the code will be made publicly available at or before the time of publication. Without the modifications, the original Kilosort and similar algorithms can split clusters according to drift of the electrode, which was a major confound for our behavioral-related analyses. Kilosort2 in comparison tracks neurons across drift levels and for longer periods of time (1 hour in our case). To further mitigate the effect of drift, we used a conservative threshold, excluding from further analysis units for which the maximal firing rate was more than twice their minimal firing rates, when the binned spikes were smoothed with a Gaussian-window filter with a standard deviation of 500 seconds. This excluded 20% of the units on average.

Computational methods

Correlations

Correlations for the two-photon neural data were computed in bins of 1.2-1.3 seconds (3 or 4 frames respectively for 12 and 10 plane recordings). To compute a shuffled distribution in Figure 1C, we circularly shifted each neuron's activity in time by a random number of bins (at least $\pm$1000), and correlated all the shifted traces with all the original traces.

Sorting neurons by correlation

In many raster plots, neurons were sorted vertically along a 1d continuum so that nearby neurons were most correlated. To do this, the binned activity of each neuron was z-scored, and for electrode data high-pass filtered by subtracting the Gaussian-filtered traces with a standard deviation of 100 seconds. The algorithm sought to order the neurons along a one-dimensional continuum according to the their firing patterns, such that each neuron's activity is most similar to the average of all nearby neurons' activity. This was achieved with a generalization of a scaled k-means clustering algorithm, where the clusters are ordered along a 1D axis and constrained to have similar means to their nearby clusters. For initialization, the neurons were ordered based on their weights onto the first principal component of the activity and then divided into 30 clusters of equal size along this ordering. We computed the mean activity of each cluster and smoothed this activity across clusters, with a Gaussian of standard deviation of 3 clusters. Then the correlation of each neuron with the smoothed activity of each cluster was computed. Neurons were reassigned to the cluster to which they were most correlated with. This process was then repeated for 75 iterations. The smoothing constant across clusters was held at 3 for the first 25 iterations and then annealed to 1 over the following 50 iterations. On the final pass, we upsampled the neurons' correlations with each cluster by a factor of 100 via kriging interpolation with a smoothing constant of 1 cluster. This allowed us to determine sub-integer assignments of neurons to clusters, resulting in a continuous distribution of neurons along the 1D axis of clustering algorithm.

Although the electrode data was high-pass filtered to compute sorting, we display the original raw activity in Figure 5.

Peer prediction: predicting single neuron activity from simultaneously recorded neurons

We binned the neural activity in time in bins of 1.2-1.3 s, and split the activity into two halves: a training half and a test half. Each half contained many temporally contiguous blocks of duration 72 s, and the blocks belonging to

training and testing sets were temporally interspersed but not overlapping. The neurons each took a turn as target for prediction.

For each target neuron we defined its "peers" as all neurons greater than 70 μm from the cell in 3D distance, in order to avoid peers that may get optical contamination of their signals from the target neuron. From these peers, we first computed the singular value decomposition of their neural activity from the training half, F_1:

$$[U\ S\ V^T] = \text{svd}[F_1].$$

We then predicted the target neuron activity from n singular value components on the training half using linear regression (f_1). The weights from the "peer" activity to the single neuron activity f_1 were

$$W_n = \left[(f_1 V_n S_n)((V_n S_n)^T (V_n S_n))^{-1}\right] U_n^T$$

where U_n, V_n, S_n are the top n singular vectors. Then the prediction of the single neuron activity on the testing half was

$$\hat{f}_2^n = W_n F_2.$$

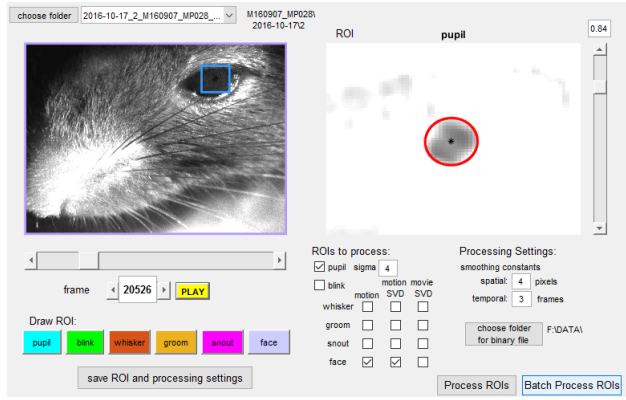
The variance explained was

$$v_{exp}^n = 1 - \frac{\text{var}\left[f_2 - \hat{f}_2^n\right]}{\text{var}\left[f_2\right]}.$$

We computed this variance explained for all target neurons and for increasing numbers of dimensions $n = 1, 2, 4, 8, 16, ..., 512, 1024$. The variance explained was averaged across all neurons predicted in Fig 2. Due to the high computational cost of this procedure, we only predicted a random subset of 10% neurons.

FaceMap: automated extraction of orofacial behaviors of mice

We developed a toolbox for videographic processing of orofacial movements of mice. The software is termed FaceMap, and is available at `www.github.com/carsen-stringer/FaceMap`. The processing time taken by the software scales linearly with the number of frames, and runs 4x faster than real-time on 30 Hz videos. A screenshot of the graphical user interface is shown below.



Motion processing of regions of interest

A graphical user interface allows for easily defining the areas of the face to be processed. The user can choose any region of the frame in which to compute the total motion energy, the SVDs of the motion energy or the SVDs of the raw frames. These regions of interest are assigned different names, but the computations for each is the same. The motion ROIs are called "groom", "whisker", "snout", and "face", and can be manually moved to overlap any region of the movie.

The motion energy G_T is computed as the absolute value of the difference between consecutive frames F_{T+1} and F_T, for all T:

$$G_T = |F_{T+1} - F_T|.$$

The full matrix G is thus number of pixels by timepoints and our goal was to (approximately) obtain its singular value decomposition. For the behavioral analyses described in this paper, we processed the whole frame and the whisker pad / eye area separately. The total motion of the region is simply the sum of non-white pixels that appear in the GUI panel on the right.

SVD processing of the movie and/or motion

In order to find the behavioral motifs in the data we reduced the dimensionality of the motion movies using SVD. The toolbox allows for computation of the SVD of the movie F, the motion G, or both. The computation is identical for F and G. We outline the procedure for G.

The matrices are too large to decompose in their raw form. We therefore computed the SVD in two stages: first we

computed the SVD of short temporal segments, then we concatenated the SVDs from different segments, scaled by the singular values, and recomputed their SVD. Each segment of frames is a matrix G_i. Since the number of pixels is very large (>1 million), we avoid computing the SVD of this matrix directly, and instead compute the time by time covariance matrix $G_i^T G_i$. We keep the top 200 eigenvectors V_i of this matrix, which are also the top 200 right singular vectors of G_i. We then compute the spatial projections of these components $U_i = G_i V_i$. Notice that U_i consists of the left singular vectors, scaled by the singular values. As such, the matrix U_i is a 200-dimensional summary of the segment G_i.

We then concatenated U_i for all segments of the movie, and re-computed the SVD:

$$[U\ S\ V^T] = \text{svd}([\hat{U}_1 ... \hat{U}_n]).$$

We kept the top 1,000 components of this U matrix as the spatial components of the face motion. We then projected the raw movies onto these spatial components, to obtain their temporal profiles:

$$W_{motion} = U^T G.$$

Pupil processing

FaceMap also computes the area and position of the pupil in the region of interest that the user defines. The minimum value in this region is subtracted from all pixels for robustness across illumination changes. In this region of interest, a user-adjusted threshold is used to keep only the darkest pixels, which correspond to the pupil. We compute the center of mass of these dark pixels as:

$$\mathbf{x}_{COM} = \frac{\mathbf{x}^T R(\mathbf{x})}{\sum_{\mathbf{x}} R(\mathbf{x})}$$

where $\mathbf{x}$ is the two-dimensional pixel location and R is the pixel's darkness level.

We compute the covariance Σ of a 2D Gaussian fit to the region of interest:

$$\Sigma = (\mathbf{x} - \mathbf{x}_{COM})(\mathbf{x} - \mathbf{x}_{COM})^T.$$

We then iterate 4 more times between re-selecting only pixels that are 2 standard deviations away from the center, and recomputing the Gaussian covariance fit. We keep as the outline of the pupil the ellipse that is 2 standard deviations from the center of mass.

Predicting neural activity from one-dimensional behavioral variables by standard linear regression

To predict neural activity from running, pupil diameter, and whisker movement, we first binned the neural activity and the behavior x in bins of 3 frames (1 or 1.2 seconds). As with the peer prediction analysis, we split the neural activity F and the behavior x into two halves in time, F_1 and F_2, x_1 and x_2. The behavioral variables $x_{1,2}$ were either single traces (running, whisking, pupil area), pairwise combinations of these, or all three traces together. We predicted F_1 from x_1 by linear regression, obtaining the weights a:

$$a = \left(x_1\, x_1^T\right)^{-1} \left(x_1\, F_1^T\right).$$

We used a to obtain the prediction on the second half of the recording and computed how much variance this prediction explained of the test data v_{exp}:

$$\hat{F}_2 = a^T x_2$$

$$v_{exp} = 1 - \frac{\text{var}\left[F_2 - \hat{F}_2\right]}{\text{var}\left[F_2\right]}.$$

Predicting neural activity from multi-dimensional behavioral variables by reduced-rank regression

We predicted the neural activity F from W, where W is a matrix of behavioral dimensions by time, using reduced-rank regression. Reduced-rank regression is a form of regularized linear regression, with the prediction weights matrix restricted to a specific rank (Izenman, 1975). Because the rank of the regression matrix is restricted, the number of parameters in the model is lower, making it more robust to overfitting. To further avoid overfitting, we first reduced the dimensionality of the neural activity F using PCA, and fit the reduced rank regression model to the top PCs only:

$$[U\ S\ V^T] = \text{svd}[F].$$

We kept 128 singular vectors of the singular value decomposition of F, and set $V = SV^T$. We split the recordings in half in time into a training set and a test set, thus splitting F into F_1 and F_2, V into V_1 and V_2, and W into W_1 and W_2.

We computed the reduced rank regression matrices A_n and B_n with rank n that minimize the expression

$$\min_{A_n, B_n} \left\| A_n B_n^T W_1 - V_1 \right\|^2 .$$

This expression can be minimized analytically (Izenman, 1975). We then reconstructed F_2 from the prediction $\hat{V}_2 = A_n B_n^T W_2$:

$$\begin{aligned} \hat{F}_2^n &= U \hat{V}_2 \\ &= U A_n B_n^T W_2. \end{aligned}$$

We then computed the variance explained v_{exp} for each rank n, up to 128 dimensions:

$$v_{exp}^n = 1 - \frac{\text{var}\left[F_2 - \hat{F}_2^n\right]}{\text{var}\left[F_2\right]}.$$

We also computed the variance explained for each of the principal components of the neural activity:

$$v_{exp}^n = 1 - \frac{\text{var}\left[V_2 - \hat{V}_2^n\right]}{\text{var}\left[V_2\right]}.$$

Stimulus variance computation

We consider the recorded neural response $f_k(n, i)$ of neuron n to stimulus i on repetition k. This can be rewritten as

$$f_k(n, i) = \mu(n, i) + \epsilon_k(n, i)$$

where $\mu(n, i)$ is the trial-averaged response of neuron n to stimulus i over an idealized, infinite number of repetitions of stimulus i, while $\epsilon_k(n, i)$ is the trial-to-trial variability, or "noise". We would like to estimate the tuning of $\mu(n, i)$ across stimuli i, and summarize this tuning by a scalar quantity

$$V_n = \mathrm{E}_i \left[(\mu(n, i) - \mathrm{E}_i[\mu(n, i)])^2 \right],$$

which we call the stimulus-driven variance. We use the notation $\mathrm{E}_i[X]$ to denote the expectation of the quantity X over an infinite number of stimuli i. By construction $\mathrm{E}_i\left[f_k(n, i)\right] = \mu(n, i)$ and $\mathrm{E}_i\left[\epsilon_k(n, i)\right] = 0$.

Suppose we have at least two repetitions of the stimuli for neuron n and all stimuli in set i. We assume that $\epsilon_k(n, i)$ does not depend on the repetition number, i.e. for all k, $\epsilon_k(n, i)$ are independent random samples from the same distribution. This condition can be approximately achieved in practice by separating the presentation of the stimulus repeats by tens of minutes. This is necessary in order to avoid temporally correlated noise. We thus assume the variance of the noise v_{noise} is the same in both repeats. We then subtract repeat 1 and 2:

$$\begin{aligned} f_1(n, i) - f_2(n, i) &= \mu(n, i) + \epsilon_1(n, i) - \mu(n, i) - \epsilon_2(n, i) \\ &= \epsilon_1(n, i) - \epsilon_2(n, i) \end{aligned}$$

Then the variance of this quantity allows the derivation of the noise variance:

$$\begin{aligned} \text{var}[f_1(n, i) - f_2(n, i)] &= \text{var}[\epsilon_1(n, i) - \epsilon_2(n, i)] \\ &= \text{var}[\epsilon_1(n, i)] + \text{var}[\epsilon_2(n, i)] \\ &= 2 v_{\text{noise}} \\ &= \text{var}[f_1(n, i)] + \text{var}[f_2(n, i)] \\ \Rightarrow v_{\text{noise}} &= \frac{1}{2}(\text{var}[f_1(n, i)] + \text{var}[f_2(n, i)]). \end{aligned}$$

And $v_{\text{signal}} = \text{var}[\mu(n, i)]$, and thus can also be derived:

$$\begin{aligned} \text{var}[f_1(n, i)] &= \text{var}[\mu(n, i) + \epsilon_1(n, i)] \\ &= \text{var}[\mu(n, i)] + \text{var}[\epsilon_1(n, i)] \\ &= v_{\text{signal}} + v_{\text{noise}} \\ \Rightarrow v_{\text{signal}} &= \text{var}[f_1(n, i)] - v_{\text{noise}}. \end{aligned}$$

We note that $\epsilon_k(n, i)$ could (and does) depend on the neuron and stimulus. However our derivation makes no assumptions on these dependencies.

To compute the tuning-related signal-to-noise ratio in our recordings, we divided the 96 repeats of 32 stimuli into two halves and computed the trial-averaged neural response from the first half of stimulus repeats ($f_1(n, i)$), and the trial-averaged neural response across the second half of stimulus repeats ($f_2(n, i)$). The signal-to-noise ratio is the ratio between v_{signal} and v_{noise} as computed above. This signal-to-noise ratio is non-zero when a neuron has responses to stimuli above its noise baseline.

We compute the signal variance of any set of traces in the same manner, for example to define the amount of signal variance in the neural subspace defined by behaviorally-predictable responses.

Stimulus versus ongoing activity analysis

Stimulus vectors

The trial-averaged stimulus responses were computed for all stimuli. This produced a matrix of neurons by number of stimuli (32). These trial-averaged stimulus components were projected onto the neural activity across all time (including the spontaneous activity). 8 traces (out of 32) are shown in Figure 6J and Figure S4.

Behavioral vectors

Reduced-rank regression was performed from the face motion energy PCs to the neural activity. The face components were defined as the columns of the A_n matrix, where n=32. These components were projected onto the neural activity across all time (including the visual stimulation periods). 8 traces are shown in Figure 6J and Figure S4.

Shared stimulus-behavior subspace

To find the shared stimulus and behavior subspace, we analyzed the covariance matrix between trial-averaged neural responses to stimuli, and the top behavioral components identified in A_n. This covariance matrix had dimensions 32 stimuli by 32 behavioral components matrix. We computed the singular value decomposition of this matrix, and projected the components back into the neural space to obtain the shared stimulus-behavior subspace. The weights of the first singular vector are plotted in Fig 6G. This procedure was performed on one third of the data, and the other two thirds were kept for quantifying the stimulus-related variance of the shared subspace, following the same procedure described above for single neurons.

Analyses on the Neuropixels data

All analyses were matched as much as possible to the corresponding analyses on the two-photon data described above: bin sizes were kept similar (1.2 s), smoothing constants were the same, the division of training and testing data was done in similar size bins. For peer prediction, we excluded neurons within plus minus 5 channels on the probe, to make sure that errors in spike sorting do not contribute to the predictability. For determining the top principal component, we first weakly high-pass filtered the data with a Gaussian windowed filter of standard deviation 100 seconds. We displayed the unfiltered binned spike trains sorted by this top PC. Similarly we high-pass filtered the data before running the manifold embedding algorithm, but for displaying the raster we still show the raw, unfiltered data.

References

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366.

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013):83–87.

Brown, A. E. X., Yemini, E. I., Grundy, L. J., Jucikas, T., and Schafer, W. R. (2013). A dictionary of behavioral motifs reveals clusters of genes affecting caenorhabditis elegans locomotion. *Proceedings of the National Academy of Sciences*, 110(2):791–796.

Chen, J. L., Pfäffli, O. A., Voigt, F. F., Margolis, D. J., and Helmchen, F. (2013). Online correction of licking-induced brain motion during two-photon imaging with a tunable lens. *The Journal of physiology*, 591(19):4689–4698.

Chen, X., Kebschull, J. M., Zhan, H., Sun, Y.-C., and Zador, A. M. (2018). Spatial organization of projection neurons in the mouse auditory cortex identified by in situ barcode sequencing. *bioRxiv*.

Cohen, M. R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819.

Constantinople, C. M. and Bruno, R. M. (2011). Effects and mechanisms of wakefulness on local cortical networks. *Neuron*, 69(6):1061–1068.

Crochet, S. and Petersen, C. C. (2006). Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nature neuroscience*, 9(5):608.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE.

Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., and Harris, K. D. (2016). Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *bioRxiv*, page 058396.

Ecker, A. S., Berens, P., Keliris, G. a., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–587.

Economo, M. N., Clack, N. G., Lavis, L. D., Gerfen, C. R., Svoboda, K., Myers, E. W., and Chandrashekar, J. (2016). A platform for brain-wide imaging and reconstruction of individual neurons. *Elife*, 5.

Engel, T. A., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., and Boahen, K. (2016). Selective modulation of cortical state during spatial attention. *Science*, 354(6316):1140–1144.

Erisken, S., Vaiceliunaite, A., Jurjut, O., Fiorini, M., Katzner, S., Busse, L., Reichardt, W., and Neuroscience, I. (2014). Article Effects of Locomotion Extend throughout the Mouse Early Visual System. *Current Biology*, 24(24):1–9.

Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data. *PLOS Computational Biology*, 13(3):e1005423.

Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll, R. A., Huang, Z. J., and Stryker, M. P. (2014). A cortical circuit for gain control by behavioral state. *Cell*, 156(6):1139–1152.

Gămănuţ, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., and Burkhalter, A. (2018). The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. *Neuron*, 97(3):698–715.

Gao, P., Trautmann, E., Yu, B. M., Santhanam, G., Ryu, S., Shenoy, K., and Ganguli, S. (2017). A theory of multineuronal dimensionality, dynamics and measurement. *bioRxiv*.

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. (2010). Membrane potential dynamics of gabaergic neurons in the barrel cortex of behaving mice. *Neuron*, 65(3):422–435.

Gentet, L. J., Kremer, Y., Taniguchi, H., Huang, Z. J., Staiger, J. F., and Petersen, C. C. (2012). Unique functional properties of somatostatin-expressing gabaergic neurons in mouse barrel cortex. *Nature Neuroscience*, 15(4):607–612.

Han, F., Caporale, N., and Dan, Y. (2008). Reverberation of recent visual experience in spontaneous cortical waves. *Neuron*, 60(2):321–327.

Han, Y., Kebschull, J. M., Campbell, R. A., Cowan, D., Imhof, F., Zador, A. M., and Mrsic-Flogel, T. D. (2018). The logic of single-cell projections from visual cortex. *Nature*.

Harris, J. A., Mihalas, S., Hirokawa, K. E., Whitesell, J. D., Knox, J., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., Feng, D., Gaudreault, N., Graddis, N., Groblewski, P. A., Henry, A., Ho, A., Howard, R., Kuan, L., Lecoq, J., Luviano, J., McConoghy, S., Mortrud, M., Naeemi, M., Ng, L., Oh, S. W., Ouellette, B., Sorensen, S., Wakeman, W., Wang, Q., Williford, A., Phillips, J., Koch, C., and Zeng, H. (2018). The organization of intracortical connections by layer and cell class in the mouse brain. *bioRxiv*.

Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424(6948):552.

Hoffman, K. and McNaughton, B. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science*, 297(5589):2070–2073.

Ito, S. and Feldheim, D. A. (2018). The mouse superior colliculus: An emerging model for studying circuit formation and function. *Frontiers in neural circuits*, 12:10.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of multivariate analysis*, 5(2):248–264.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232.

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961):954.

Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62(3):413–425.

Ludwig, K. A., Miriani, R. M., Langhals, N. B., Joseph, M. D., Anderson, D. J., and Kipke, D. R. (2009). Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of neurophysiology*, 101(3):1679–1689.

Machado, A. S., Darmohray, D. M., Fayad, J., Marques, H. G., and Carey, M. R. (2015). A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. *Elife*, 4.

McGinley, M. J., David, S. V., and McCormick, D. A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, 87(1):179–192.

Niell, C. M. and Stryker, M. P. (2010). Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron*, 65(4):472–479.

Okun, M. and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience*, 11(5):535.

Okun, M., Steinmetz, N. A., Cossell, L., Iacaruso, M. F., Ko, H., Barthó, P., Moore, T., Hofer, S. B., Mrsic-Flogel, T. D., Carandini, M., et al. (2015). Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515.

Okun, M., Yger, P., Marguet, S. L., Gerard-Mercier, F., Benucci, A., Katzner, S., Busse, L., Carandini, M., and Harris, K. D. (2012). Population rate dynamics and multineuron firing patterns in sensory cortex. *Journal of Neuroscience*, 32(48):17108–17119.

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Harris, K. D. (2016a). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, page 061481.

Pachitariu, M., Stringer, C., and Harris, K. (2017). Robustness of spike deconvolution for calcium imaging of neural spiking. *bioRxiv*.

Pachitariu, M., Stringer, C., Schröder, S., Dipoppa, M., Rossi, L. F., Carandini, M., and Harris, K. D. (2016b). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*, page 061507.

Pakan, J. M., Lowe, S. C., Dylda, E., Keemink, S. W., Currie, S. P., Coutts, C. A., and Rochefort, N. L. (2016). Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex. *eLife*, 5:e14985.

Parga, N. and Abbott, L. F. (2007). Network model of spontaneous activity exhibiting synchronous transitions between up and down states. *Frontiers in neuroscience*, 1:4.

Peron, S. P., Freeman, J., Iyer, V., Guo, C., and Svoboda, K. (2015). A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron*, 86(3):783 – 799.

Petreanu, L., Gutnisky, D. A., Huber, D., Xu, N.-l., O'Connor, D. H., Tian, L., Looger, L., and Svoboda, K. (2012). Activity in motor–sensory projections reveals distributed coding in somatosensation. *Nature*, 489(7415):299.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999.

Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nature Neuroscience*, 16(9):1331–9.

Powell, K., Mathy, A., Duguid, I., and Häusser, M. (2015). Synaptic representation of locomotion in single cerebellar granule cells. *Elife*, 4.

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., and Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature communications*, 7:13289.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–590.

Ringach, D. L. (2009). Spontaneous and driven cortical activity: implications for computation. *Current opinion in neurobiology*, 19(4):439–444.

Robie, A. A., Hirokawa, J., Edwards, A. W., Umayam, L. A., Lee, A., Phillips, M. L., Card, G. M., Korff, W., Rubin, G. M., Simpson, J. H., et al. (2017). Mapping the neural substrates of behavior. *Cell*, 170(2):393–406.

Saleem, A. B., Ayaz, A., Jeffery, K. J., Harris, K. D., and Carandini, M. (2013). Integration of visual motion and locomotion in mouse visual cortex. *Nature neuroscience*, 16(12):1864.

Schneider, D. M. and Mooney, R. (2015). Motor-related signals in the auditory system for listening and learning. *Current opinion in neurobiology*, 33:78–84.

Schneider, D. M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517):189–194.

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896.

Shimaoka, D., Harris, K. D., and Carandini, M. (2018). Effects of arousal on mouse sensory cortex depend on modality. *Cell Reports*, 22(12):3160 – 3167.

Stringer, C., Pachitariu, M., Steinmetz, N. A., Okun, M., Bartho, P., Harris, K. D., Sahani, M., and Lesica, N. A. (2016). Inhibitory control of correlated intrinsic variability in cortical networks. *Elife*, 5:e19695.

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726.

Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J. A. (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron*, 86(3):740–754.

Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., and Tononi, G. (2011). Local sleep in awake rats. *Nature*, 472(7344):443.

Williamson, R. S., Hancock, K. E., Shinn-Cunningham, B. G., and Polley, D. B. (2015). Locomotion and task demands differentially modulate thalamic audiovisual processing during active search. *Current Biology*, 25(14):1885–1891.

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., Abraira, V. E., Adams, R. P., and Datta, S. R. (2015). Mapping sub-second structure in mouse behavior. *Neuron*, 88(6):1121 – 1135.

Wimmer, R. D., Schmitt, L. I., Davidson, T. J., Nakajima, M., Deisseroth, K., and Halassa, M. M. (2015). Thalamic control of sensory selection in divided attention. *Nature*, 526(7575):705.

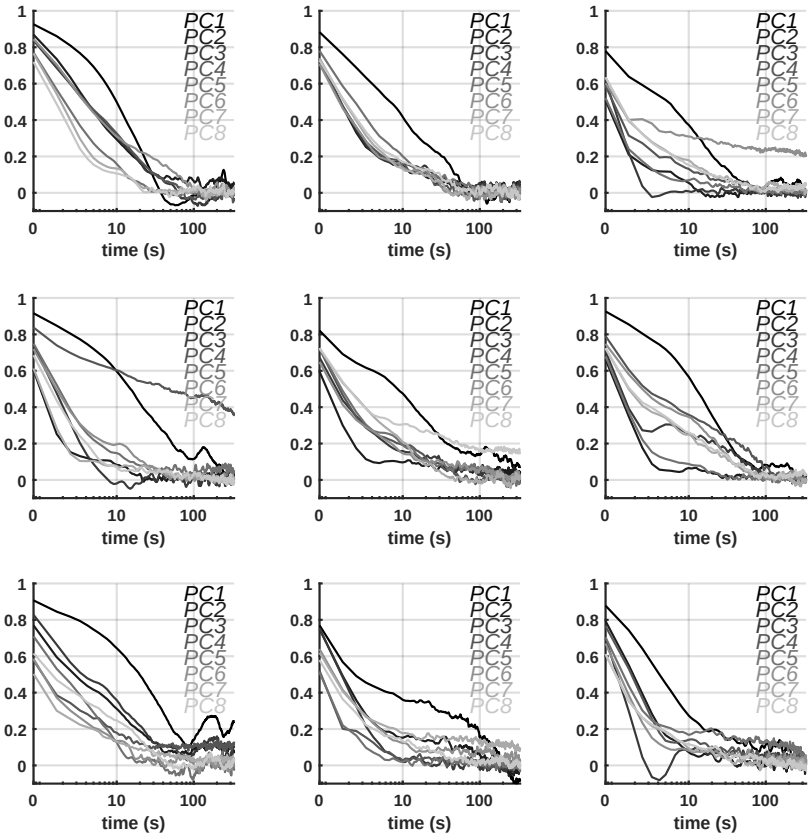


Figure S1. Temporal autocorrelation of ongoing neural activity. Each panel shows data for one recording; within each plot, each curve shows the temporal autocorrelation of a single principal component of ongoing population activity (1.2 second bins; log x-scale). The first plot is for the example recording shown in Figure 1.

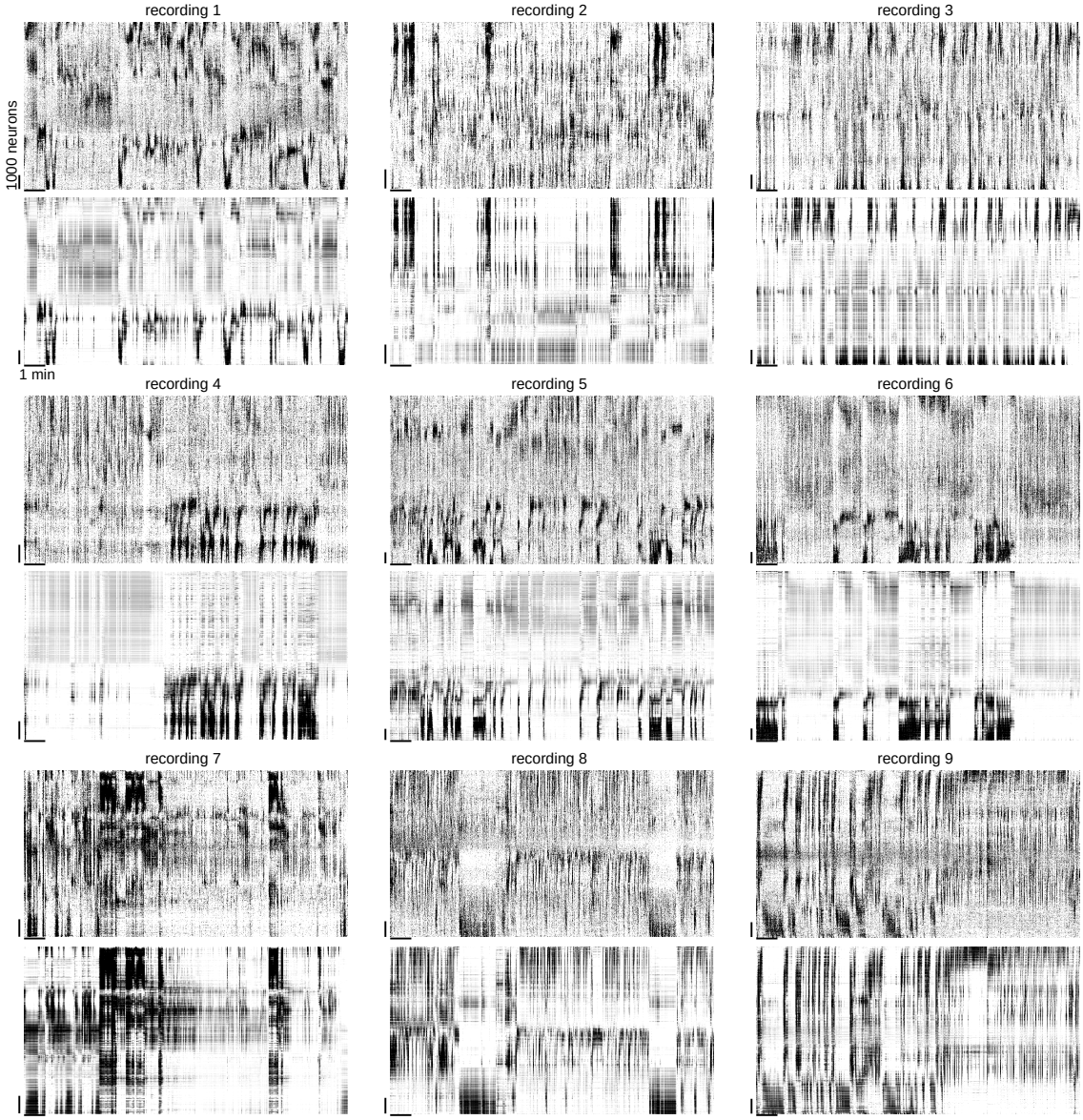


Figure S2. Nine additional examples of predictions of neural population activity from facial videography. For each panel, the top plot shows a raster diagrams with neurons sorted to place correlated neurons together. Bottom plots show predictions from facial videography.

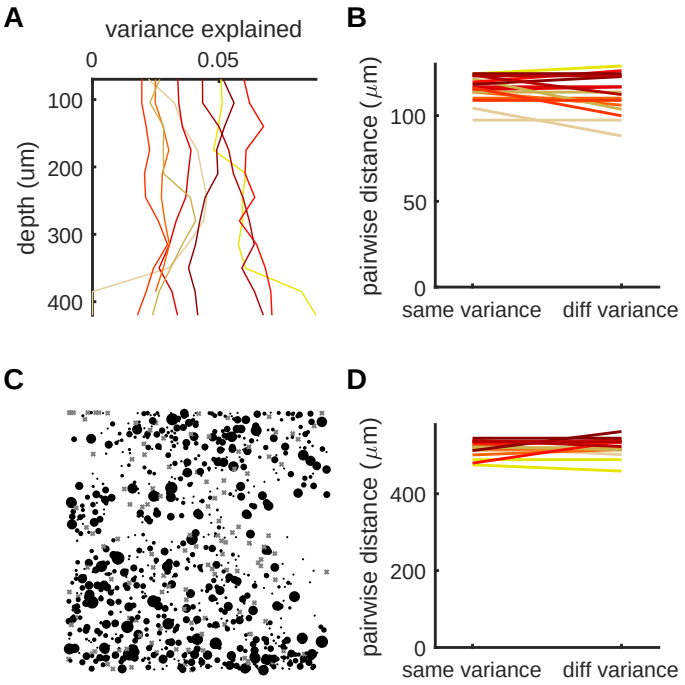


Figure S3. Predictability of neural activity from facial motion is uniform across V1. (A) Fraction of variance explained by facial motion, averaged over cells as a function of depth within cortex (16-dimensional reduced rank regression, cross-validated, 1.2 second bins). Each line represents a different experiment. (B) Neurons were split into two groups: low variance explained (<0.03) and high variance explained (>0.10). The average vertical distance between neurons in the same group ("same variance") was similar to the distance between neurons with different variance levels ("diff variance") (115 m vs 117 m). The lack of difference indicates that predictability from facial motion does not depend systematically on cortical depth. (C) Fraction of variance explained as a function of XY position in V1, for an example recording plane. The size of the dot is proportional to the explained variance, with crosses indicating negative test set variance. (D) Mean XYZ distance, for cells with similar vs. different fractions of variance explained. The lack of difference (516 μm vs 525 μm) indicates that variance explained does not depend systematically on XY position.

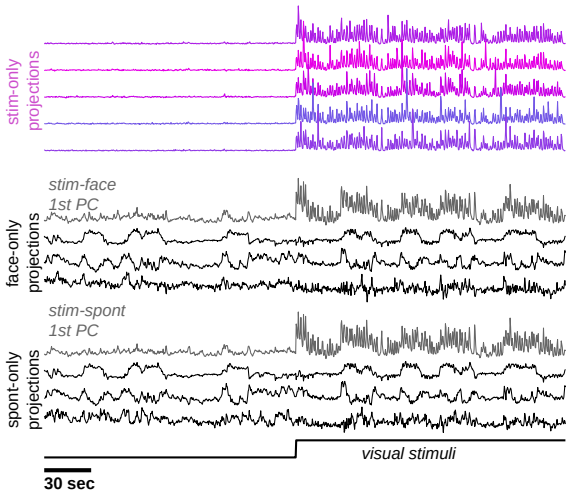


Figure S4. 10x zoom in at stimulus onset of Figure 6J. Projections of population activity along several types of dimensions: stimulus, stim-face, face, stim-spont, spontaneous.