Article

Neuron

Cortical Areas Interact through a Communication Subspace

Highlights

- Visual cortical areas interact through a communication subspace (CS)
- The CS defines which activity patterns in a source area relate to downstream activity
- The largest activity patterns in a source area are not matched to the CS
- The CS allows for selective and flexible routing of population signals between areas

Authors

João D. Semedo, Amin Zandvakili, Christian K. Machens, Byron M. Yu, Adam Kohn

Correspondence

jsemedo@cmu.edu

In Brief

Most brain functions require the selective and flexible routing of neuronal activity between cortical areas. Using paired population recordings from multiple visual cortical areas, Semedo et al. find a population-level mechanism that can achieve this routing, termed a communication subspace.

Semedo et al., 2019, Neuron *102*, 249–259
April 3, 2019
https://doi.org/10.1016/j.neuron.2019.01.026

CellPress

Neuron
Article

Cortical Areas Interact through a Communication Subspace

João D. Semedo,[1,2,3,*] Amin Zandvakili,[5] Christian K. Machens,[2,8] Byron M. Yu,[1,4,8] and Adam Kohn[5,6,7,8]
[1]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal
[3]Department of Electrical and Computer Engineering, Instituto Superior Técnico, Lisbon, Portugal
[4]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[5]Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA
[6]Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine, Bronx, NY, USA
[7]Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY, USA
[8]These authors contributed equally
*Correspondence: jsemedo@cmu.edu
https://doi.org/10.1016/j.neuron.2019.01.026

SUMMARY

Most brain functions involve interactions among multiple, distinct areas or nuclei. For instance, visual processing in primates requires the appropriate relaying of signals across many distinct cortical areas. Yet our understanding of how populations of neurons in interconnected brain areas communicate is in its infancy. Here we investigate how trial-to-trial fluctuations of population responses in primary visual cortex (V1) are related to simultaneously recorded population responses in area V2. Using dimensionality reduction methods, we find that V1-V2 interactions occur through a communication subspace: V2 fluctuations are related to a small subset of V1 population activity patterns, distinct from the largest fluctuations shared among neurons within V1. In contrast, interactions between subpopulations within V1 are less selective. We propose that the communication subspace may be a general, population-level mechanism by which activity can be selectively routed across brain areas.

INTRODUCTION

Interactions among brain areas are widely assumed to be essential to most brain functions, yet we are only beginning to understand how neurons in distinct brain areas mediate these interactions. Previous studies of inter-areal interactions have related the spiking activity of pairs of neurons in different areas (Nowak et al., 1999; Roe and Ts'o, 1999; Jia et al., 2013; Pooresmaeili et al., 2014; Oemisch et al., 2015; Ruff and Cohen, 2016), the spiking activity of a neuronal population in one area and a single neuron in another (Truccolo et al., 2010; Zandvakili and Kohn, 2015), the spiking activity of a neuron or group of neurons in one area and the local field potential (LFP) in another (Gregoriou et al., 2009; Salazar et al., 2012; Menzer et al., 2014; Arce-McShane et al., 2016; Wong et al., 2016), the LFPs recorded in different areas (Gregoriou et al., 2009; Salazar et al., 2012; Bosman et al., 2012; Jia et al., 2013; Roberts et al., 2013), or the trial-averaged population activity in distinct areas (Kaufman et al., 2014). These approaches have provided insight into how interaction strength changes with stimulus drive (Nowak et al., 1999; Jia et al., 2013; Roberts et al., 2013), attentional state (Gregoriou et al., 2009; Bosman et al., 2012; Oemisch et al., 2015; Ruff and Cohen, 2016), or task demands (Salazar et al., 2012; Kaufman et al., 2014; Menzer et al., 2014; Pooresmaeili et al., 2014; Arce-McShane et al., 2016; Wong et al., 2016).

These previous approaches fall short, however, of elucidating how the spiking activity of neuronal populations—the signals thought to encode information in the brain—is related across areas on a trial-by-trial basis (Semedo et al., 2014). Pairwise correlations, by definition, ignore structure not evident in the interactions between two individual neurons. LFPs lump the activity of spiking populations into a single summary signal and thereby risk losing much of the richness of area-to-area interactions (Jia et al., 2011; Ray and Maunsell, 2011). Trial-averaging allows one to study how mean signals (e.g., receptive field structure) are related, but not to understand how the moment-by-moment changes in activity in one area relate to those in another area (Saalmann et al., 2012; Salazar et al., 2012).

Here we leverage trial-to-trial co-fluctuations of V1 and V2 neuronal population responses, recorded simultaneously in macaque monkeys, to understand the nature of population-level interaction between cortical areas. Within individual brain areas, trial-to-trial fluctuations in activity have yielded important insight into the effects of attention (Cohen and Maunsell, 2009; Mitchell et al., 2009), learning (Gu et al., 2011; Jeanne et al., 2013), stimulus drive (Smith and Kohn, 2008; Churchland et al., 2010), and more. These fluctuations involve multiple dimensions of activity shared among neurons (Yu et al., 2009; Harvey et al., 2012; Ecker et al., 2014; Sadtler et al., 2014; Kaufman et al., 2015; Lin et al., 2015; Rabinowitz et al., 2015; Mazzucato et al., 2016; Williamson et al., 2016), as identified using dimensionality reduction (Cunningham and Yu, 2014). Each of these dimensions represents a characteristic way in which the activities of the recorded neurons covary (referred to as a population activity pattern). It is

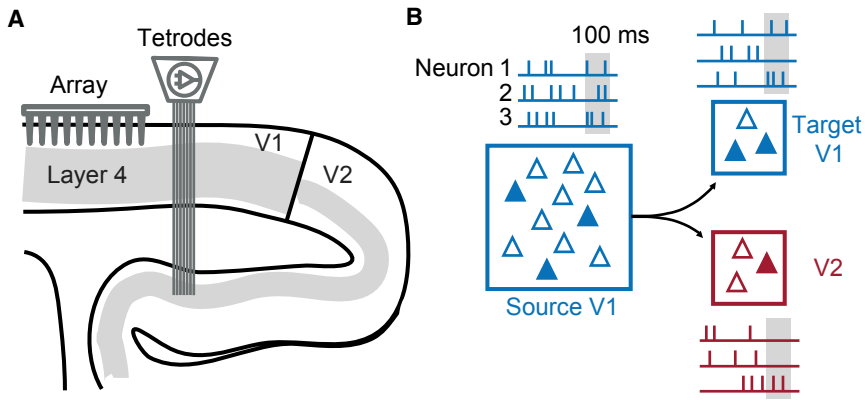

Figure 1. V1 and V2 Recordings

(A) Schematic showing a sagittal section of occipital cortex and the arrangement of the recording apparatus. We simultaneously recorded V1 population activity using a 96-channel Utah array and V2 population activity using a set of movable electrodes and tetrodes.

(B) We related activity of the same V1 source population to a target V1 population and a V2 population. In this illustration, each triangle represents a neuron and the filled triangles indicate active neurons. Spike counts were taken in 100 ms bins.

currently unknown whether all or only a subset of these dimensions are related across brain areas, and which dimensions are involved.

In anesthetized macaque monkeys, we find that interactions between V1 and V2 are similar in strength to those between subpopulations within V1, but that the structure of those interactions is strikingly distinct. V2 activity is related to a small subset of V1 population activity patterns, which are distinct from the largest shared fluctuations among V1 neurons. The selective routing of specific population activity patterns between V1 and V2 can be described by a low-dimensional communication subspace, which defines which activity patterns are effectively relayed between areas. We found that the same low-dimensional structure was present in paired V1-V4 recordings in awake animals, suggesting a general principle of inter-areal interactions. We propose that the communication subspace can be a population-level mechanism by which activity is selectively and flexibly routed between distinct neuronal populations.

RESULTS

We simultaneously recorded the activity of neuronal populations in the output layers (2/3-4B) of V1 (88 to 159 neurons; mean: 112.8) and their primary downstream target, the middle layers of V2 (24 to 37 neurons; mean: 29.4) (Felleman and Van Essen, 1991), in three sufentanil-anesthetized monkeys (Figure 1A). Neurons consisted of both well-isolated single units and small multi-unit clusters. The recorded V1 and V2 populations had retinotopically aligned receptive fields, maximizing the probability of direct feedforward interactions (Zandvakili and Kohn, 2015).

We measured neuronal activity as spike counts in 100 ms bins during the presentation of drifting sinusoidal gratings of different orientations. To study how neuronal activity in the two areas is related, we analyzed trial-to-trial response fluctuations to repeated presentations of each grating. These fluctuations involve spiking activity that could propagate between areas, and thus provides a useful window for understanding inter-areal interactions (Fries et al., 2001; Pesaran et al., 2008; Bosman et al., 2012; Saalmann et al., 2012; Salazar et al., 2012). Specifically, we subtracted the appropriate peri-stimulus time histogram from each single-trial response and then analyzed the residuals for each stimulus orientation (henceforth referred to as a dataset) separately.

To determine how V1-V2 interactions differ from interactions within V1, we divided the recorded V1 neurons into *source* and *target* populations (Figure 1B). For each dataset, we matched the target V1 population to the neuron count and firing rate distribution of the measured V2 population (see STAR Methods). We then related the activity of the same source V1 population separately to the activity of the target V1 population (V1-V1 interaction) and that of the V2 population (V1-V2 interaction).

Strength of Population Interactions

We first characterized V1-V2 interactions by measuring the degree to which response fluctuations were shared between pairs of neurons (i.e., noise correlations), as in previous inter-areal studies (Nowak et al., 1999; Ruff and Cohen, 2016). The vast majority of V1-V2 pairs had correlations between 0 and 0.2 (Figure 2A, red histogram; average correlation: 0.07 ± 0.06 SD). V1-V1 correlations were remarkably similar to those of V1-V2 pairs (Figure 2A, blue histogram; average correlation: 0.07 ± 0.06 SD; two-sided Monte Carlo permutation test, $p < 0.01$ for difference between V1-V1 and V1-V2). These weak correlations indicate that only a small fraction of a neuron's response variability can be explained by another individual neuron. Indeed, individual source V1 neurons could predict only 1.11% ± 0.03% and 1.35% ± 0.03% of the variability of the target V1 and V2 neurons, respectively (Figure 2B, solid lines).

We next asked how well the variability of the target V1 and V2 neurons could be explained by the source V1 population using multivariate linear regression (see STAR Methods). On average, the source V1 population predicted 15.2% ± 0.7% of the V2 variability (Figure 2B, red histogram), a substantial improvement over the performance afforded by individual V1 neurons. V1-V1 prediction quality was similar to that of the V1-V2 prediction (Figure 2B, blue histogram; 12.9% ± 0.8%; two-sided Monte Carlo paired permutation test, $p < 0.01$ for difference between V1-V1 and V1-V2).

To assess whether the performance of the regression models is reasonable in absolute terms, we implemented a basic model of population interactions using a linear feedforward network. Regression performance for these simulated data was similar to performance on the physiological data either when the target

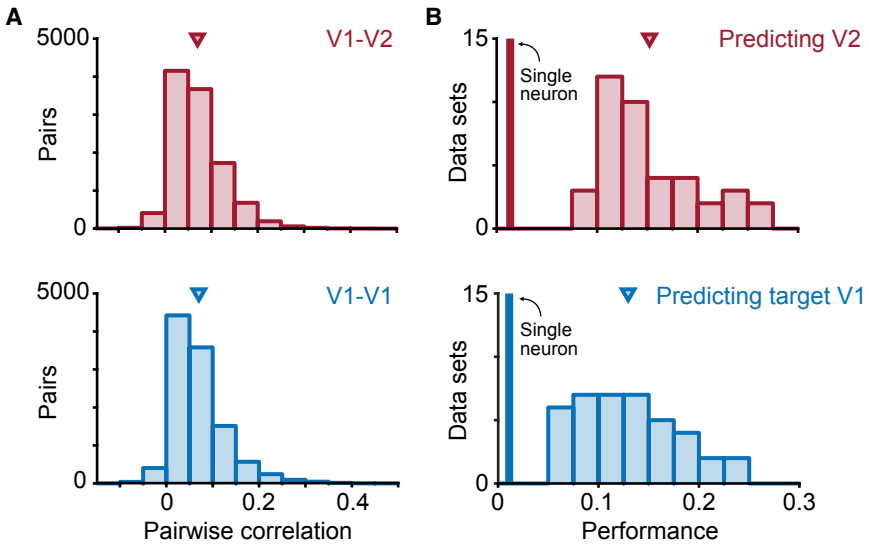

Figure 2. V1-V1 and V1-V2 Interactions Are Similar in Strength

(A) Pairwise correlation histograms for pairs of V1-V2 (red) and V1-V1 (blue) neurons. Triangles indicate average pairwise correlation. Total number of pairs in each histogram n = 10,944.

(B) Prediction performance for V1-V2 (red) and V1-V1 (blue). Prediction was performed using a single V1 neuron at a time (solid lines) or using the entire source V1 population (histograms; triangles indicate mean). Prediction performance for each dataset is defined as the average cross-validated r^2 across all selections of the target and source V1 populations.

population had Poisson variability or when the observed source population was a subset of the full input population (Figure S1).

In summary, both pairwise analysis and population-based regression models indicate that interactions between areas are similar in strength to those within a cortical area: fluctuations in the source V1 population can be used as effectively for predicting V2 activity as for predicting the fluctuations of other V1 neurons. We next asked whether the structure of these interactions is similar as well.

Structure of Population Interactions

Consider predicting the activity of a V2 neuron from a population of three V1 neurons using linear regression, as in the preceding section:

$$V2^k = w_1 V1_1^k + w_2 V1_2^k + w_3 V1_3^k$$

where $V2^k$ is the predicted activity of a V2 neuron on the kth trial, $V1_1^k$, $V1_2^k$ and $V1_3^k$ are the corresponding activities of the three V1 neurons on the same trial, and w_1, w_2, and w_3 are the regression weights. We can plot the activity of the V1 population on each trial as a point in a three-dimensional space, where each axis represents the activity of one of the V1 neurons (Figure 3A). The weights can be represented as a *regression dimension*, which captures which aspects of the V1 population activity are predictive of the V2 neuron's activity. Specifically, the location of the V1 activity along the regression dimension is the predicted activity of the V2 neuron (Figure 3A, shading).

In a basic multivariate regression model, each V2 neuron has its own regression dimension. These regression dimensions could, in principle, fully span the V1 activity space (Figure 3B). If this were the case, any fluctuation in V1 population activity would be predictive of the fluctuations of one or more V2 neurons (i.e., changing the V1 population activity would change the location of the activity along at least one of the regression dimensions). Alternatively, if the regression dimensions span only a subspace of the V1 activity space (shown as a plane in Figure 3C), certain V1 fluctuations (i.e., those orthogonal to the plane, Figure 3C, dashed line) would not be predictive of V2 fluctuations. We define *predictive dimensions* to be those which reside within the V1 subspace that is predictive of V2 fluctuations, and *private dimensions* as those which do not. The existence of private dimensions within the source population would allow for specific population activity fluctuations to be relayed downstream; any fluctuations along the private dimensions would be hidden from the target population.

To test whether our ability to predict V2 fluctuations involves only a subspace of V1 population activity, we used reduced-rank regression (Izenman, 1975; Kobak et al., 2016), a variant of linear regression in which the regression dimensions are constrained to lie in a low-dimensional subspace (see STAR Methods). If only a few dimensions of V1 activity are predictive of V2, then using a low-dimensional subspace should achieve the same prediction performance as the full regression model. For a representative dataset (Figure 4A), only two dimensions were needed to achieve a prediction performance that was indistinguishable from the full regression model (triangle). In contrast, when we applied the same analysis to the target V1 population, six dimensions of the source V1 population activity were needed to reach the performance of the full model (Figure 4B). Across all datasets, consistently fewer dimensions were needed to predict fluctuations in the V2 population (2.2 ± 0.1) compared to the target V1 population (3.5 ± 0.1; one-sided Monte Carlo paired permutation test, $p < 10^{-8}$; Figure 4C).

These results indicate that the V1 fluctuations that are predictive of V2 are confined to a small number of V1 dimensions. Notably, the number of dimensions needed to account for interactions between areas was smaller than the number of dimensions involved in interactions within an area.

The Influence of Target Population Dimensionality

A possible explanation for the lower-dimensional interaction between V1-V2 compared to within V1 is that the V2 population activity is itself less complex, or lower dimensional, than the target V1 activity. For example, if the measured V2 population consisted of neurons with identical responses, then predicting those responses would involve the same weighting of V1 activity (i.e., one predictive dimension). More generally, the number of

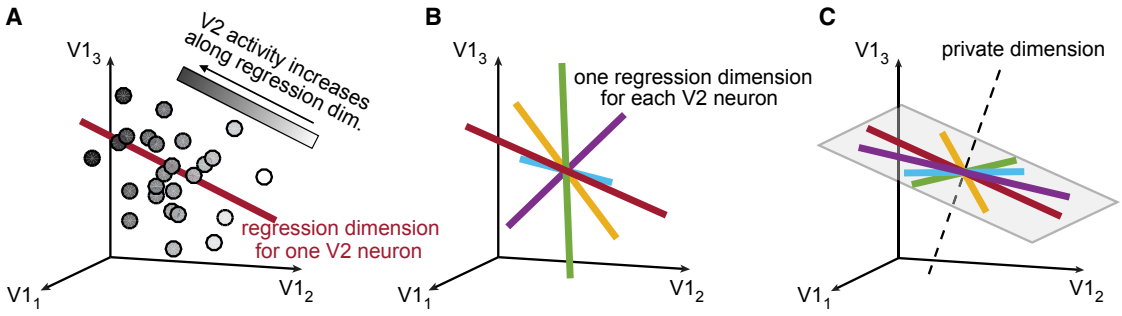

Figure 3. Illustration of a Low-Dimensional Interaction

(A) Graphical depiction of linear regression between a population of V1 neurons and one V2 neuron. Each circle represents the activity recorded simultaneously in V1 (three neurons) and V2 (one neuron) during one timestep (100 ms). The position of the circle represents the V1 population activity and its shading represents the activity of the V2 neuron. The activity of the V2 neuron increases along the regression dimension (red line).

(B) High-dimensional interaction. The regression dimensions for different V2 neurons (one regression dimension per V2 neuron) span the entire V1 population space.

(C) Low-dimensional interaction. The regression dimensions for different V2 neurons span a subspace of the V1 population space. In this illustration, all regression dimensions lie in a 2-dimensional subspace (the gray plane). The basis vectors for this subspace are called predictive dimensions. Thus, two predictive dimensions are sufficient to capture the between-area interaction. All dimensions that are not predictive of V2, and therefore lie outside of this subspace, are called private dimensions.

predictive dimensions will depend in part on the dimensionality of the target population activity. All else being equal, the lower the dimensionality of the target population activity, the smaller the number of predictive dimensions will be.

We used factor analysis to test whether the V2 population activity was lower dimensional than the target V1 population activity. Factor analysis identifies factors (or dimensions) which capture shared activity fluctuations among neurons (Santhanam et al., 2009; Yu et al., 2009; Churchland et al., 2010; Harvey et al., 2012; Ecker et al., 2014; Sadtler et al., 2014; Semedo et al., 2014; Williamson et al., 2016). This analysis revealed that the dimensionality of the V2 activity was higher than that of the target V1 activity (Figure 5A; 5.0 ± 0.2 for V2; 3.7 ± 0.1 for target V1; mean ± SEM; one-sided Monte Carlo paired permutation test, $p < 10^{-8}$). Thus, the smaller number of V2 predictive dimensions cannot be explained by the V2 population responses being less complex than the target V1 population responses.

To assess how the complexity of the target population influenced the dimensionality of the interactions, we compared the number of predictive dimensions to the dimensionality of the target population activity. For V1-V1 interactions, the number of predictive dimensions closely matched the dimensionality of the target population activity in each dataset (Figure 5B, blue points). Although these two estimates of dimensionality are based on different analyses, their similarity suggests that the number of V1 predictive dimensions is as large as possible, given the complexity of the target population responses. In contrast, for V1-V2 interactions, the number of predictive dimensions was consistently lower than the dimensionality of the target population (Figure 5B, red points).

The finding that the V1-V2 interaction is lower dimensional than the V2 population activity could arise because the reduced rank-regression model predicted the activity of only a few V2 neurons, ignoring the others. To assess this possibility, we refit the model after removing the three V2 neurons whose activity was best captured by the regression model, a number which corresponded to the largest number of V2 predictive dimensions we observed. Refitting the model after removing these neurons had little effect on the number of estimated predictive dimensions (2.20 ± 0.11 in the original analysis versus 2.20 ± 0.09 after removing top three V2 neurons; two-sided Monte Carlo paired permutation test, $p > 0.05$), indicating that the small number of predictive dimensions reflects a population-level effect.

We conclude that the difference in the number of V1 and V2 predictive dimensions cannot be explained by the complexity of the respective target population responses, but rather reflects the nature of the interaction between these areas. Whereas the V1-V1 interaction uses as many predictive dimensions as possible, the V1-V2 interaction is more selective and is confined to a small subspace of source V1 population activity, which we term a *communication subspace*.

Notably, this low-dimensional interaction structure was also present in simultaneous population recordings in V1 and V4 of awake monkeys (Figure S2), suggesting that the communication subspace is a general property of population-level interactions between brain areas.

Relationship to Source Population Activity

We next sought to understand the structure of the V1-V2 communication subspace. Specifically, we asked two related questions. First, we examined how the V1 and V2 predictive dimensions are related. Are the predictive dimensions for these target populations aligned or do they capture distinct activity fluctuations within the source V1 population? Second, we examined how the V1-V2 communication subspace relates to the structure of activity within the source V1 population. Is V2 activity predicted by the most dominant fluctuations within V1?

To characterize the relationship between V1 and V2 predictive dimensions, we made use of the fact that these dimensions are both defined within the source V1 activity space and capture the parts of the source population activity that are most relevant for predicting each target population. We therefore removed the

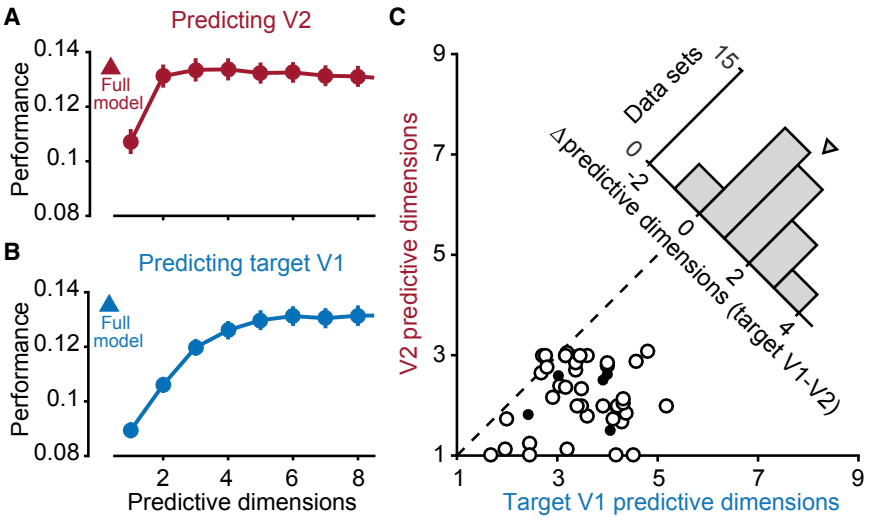

Figure 4. V1-V2 Interactions Use Only a Small Number of Dimensions

(A) Predicting V2 activity. The number of predictive dimensions (red circles; reduced-rank regression) needed to achieve full predictive performance (red triangle; ridge regression) is small (in this case, two dimensions). Across all datasets, reduced-rank regression achieved nearly the same performance as the full regression model (0.150 ± 0.007 for reduced-rank regression versus 0.152 ± 0.007 for the full regression model). The predictive performance slightly decreases with the number of predictive dimensions due to cross-validation. Error bars indicate SEM across cross-validation folds.

(B) Predicting target V1 activity. The number of predictive dimensions (blue circles; reduced-rank regression) needed to achieve full predictive performance (blue triangle; ridge regression) is large (in this case, six dimensions). Across all datasets, predictive performance was again similar for reduced-rank regression (0.123 ± 0.008) and the full regression model (0.129 ± 0.008).

(C) The optimal number of predictive dimensions is smaller for predicting V2 than target V1. Each open circle corresponds to one dataset. Filled circles indicate averages across datasets for each of the five sessions (see STAR Methods). Inset shows the difference between the optimal number of predictive dimensions needed when predicting the target V1 and V2 populations (target V1 minus V2).

source V1 activity along the different predictive dimensions (see STAR Methods) and assessed whether the remaining source activity could still be used to predict activity in the target V1 and V2 populations.

We first confirmed that our method for removing activity along predictive dimensions was effective. As expected, our ability to predict V2 fluctuations quickly decreased as we removed the source V1 activity along the dimensions that were most predictive of V2 (Figure 6A, filled circles). Across datasets, average predictive performance vanished when all source activity aligned with the V1-V2 communication subspace had been removed (Figure 6B, filled bars; average normalized performance: −0.005 ± 0.001; value is negative due to cross-validation).

In contrast, after removing the source V1 activity that fell along the top V1 predictive dimensions, we were still able to predict V2 fluctuations (Figure 6A, open circles). Across datasets, we retained a substantial fraction of our ability to predict fluctuations in V2 after removing the same number of V1 predictive dimensions as the number of predictive dimensions in the V1-V2 communication subspace (Figure 6B, open bars; average normalized performance: 0.24 ± 0.01; one-sided Monte Carlo paired permutation test, $p < 10^{-8}$). This indicates that the V2 predictive dimensions are not well aligned with the leading V1 predictive dimensions.

We obtained similar results when predicting fluctuations in the target V1 population (Figure 6C). Across datasets, predictive performance was significantly higher after removing source activity along V2 predictive dimensions (Figure 6D, open bars; 0.31 ± 0.01) than after removing activity along the same number of V1 predictive dimensions (Figure 6D, filled bars; 0.06 ± 0.01; one-sided Monte Carlo paired permutation test, $p < 10^{-8}$). Even after removing all source activity that fell within the V1-V2 communication subspace, we could still predict fluctuations in the target V1 population. Together, these analyses indicate that the V1-V2 and V1-V1 interactions not only differ in the number of predictive dimensions, but also involve different patterns of source population activity.

To understand how the V1-V2 communication subspace is related to the structure of the source V1 population activity, we used factor analysis to identify the dimensions of largest shared fluctuations within the source V1 population (termed *dominant dimensions*). We then predicted the activity of V2 neurons using linear regression based on the dominant dimensions only. This analysis is conceptually related to reduced-rank regression, which was used to identify the predictive dimensions. However, rather than identifying the subspace that is best for predicting fluctuations in the target population (as in reduced-rank regression), this analysis identifies a subspace that captures the largest shared fluctuations within the source population and then performs regression in that space.

If the dominant source V1 dimensions are able to predict V2 activity as well as the V2 predictive dimensions, for the same number of dimensions, this would indicate that the V1-V2 communication subspace preferentially involves the largest activity fluctuations of the V1 population. However, as shown for a representative dataset, the dominant V1 dimensions (Figure 7A, open circles) were not able to predict V2 as well as the predictive dimensions (Figure 7A, filled circles). In contrast, within V1, the predictive and dominant dimensions performed similarly (Figure 7B). Across datasets, predicting V2 fluctuations almost always required more dominant V1 dimensions than V2 predictive dimensions (Figure 7C, red). However, for target V1 fluctuations, dominant dimensions of the source V1 population were nearly as informative as the predictive dimensions (Figure 7C, blue; one-sided Monte Carlo permutation test for difference in the minimum number of dominant dimensions when predicting target V1 and V2, $p < 10^{-8}$ for 1 predictive dimension; $p < 10^{-8}$ for 2 predictive dimensions; $p < 0.01$ for 3 predictive dimensions).

These results indicate that the V1 predictive dimensions are aligned with the largest source V1 fluctuations. The V2 predictive

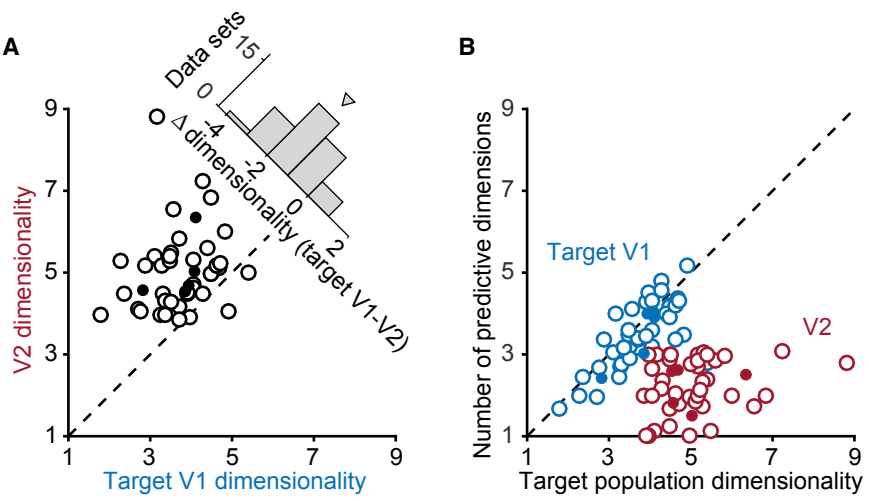

Figure 5. Low-Dimensional V1-V2 Interaction Is Not Due to Low-Dimensional V2 Activity

(A) Population activity is more complex in V2 than in target V1. Each open circle corresponds to one dataset. Filled circles indicate averages across datasets for each of the five sessions. Inset shows the difference between the dimensionality (target V1 minus V2) of the population activity in target V1 and V2.

(B) V1 and V2 interact through a communication subspace. The number of predictive dimensions identified for the V1-V2 interaction was always smaller than the dimensionality of the V2 population activity (red circles). The number of predictive dimensions required when predicting target V1 population activity was similar to the dimensionality of the target V1 population (blue circles). Each open circle corresponds to one dataset. Filled circles indicate averages across datasets for each of the five sessions.

dimensions, however, are distinct: not only are they less numerous, they are not well aligned with the V1 predictive dimensions nor with the largest source V1 fluctuations.

DISCUSSION

Nearly all previous studies of interactions between brain areas have used pairwise spike-spike, spike-LFP, or LFP-LFP analyses. Here we investigated the structure of interactions between areas at the level of neuronal population spiking responses. We found a striking difference in the nature of V1-V1 and V1-V2 interactions, summarized in Figure 7D. V2 activity was related to a small subset of population activity patterns in the source V1 population, and these patterns were distinct from the most dominant shared V1 fluctuations. In contrast, more activity patterns in the source V1 population were relevant for predicting the activity of other V1 neurons, and the dominant fluctuations in the source population were the most predictive. Interactions between areas are thus defined by a communication subspace: V1 activity that lies within the communication subspace is communicated with V2, whereas V1 activity that lies outside this subspace is not.

Our analyses were designed to ensure a fair comparison of V1-V1 and V1-V2 interactions. First, we used the same V1 population to predict target V1 and V2 responses, ruling out any potential differences in the source population. Second, we matched the sizes of the target V1 and V2 populations as well as their firing rate distributions, ruling out differences in these basic target population properties. Third, we were able to predict fluctuations in the target V1 and V2 populations equally well (Figure 2), so our results cannot be attributed to differences in the strength of V1-V1 and V1-V2 interactions. Finally, the spatial receptive fields of both the target V1 and V2 population overlapped those of the source population, and subtle variations in alignment could not explain the differences between V1-V1 and V1-V2 interactions (Figure S3).

It is important to note that the estimated number of predictive and dominant dimensions likely depends on the number of recorded neurons and trials (Williamson et al., 2016). Accordingly, our results do not define the dimensionality of V1-V2 interactions in absolute terms; rather, they indicate that those interactions are low dimensional relative to V1-V1 interactions. We found that if we analyzed only a portion of the recorded populations or trials, the difference between V1-V1 and V1-V2 interactions was less prominent (Figures S2D–S2F). Thus, with larger datasets, the difference between these interactions is likely even larger than that we identified.

Dimensionality reduction analyses have provided important insights into neuronal population activity structure and its function (see Cunningham and Yu, 2014, for a review). However, such analyses have been applied almost uniquely to population responses recorded in a single brain area, rather than to the study of interactions between areas, as we have done. Two important recent studies have investigated the relationship between activity in motor cortex and muscles (Kaufman et al., 2014; Elsayed et al., 2016). They found that preparatory motor activity avoids the potent (i.e., predictive) dimensions which relate cortical activity to muscles during movement, akin to our finding of private dimensions for V1-V2 interactions. Our work builds upon the strength of those studies by relating trial-to-trial fluctuations in directly connected neuronal populations (i.e., those with functional alignment and in specific cortical laminae; Zandvakili and Kohn, 2015). In addition, we studied the difference in interactions within and between areas, as well as the relationship between predictive and dominant dimensions in the source population.

V2 likely performs non-linear operations on inputs received from V1 (Freeman et al., 2013; Yu et al., 2015). Our approach to understanding V1-V2 interactions was to study local fluctuations around different set points (i.e., the trial-to-trial variability around the mean responses to a particular grating)—which function effectively as local linear perturbations in the non-linear transformation between V1 and V2. Our use of trial-to-trial fluctuations is consistent with most previous studies of inter-areal interactions (Fries et al., 2001; Pesaran et al., 2008; Bosman et al., 2012; Saalmann et al., 2012; Salazar et al., 2012), although these have used entirely distinct analyses such as spike-field

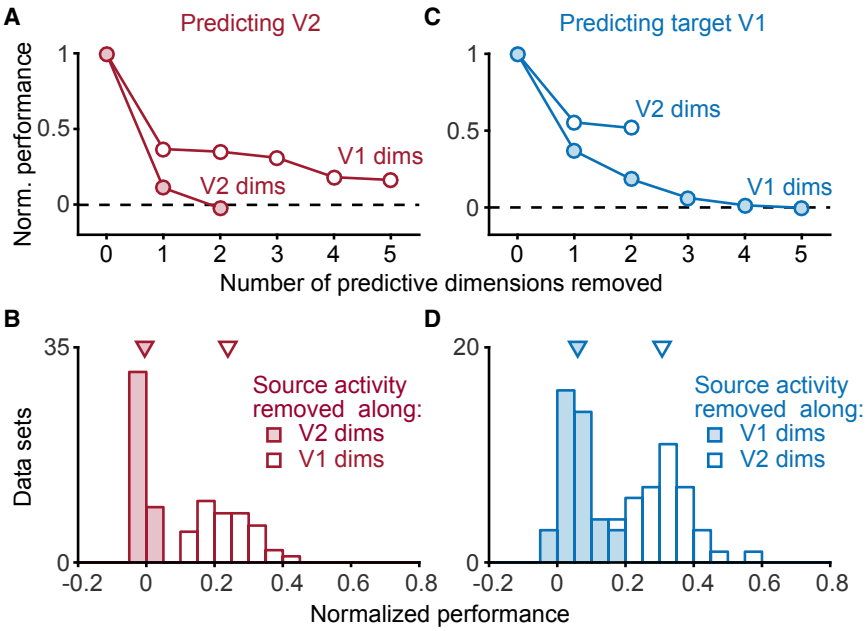

Figure 6. V2 Predictive Dimensions Are Not Aligned with Target V1 Predictive Dimensions

(A) Source V1 activity outside of the V1 predictive dimensions is still predictive of V2 activity. V2 predictive performance quickly decreased as we removed the source V1 activity along the V2 predictive dimensions (filled circles). Removing the source V1 activity along the target V1 predictive dimensions had a smaller impact on V2 predictive performance (open circles). Predictive performance was normalized by the performance of the reduced-rank regression model when no activity was removed. SEM is smaller than plotted circles.

(B) Across all datasets, removing all V2 predictive dimensions drove the V2 predictive performance to 0, as expected (red histogram). Removing the same number of V1 predictive dimensions, had a smaller impact on performance (white histogram), as we could still account for roughly one fifth of the predictable activity in V2.

(C) Source V1 activity outside of the V1-V2 communication subspace still accounts for a substantial part of the explained activity in target V1. Target V1 predictive performance decreased faster when removing source V1 activity along the target V1 predictive dimensions (filled circles), when compared to removing source activity along the V2 predictive dimensions (open circles). SEM is smaller than plotted circles.

(D) Across all datasets, even after removing all source activity that was predictive of V2, we could still account for approximately a third of the predictable activity in target V1 (white histogram). Removing the same number of target V1 predictive dimensions had a much larger effect on target V1 predictive performance (blue histogram).

coherence. To ensure that our estimates of V1-V2 interactions were not distorted by simple downstream non-linearities, we implemented several feedforward network models with standard non-linearities (e.g., squaring). In all cases, we found that our analyses recovered interaction dimensionality that closely matched the dimensionality of the linear weights (Figure S4).

Given this reasoning, how can we be sure that the communication subspaces are not an oddity, perhaps defining private and communicated V1 fluctuations differently for each grating stimulus? First, we confirmed that a communication subspace was evident when we analyzed our grating datasets together (Figure S5). Thus, it is not the case that all V1 population fluctuations that are private during the presentation of one grating stimulus are relayed to V2 during the presentation of another. Consistent with the existence of a shared communication subspace, we also found that the communication subspace defined for responses to one grating could effectively predict responses to other gratings (Figure S6). Second, we analyzed V1-V2 interactions during repeated presentations of brief naturalistic movies. These responses also revealed a communication subspace (Figure S7), indicating that the low-dimensional V1-V2 interactions do not arise from the use of grating stimuli. Finally, we analyzed the relationship between the communication subspace and the mapping of stimulus-driven activity from V1 to V2 (i.e., the PSTHs) and found that the communication subspace was able to capture responses that included stimulus information (Figure S8). Thus, the communication subspace identified using trial-to-trial fluctuations captures important aspects of the inter-areal circuitry that is used to relay stimulus information. These lines of evidence together indicate that the communication subspace is a fundamental aspect of V1-V2 interactions.

What is the basis of the communication subspace? One possibility might be that our results reflect global population fluctuations, which involve all neurons increasing and decreasing their activity together (Ecker et al., 2014; Schölvinck et al., 2015; Williamson et al., 2016) and may be more prevalent under anesthesia (Ecker et al., 2014; but see Rabinowitz et al., 2015). However, since global fluctuations are one dimensional, they cannot by themselves explain the V1-V2 interactions reported here, which typically involved more than a single dimension. In addition, the most predictive dimensions for the V1-V2 interaction were not well aligned with the largest shared fluctuations in V1, nor with the dimensions that were most predictive of the target V1 activity. Notably, we observed a similar communication subspace in simultaneous population recordings in V1 and V4 of awake monkeys (Figure S2), ruling out any confounding influence of anesthesia.

Another possibility might be that the communication subspace between V1 and V2 reflects feedback from higher cortical areas (e.g., feedback from V4 or MT to V2 and V1). In highly interconnected networks, such as the visual cortex, it is difficult to infer with certainty the source of inputs, especially with correlational methods such as those employed here. However, two pieces of evidence suggest that feedback cannot explain our findings. (1) Feedback connections have coarse retinotopic specificity, with individual axons spanning a relatively large portion of the visual field (Angelucci et al., 2002; Stettler et al., 2002). Our effects, however, are retinotopically specific. That is, when we analyzed additional recording sessions where the

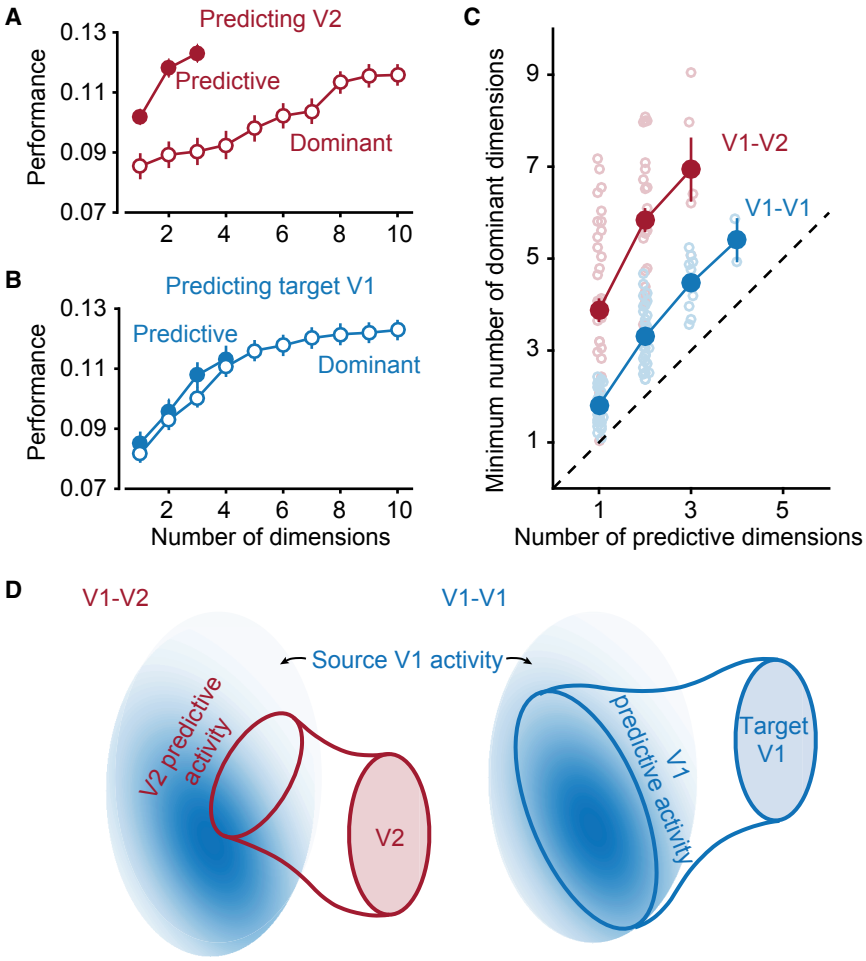

Figure 7. The Dominant Dimensions of V1 Are Not the Most Predictive of V2

(A) Predicting V2 activity using dominant and predictive dimensions. Dominant dimensions (open circles, factor-analysis regression) carried less predictive power than the same number of predictive dimensions (filled circles, reduced-rank regression). Error bars indicate SEM across cross-validation folds.

(B) Predicting target V1 activity using dominant and predictive dimensions. Predictive performance using dominant dimensions (open circles, factor-analysis regression) was similar to the predictive performance obtained for the same number of predictive dimensions (filled circles, reduced-rank regression). Error bars indicate SEM across cross-validation folds.

(C) For a given number of predictive dimensions, a larger number of dominant dimensions was required to reach (within a SEM, across folds) the same V2 predictive performance (red circles). When predicting target V1 activity, the number of dominant dimensions needed was only slightly greater than the number of predictive dimensions (blue circles). Error bars indicate SEM across datasets. Faded circles show results for each dataset and were horizontally jittered for visual clarity.

(D) Left: Schematic of V1-V2 results. Only a small number of activity patterns in the source V1 population was predictive of the V2 population. These predictive activity patterns did not correspond to the dominant patterns in the source V1 population. Large blue ellipse represents the set of all activity patterns observed in the source V1 population. Darker shading indicates more dominant activity patterns. Right: Schematic of V1-V1 results. A large number of activity patterns in the source V1 population was predictive of the target V1 population. These predictive activity patterns corresponded to the dominant patterns in the source V1 population.

V1 and V2 populations had receptive fields that were misaligned by several degrees, we found a much weaker V1-V2 interaction that was often well captured by a single predictive dimension (Figure S3). Furthermore, this dimension was well aligned with the largest shared fluctuations of the source V1 population. The mismatch between the retinotopic specificity of our results and that of feedback connections suggests that the communication subspace does not arise primarily from feedback. (2) Our V2 recordings were performed in the middle layers, which do not receive feedback and are driven almost exclusively by the superficial layers of V1 (Felleman and Van Essen, 1991). Consistent with this, we found in these data an elevated probability of V2 spiking several milliseconds after the occurrence of a spike in V1 (Zandvakili and Kohn, 2015). This functional signature also suggests a strong feedforward component to the V1-V2 interaction, though it does not exclude the possibility that feedback signals contribute as well.

An alternative possibility might be that the low-dimensional communication subspace arises because only a small subset of V1 neurons project to V2. There is good evidence for selective connections between V1 and V2 (Sincich et al., 2010) and between V1 and other areas (Glickfeld et al., 2013, but see Han et al., 2018 for a contrary view). However, additional simulations confirmed that our observations do not arise trivially from sparse anatomical projection between source and target areas (Figure S9). Further, we emphasize that while anatomy constrains how activity can be routed in cortex, the flexibility of cognition and perception requires additional mechanisms that allow inter-areal signaling to be adjusted from moment to moment, based on task demands (e.g., Saalmann et al., 2012; Salazar et al., 2012).

We propose instead that the communication subspace is an advantageous design principle of inter-area communication. The ability of a source area to communicate only certain activity patterns while keeping others "private" could be a means for the selective routing of signals between areas. To understand the computational benefit of structuring inter-areal communication in this way, we implemented a simulation that captures the common scenario of a source area projecting to two downstream areas, areas A and B (Figure 8). If each downstream area reads from the source area using a different communication subspace, there will be dimensions of the source population activity that are relayed to area A but not to area B (Figure 8A) and vice versa (Figure 8B). Crucially, if the interaction between these areas does not

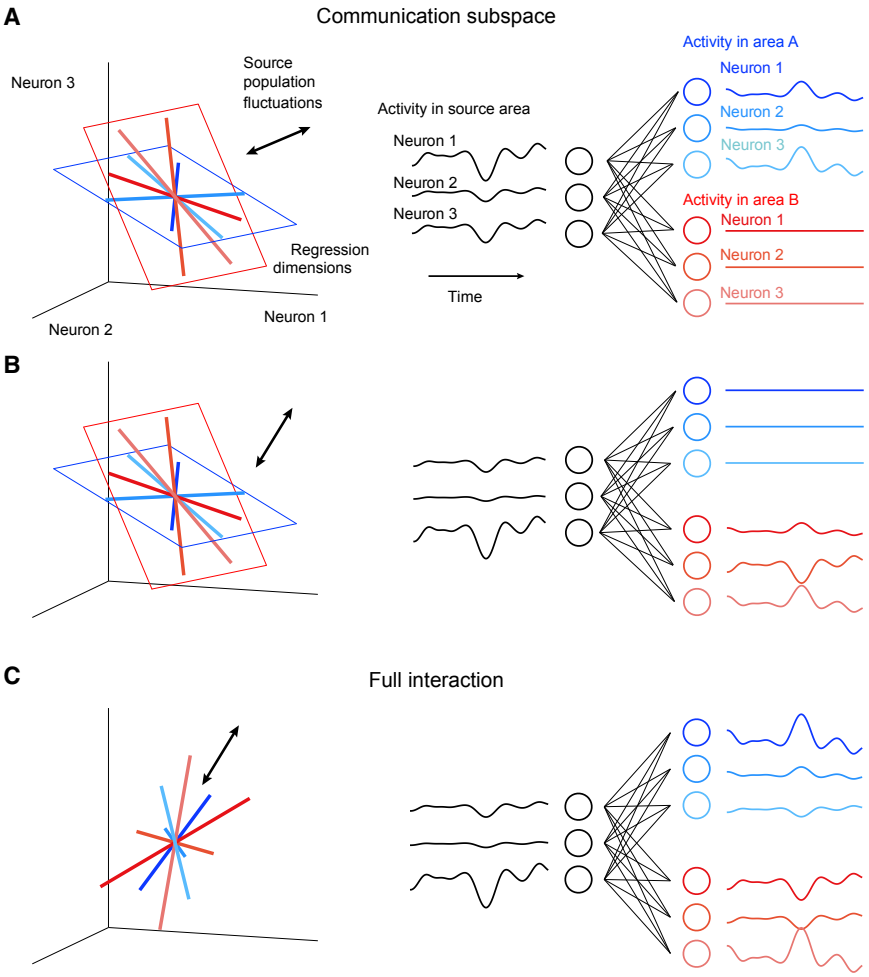

Figure 8. Communication Subspaces Enable Selective Communication with Multiple Downstream Areas

(A) Communication subspace: activity in source area influences downstream area A but not area B. Left: Each of the coordinate axes represents the activity of one source neuron. Each colored line represents the regression dimension of one neuron in area A (blue) or area B (red). Blue plane corresponds to the communication subspace for area A. Red plane corresponds to the communication subspace for area B. Source population fluctuations are indicated by the black arrow. Right: The activity of three source neurons (black, corresponding to the source population fluctuations indicated by the black arrow in the left panel) is mapped to the activity of three neurons in area A (blue) and three neurons in area B (red).

(B) Communication subspace: activity in source area influences downstream area B but not area A. Same conventions as in (A). The only difference from (A) is the direction of source population fluctuations (i.e., the population activity patterns produced in the source area).

(C) Full interaction: activity in source area influences both downstream areas. Same conventions as in (A).

involve communication subspaces, then all fluctuations in the source population will be relayed to both downstream areas (Figure 8C). The communication subspace is consequently a population-level mechanism whereby activity can be selectively routed between brain areas.

The selective routing allowed by the communication subspace could be adjusted dynamically, allowing moment-to-moment modulation of interactions between cortical areas. Dynamic routing could be accomplished by altering the structure of population activity in a source area; it need not involve changing the communication subspace itself. Much recent work has shown that the structure of population activity is highly and rapidly malleable, by stimulus drive (Kohn and Smith, 2005; Churchland et al., 2010; Mazzucato et al., 2016), task demands (Cohen and Newsome, 2008; Elsayed et al., 2016; Bondy et al., 2018), attention (Cohen and Maunsell, 2009; Mitchell et al., 2009), learning (Gu et al., 2011; Jeanne et al., 2013), and other factors (Kohn et al., 2016).

Allowing interactions between areas to be modulated by the alignment of population activity with a relevant communication subspace has several advantages over a well-known alternative: defining interaction strength by the phase-alignment of spikes to ongoing oscillations (Fries, 2015) (termed "communication through coherence," CTC). First, the communication subspace hypothesis does not require coordinated oscillations between the source and target areas, which can be difficult to achieve in practice (Ray and Maunsell, 2015). Instead, the implementation of a communication subspace requires only that the target area takes a particular type of weighted combination of its inputs, namely a linear readout that is low dimensional. This can be implemented in a linear feedforward network (Salinas and Abbott, 1994; Jazayeri and Movshon, 2006), if the weights for each downstream neuron are defined as linear combinations of the same set of basis "weights" (or predictive dimensions). Second, different target areas (or subpopulations within the same target area) can have different communication subspaces in the same source area (Figure 8). CTC can also route distinct signals to downstream targets by using different oscillations within the source area, each of which is coherent with a different downstream target. However, the number of oscillations that can be distinguished by phase is limited by the temporal precision of neurons, and it is not clear whether the same source neurons can entrain to different oscillations at the same time (Remme et al., 2010).

Our framework for understanding inter-areal population interactions makes clear predictions of how the communication subspace could contribute to behavior, which can be tested in future work. For instance, if attention involves altered inter-areal communication, this could be achieved by better alignment between population responses in a source area and the communication subspace relaying those responses to a relevant

downstream area. Similarly, learning could involve achieving population activity patterns that are better aligned with an existing communication subspace (Sadtler et al., 2014), or perhaps altering the communication subspace itself. Finally, the degree to which the effects of perturbation experiments (e.g., patterned optogenetic stimulation) would propagate across areas could depend on their alignment with the relevant communication subspaces. A critical implication of our work is thus that studying how experimental manipulations alter population responses in a given cortical area can be misleading. One must also understand how these altered population responses align with the mapping to downstream areas.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Visual stimulation and recordings
  - Data preprocessing
  - Regression models
  - Factor analysis
  - Removing activity along the predictive dimensions
  - Comparing dominant and predictive dimensions
  - Selective communication simulation
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes nine figures and can be found with this article online at https://doi.org/10.1016/j.neuron.2019.01.026.

ACKNOWLEDGMENTS

We thank A. Aschner and S. Tanabe for graciously providing the V1-V4 awake recordings in Figure S2, A. Jasper for help in analyzing the V1-V4 awake recordings, and A. Motiwala, B. Cowley, and A. Kepecs for invaluable discussions. We also thank M. Cohen, S. Kuhlman, and R. Coen-Cagli for providing feedback on the manuscript. This work was supported by the Fundação para a Ciência e a Tecnologia graduate scholarship SFRH/BD/52069/2012 (J.D.S.), John and Claire Bertucci Graduate Fellowship (J.D.S.), NIH U01 NS094288 (C.K.M.), Simons Collaboration on the Global Brain 364994 (B.M.Y., A.K.), 543009 (C.K.M.), 543065 (B.M.Y.), and 542999 (A.K.), NIH CRCNS R01 HD071686 (B.M.Y.), NIH CRCNS R01 NS105318 (B.M.Y.), NSF NCS BCS 1533672 and 1734916 (B.M.Y.), Pennsylvania Department of Health Research Formula Grant SAP 4100077048 under the Commonwealth Universal Research Enhancement program (B.M.Y.), NIH R01 EY016774 (A.K.), and Irma T. Hirschl Trust (A.K.).

AUTHOR CONTRIBUTIONS

J.D.S., C.K.M., B.M.Y., and A.K. designed the analyses. J.D.S. performed all the analyses. A.Z. and A.K. designed and performed the experiments. J.D.S., C.K.M., B.M.Y., and A.K. wrote the manuscript. C.K.M., B.M.Y., and A.K. contributed equally to this work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 12, 2018
Revised: October 12, 2018
Accepted: January 14, 2019
Published: February 12, 2019

REFERENCES

Angelucci, A., Levitt, J.B., Walton, E.J., Hupe, J.-M., Bullier, J., and Lund, J.S. (2002). Circuits for local and global signal integration in primary visual cortex. J. Neurosci. *22*, 8633–8646.

Arce-McShane, F.I., Ross, C.F., Takahashi, K., Sessle, B.J., and Hatsopoulos, N.G. (2016). Primary motor and sensory cortical areas communicate via spatiotemporally coordinated networks at multiple frequencies. Proc. Natl. Acad. Sci. USA *113*, 5083–5088.

Bondy, A.G., Haefner, R.M., and Cumming, B.G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. Nat. Neurosci. *21*, 598–606.

Bosman, C.A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., and Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron *75*, 875–888.

Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat. Neurosci. *13*, 369–378.

Cohen, M.R., and Maunsell, J.H.R. (2009). Attention improves performance primarily by reducing interneuronal correlations. Nat. Neurosci. *12*, 1594–1600.

Cohen, M.R., and Newsome, W.T. (2008). Context-dependent changes in functional circuitry in visual area MT. Neuron *60*, 162–173.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.

Ecker, A.S., Berens, P., Cotton, R.J., Subramaniyan, M., Denfield, G.H., Cadwell, C.R., Smirnakis, S.M., Bethge, M., and Tolias, A.S. (2014). State dependence of noise correlations in macaque primary visual cortex. Neuron *82*, 235–248.

Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., and Cunningham, J.P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nat. Commun. *7*, 13239.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., and Movshon, J.A. (2013). A functional and perceptual signature of the second visual area in primates. Nat. Neurosci. *16*, 974–981.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). The Elements of Statistical Learning, volume 1. Springer series in statistics (Berlin: Springer).

Fries, P. (2015). Rhythms for cognition: communication through coherence. Neuron *88*, 220–235.

Fries, P., Reynolds, J.H., Rorie, A.E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science *291*, 1560–1563.

Glickfeld, L.L., Andermann, M.L., Bonin, V., and Reid, R.C. (2013). Cortico-cortical projections in mouse visual cortex are functionally target specific. Nat. Neurosci. *16*, 219–226.

Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science *324*, 1207–1210.

Gu, Y., Liu, S., Fetsch, C.R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Perceptual learning reduces interneuronal correlations in macaque visual cortex. Neuron *71*, 750–761.

Han, Y., Kebschull, J.M., Campbell, R.A.A., Cowan, D., Imhof, F., Zador, A.M., and Mrsic-Flogel, T.D. (2018). The logic of single-cell projections from visual cortex. Nature *556*, 51–56.

Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature *484*, 62–68.

Izenman, A.J. (1975). Reduced-rank regression for the multivariate linear model. J. Multivariate Anal. *5*, 248–264.

Jazayeri, M., and Movshon, J.A. (2006). Optimal representation of sensory information by neural populations. Nat. Neurosci. *9*, 690–696.

Jeanne, J.M., Sharpee, T.O., and Gentner, T.Q. (2013). Associative learning enhances population coding by inverting interneuronal correlation patterns. Neuron *78*, 352–363.

Jia, X., Smith, M.A., and Kohn, A. (2011). Stimulus selectivity and spatial coherence of gamma components of the local field potential. J. Neurosci. *31*, 9390–9403.

Jia, X., Tanabe, S., and Kohn, A. (2013). γ and the coordination of spiking activity in early visual cortex. Neuron *77*, 762–774.

Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2014). Cortical activity in the null space: permitting preparation without movement. Nat. Neurosci. *17*, 440–448.

Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2015). Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. eLife *4*, e04677.

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.-L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. eLife *5*, e10989.

Kohn, A., and Smith, M.A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. J. Neurosci. *25*, 3661–3673.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and neuronal population information. Annu. Rev. Neurosci. *39*, 237–256.

Lin, I.-C., Okun, M., Carandini, M., and Harris, K.D. (2015). The nature of shared cortical variability. Neuron *87*, 644–656.

Mazzucato, L., Fontanini, A., and La Camera, G. (2016). Stimuli reduce the dimensionality of cortical activity. Front. Syst. Neurosci. *10*, 11.

Menzer, D.L., Rao, N.G., Bondy, A., Truccolo, W., and Donoghue, J.P. (2014). Population interactions between parietal and primary motor cortices during reach. J. Neurophysiol. *112*, 2959–2984.

Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron *63*, 879–888.

Nowak, L.G., Munk, M.H.J., James, A.C., Girard, P., and Bullier, J. (1999). Cross-correlation study of the temporal interactions between areas V1 and V2 of the macaque monkey. J. Neurophysiol. *81*, 1057–1074.

Oemisch, M., Westendorff, S., Everling, S., and Womelsdorf, T. (2015). Interareal spike-train correlations of anterior cingulate and dorsal prefrontal cortex during attention shifts. J. Neurosci. *35*, 13076–13089.

Pesaran, B., Nelson, M.J., and Andersen, R.A. (2008). Free choice activates a decision circuit between frontal and parietal cortex. Nature *453*, 406–409.

Pooresmaeili, A., Poort, J., and Roelfsema, P.R. (2014). Simultaneous selection by object-based attention in visual and frontal cortex. Proc. Natl. Acad. Sci. USA *111*, 6467–6472.

Rabinowitz, N.C., Goris, R.L., Cohen, M., and Simoncelli, E.P. (2015). Attention stabilizes the shared gain of V4 populations. eLife *4*, e08998.

Ray, S., and Maunsell, J.H.R. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. *9*, e1000610.

Ray, S., and Maunsell, J.H.R. (2015). Do gamma oscillations play a role in cerebral cortex? Trends Cogn. Sci. *19*, 78–85.

Remme, M.W.H., Lengyel, M., and Gutkin, B.S. (2010). Democracy-independence trade-off in oscillating dendrites and its implications for grid cells. Neuron *66*, 429–437.

Roberts, M.J., Lowet, E., Brunet, N.M., Ter Wal, M., Tiesinga, P., Fries, P., and De Weerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. Neuron *78*, 523–536.

Roe, A.W., and Ts'o, D.Y. (1999). Specificity of color connectivity between primate V1 and V2. J. Neurophysiol. *82*, 2719–2730.

Ruff, D.A., and Cohen, M.R. (2016). Attention increases spike count correlations between visual cortical areas. J. Neurosci. *36*, 7523–7534.

Saalmann, Y.B., Pinsk, M.A., Wang, L., Li, X., and Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. Science *337*, 753–756.

Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., and Batista, A.P. (2014). Neural constraints on learning. Nature *512*, 423–426.

Salazar, R.F., Dotson, N.M., Bressler, S.L., and Gray, C.M. (2012). Content-specific fronto-parietal synchronization during visual working memory. Science *338*, 1097–1100.

Salinas, E., and Abbott, L.F. (1994). Vector reconstruction from firing rates. J. Comput. Neurosci. *1*, 89–107.

Santhanam, G., Yu, B.M., Gilja, V., Ryu, S.I., Afshar, A., Sahani, M., and Shenoy, K.V. (2009). Factor-analysis methods for higher-performance neural prostheses. J. Neurophysiol. *102*, 1315–1330.

Schölvinck, M.L., Saleem, A.B., Benucci, A., Harris, K.D., and Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. J. Neurosci. *35*, 170–178.

Semedo, J., Zandvakili, A., Kohn, A., Machens, C.K., and Yu, B.M. (2014). Extracting latent structure from multiple interacting neural populations. In Advances in Neural Information Processing Systems, *27*, pp. 2942–2950.

Sincich, L.C., Jocson, C.M., and Horton, J.C. (2010). V1 interpatch projections to v2 thick stripes and pale stripes. J. Neurosci. *30*, 6963–6974.

Smith, M.A., and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. J. Neurosci. *28*, 12591–12603.

Smith, M.A., Jia, X., Zandvakili, A., and Kohn, A. (2013). Laminar dependence of neuronal correlations in visual cortex. J. Neurophysiol. *109*, 940–947.

Stettler, D.D., Das, A., Bennett, J., and Gilbert, C.D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron *36*, 739–750.

Truccolo, W., Hochberg, L.R., and Donoghue, J.P. (2010). Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. Nat. Neurosci. *13*, 105–111.

Williamson, R.C., Cowley, B.R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M.A., and Yu, B.M. (2016). Scaling properties of dimensionality reduction for neural populations and network models. PLoS Comput. Biol. *12*, e1005141.

Wong, Y.T., Fabiszak, M.M., Novikov, Y., Daw, N.D., and Pesaran, B. (2016). Coherent neuronal ensembles are rapidly recruited when making a look-reach decision. Nat. Neurosci. *19*, 327–334.

Yu, B.M., Cunningham, J.P., Santhanam, G., Ryu, S.I., Shenoy, K.V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. J. Neurophysiol. *102*, 614–635.

Yu, Y., Schmid, A.M., and Victor, J.D. (2015). Visual processing of informative multipoint correlations arises primarily in V2. eLife *4*, e06604.

Zandvakili, A., and Kohn, A. (2015). Coordinated neuronal activity enhances corticocortical communication. Neuron *87*, 827–839.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Zandvakili and Kohn (2019). Paired V1-V2 neuronal spiking responses in anesthetized macaque monkey. CRCNS.org.	Albert Einstein College of Medicine	https://doi.org/10.6080/K0B27SHN
Experimental Models: Organisms/Strains		
Macaca fascicularis, Male, 2-3 years old	Charles River	N/A
Software and Algorithms		
MATLAB	Mathworks	https://www.mathworks.com/products/matlab.html
Analysis code	João Semedo	https://github.com/joao-semedo/communication-subspace

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, João D. Semedo (jsemedo@cmu.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal procedures and recording details have been described in previous work (Smith and Kohn, 2008; Zandvakili and Kohn, 2015). Briefly, animals (macaca fascicularis, male, 2-3 years old) were anesthetized with ketamine (10 mg/kg) and maintained on isoflurane (1%–2%) during surgery. Recordings were performed under sufentanil (typically 6-18 μg/kg/hr) anesthesia. Vecuronium bromide (150 μg/kg/hr) was used to prevent eye movements. All procedures were approved by the IACUC of the Albert Einstein College of Medicine.

METHOD DETAILS

Visual stimulation and recordings

The data analyzed here represent a subset of those reported in Zandvakili and Kohn (2015), namely those that involved the largest and best retinotopically-aligned populations. V1 activity was recorded using a 96 channel Utah array (400 micron inter-electrode spacing, 1 mm length, inserted to a nominal depth of 600 microns; Blackrock, UT). We recorded V2 activity using a set of electrodes/tetrodes (interelectrode spacing 300 microns) whose depth could be controlled independently (Thomas Recording, Germany). These electrodes were lowered through V1, the underlying white matter, and then into V2. Within V2, we targeted neurons in the input layers. We verified the recordings were performed in the input layers using measurements of the depth in V2 cortex, histological confirmation (in a subset of recordings), and correlation measurements. For complete details see Smith et al. (2013) and Zandvakili and Kohn (2015). Voltage snippets that exceeded a user-defined threshold were digitized and sorted offline. The sampled neurons had spatial receptive fields within $2-4^{\circ}$ of the fovea, in the lower visual field. Average receptive field size (defined as $\pm$ 2 S.D.s of a Gaussian function fit to the data) was $1.22 \pm 0.01^{\circ}$ for V1 and $2.33 \pm 0.09^{\circ}$ for V2.

We measured responses evoked by drifting sinusoidal gratings (1 cyc/$^{\circ}$; drift rate of $3-6.25$ Hz; $2.6-4.9^{\circ}$ in diameter; full contrast, defined as Michelson contrast (Lmax-Lmin/Lmax+Lmin) where Lmin is 0 cd/m^2 and Lmax is 80 cd/m^2) at 8 different orientations (22.5° steps), on a calibrated CRT monitor placed 110 cm from the animal (1024 × 768 pixel resolution at 100 Hz refresh). Each stimulus was presented 300-400 times for 1.28 s. Each presentation was preceded by an interstimulus interval of 1.5 s. The duration of each experiment varied from 5 to 7 days.

We recorded neuronal activity in three animals. In two of the animals, we recorded in two different but nearby locations in V2, providing distinct middle-layer populations. We refer to each of these five recordings as a session. We treated responses to each of the 8 stimuli in each session separately, yielding a total of 40 datasets.

Data preprocessing

We counted spikes in 100 ms bins, beginning 160 ms after stimulus onset and spanning a total of 1 s (10 bins per trial). To study how neuronal activity in the two areas is related, we reasoned that any fluctuations in the V1 responses, whether due to changes in the visual stimulus or not, could relate to fluctuations in V2. We therefore subtracted the appropriate peri-stimulus time histogram (PSTH) from each single-trial response, and then analyzed the residuals for each orientation (termed datasets) separately. We confirmed that the temporal structure had little effect on our results by shuffling the data across trials while maintaining the temporal identity; doing so reduced the predictive performance for the V2 population to 0. We found qualitatively similar results for a wide range of bin widths (20 ms - 1 s). Furthermore, we obtained similar results after z-scoring both the source and target population responses, ruling out the possibility that our results were driven by a few high-firing neurons. For all analyses, we excluded neurons that fired less than 0.5 spikes/s on average, across all trials.

We compared our analyses of V1-V2 interactions to the results of applying the same analyses to a held-out V1 population (V1-V1). The target population in the V1-V1 analyses was a held-out subset of the originally recorded population, which was matched in neuron count to the corresponding V2 population. We also matched the firing rate distribution (mean-matched) to the V2 population separately for each stimulus condition (as in Churchland et al. (2010)). To do so, we binned the firing rate distribution of the V1 and V2 populations (for each neuron, the average firing rate was taken across time and trials for each dataset), and determined the common firing rate distribution (i.e., for each firing rate interval, we took the minimum neuron count between the two populations). For each firing rate interval, we then randomly picked this minimum number of neurons from the corresponding bin in each population, without replacement. Because we had many more V1 than V2 neurons, the common distribution usually matched the V2 distribution and we selected an equal number of V1 neurons. The size of the matched populations ranged from 15 to 31 units across datasets (mean: 22.3). The V1 neurons that were not selected for the held-out population defined the source V1 population. V2 neurons that were not selected for the V2 mean-matched population were not used in the analysis. We repeated the mean-matching procedure 25 times, using different random, mean-matched subsets of neurons (and consequently producing a different source population). Results for each dataset are based on averages across these repeats. The pairwise correlation (r_{sc}) analysis in Figure 2A was based on a single mean-matching procedure which was done jointly for all stimulus conditions. Statistical evaluation for this analysis was performed after converting r_{sc} to z-scores using the Fisher transformation (Kohn and Smith, 2005):

$$z = \frac{1}{2}\ln\left(\frac{1+r_{sc}}{1-r_{sc}}\right)$$

Regression models

We first related trial-to-trial fluctuations in the source V1 population to those in the target populations using a linear model of the form:

$$Y = XB$$

where X is a $n \times p$ matrix containing the residual activity of the source V1 population and Y is a $n \times q$ matrix containing the residual activity of the target (V1 or V2) population (n represents the number of data points, p and q are the number of neurons in the source and target populations, respectively). The coefficient matrix B is of size $p \times q$. Each of the q columns of B linearly combines the activity of the p neurons in X to predict the activity of one neuron in Y. B can be found using the ordinary least-squares (OLS) solution which minimizes the squared prediction error:

$$B_{OLS} = \left(X^T X\right)^{-1} X^T Y$$

To reduce overfitting, we used ridge regression (referred to as *full regression model* in the main text), a variant of classical linear regression, which gives the solution $B_{Ridge} = (X^T X + \lambda I)^{-1} X^T Y$, where I is a $p \times p$ identity matrix and λ is a constant that determines the strength of regularization. We chose the value of λ using 10-fold cross-validation. Specifically, we selected the largest λ for which mean performance (across folds) was within one SEM of the best performance, separately for each dataset (i.e., for each stimulus condition in each recording session). To quantify model performance, we employed 10-fold nested cross-validation (Friedman et al., 2001)

We sought to test whether the target population activity (V1 or V2) could be predicted using a subspace of the source V1 population activity. In other words, we asked if the linear model $Y = XB$ was still accurate when we impose B to be of a given rank, $\text{rank}(B) = m$. This constrained linear regression problem is known as reduced-rank regression (RRR) (Izenman, 1975; Kobak et al., 2016), and can be solved using the singular value decomposition:

$$B_{RRR} = B_{OLS} V V^T$$

where B_{OLS} is the ordinary least-squares solution and the columns of the $q \times m$ matrix V contain the top m principal components of the optimal linear predictor $\hat{Y}_{OLS} = XB_{OLS}$. To predict target population activity using RRR, we computed:

$$\hat{Y}_{RRR} = XB_{RRR} = XB_{OLS} V V^T = X\bar{B}V^T$$

where $\overline{B} = B_{OLS}V$ is a matrix of size $p \times m$. The columns of $\overline{B}$ define which dimensions of the source population activity are used when generating predictions: they are the predictive dimensions. The sets of weights used to predict each target neuron (the columns of B_{RRR}) are themselves linear combinations of the columns of $\overline{B}$. Note also that the columns of $\overline{B}$ do not form an orthonormal basis. Rather, they are uncorrelated with respect to the source activity, i.e., $\overline{B}^T \sum \overline{B} = D$, where Σ is the covariance matrix of the source population activity and D is a diagonal matrix. Thus, the columns of $\overline{B}$ are linearly independent and $\text{rank}(\overline{B}) = m$.

To find the optimal dimensionality for the RRR model (the value of m), we used 10-fold cross-validation and found the smallest number of dimensions for which predictive performance was within one SEM of the peak performance.

Factor analysis

To quantify the dimensionality of the activity in the target populations we used Factor Analysis (FA) (Yu et al., 2009; Williamson et al., 2016). FA is defined by:

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$$
$$\mathbf{y} \mid \mathbf{z} \sim \mathcal{N}(L\mathbf{z} + \boldsymbol{\mu}, \Psi)$$

where $\mathbf{y}$ is a q-dimensional vector containing the observed residuals at a given time point, L is the $q \times m$ loading matrix that defines the relationship between the m-dimensional ($m < q$) latent variable $\mathbf{z}$ and $\mathbf{y}$, $\boldsymbol{\mu}$ is a q-dimensional vector and Ψ is a $q \times q$ diagonal matrix. We estimated the dimensionality of the latent variable $\mathbf{z}$ in two steps: (1) we found the number of dimensions m_{peak} that maximized the cross-validated log-likelihood of the observed residuals; (2) we fitted a FA model with m_{peak} dimensions and chose m, using the eigenvalue decomposition, as the smallest dimensionality that captured 95% of the variance in the shared covariance matrix LL^T. This procedure provides more robust estimates of the FA model dimensionality (Williamson et al., 2016).

Removing activity along the predictive dimensions

In order to remove the source population activity along the predictive dimensions, we projected the source activity onto the subspace that is uncorrelated with the predictive dimensions. Formally, we state that two dimensions defined by the vectors $\mathbf{u}$ and $\mathbf{v}$ are uncorrelated with respect to the source activity matrix X if:

$$\mathbf{u}^T \boldsymbol{\Sigma} \mathbf{v} = 0$$

where Σ is the covariance matrix of the source activity. Let matrix $\overline{B}$ contain the predictive dimensions. The set of vectors in the uncorrelated subspace is:

$$\left\{ \mathbf{v} : \overline{B}^T \Sigma \mathbf{v} = 0 \right\}$$

In particular, it will be useful to find an orthonormal basis for this subspace:

$$\left\{ Q : \overline{B}^T \Sigma Q = 0,\ Q^T Q = I \right\}$$

This can be accomplished using the singular value decomposition (SVD). Start by defining $M = \overline{B}^T \Sigma$ and consider its SVD $M = UDV^T$. Choosing Q as the last $p - m$ columns of V (corresponding to the 0 singular values) yields $MQ = 0$, $Q^T Q = I$, which makes Q an orthonormal basis for the uncorrelated subspace. We then projected the source population onto the uncorrelated subspace, $\widehat{X} = XQ$, and predicted target activity using ridge regression between $\widehat{X}$ and Y.

We also tested the effect of removing all population activity that was predictive of the target population activity under any stimulus condition by analyzing responses to all stimulus conditions together (Figure S5).

Comparing dominant and predictive dimensions

To identify the dominant dimensions in the source population, we fit a FA model, and determined the optimal dimensionality (as described above). Using this FA model, we estimated the latent variables $\widehat{\mathbf{z}} = \mathbb{E}[\mathbf{z} \mid \mathbf{y}]$ for each $\mathbf{z}$, then performed an orthonormalization procedure to order the elements of $\widehat{\mathbf{z}}$ by the amount of shared variance explained (Williamson et al., 2016). This allowed us to predict the target population activity using only the most dominant V1 dimension (first element of orthonormalized $\widehat{\mathbf{z}}$), the top two most dominant V1 dimensions (first two elements of orthonormalized $\widehat{\mathbf{z}}$), etc. We then compared the performance of the dominant and predictive dimensions for predicting activity of the target populations.

Selective communication simulation

In order to show how a communication subspace can subserve selective communication (Figure 8), we simulated responses in a source population (3 neurons), as well as in two downstream populations (3 neurons each). The responses of each downstream neuron were generated as a linear combination of the activity of the neurons in the source population. In Figure 8A-B, where both downstream areas interact with the source area via communication subspaces, the predictive dimensions for all neurons in each area were chosen to lie within a 2-dimensional subspace. Specifically, we generated these dimensions by creating randomly oriented unit vectors in the xy plane and then rotating these vectors 20° around the x axis for the downstream area A neurons and 40° for the downstream area B neurons. Predictive dimensions for both downstream areas were then rotated 20° around the z axis. In Figure 8C,

all predictive dimensions were generated by creating randomly oriented unit vectors in the 3-dimensional source activity space. To generate source activity we drew a sample from a Gaussian process with a squared exponential kernel (length scale $\ell^2 = 0.05$). This sample was then embedded into the 3-dimensional source activity space by projecting the activity along a chosen dimension, i.e., if the Gaussian process sample is represented as a $T \times 1$ vector $\mathbf{x}$ (where T represents the number of time points), and the chosen dimension is represented by the 3×1 vector $\mathbf{v}$, then the 3-dimensional source activity is given by $\mathbf{x}\mathbf{v}^T$. When a communication subspace was present, the source activity was chosen to lie along the private dimension of the relevant area. In Figure 8C, any choice of $\mathbf{v}$ leads to qualitatively similar results, so we chose it to align with the dimension used in Figure 8B. The response for each downstream neuron is given by the projection of the source activity onto the corresponding predictive dimension.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details can be found in the Results and figure legends. All statistical tests reported in the main text treat the datasets as independent (with the exception of Figure 2A, in which data are pooled across all stimuli, resulting in a single pairwise correlation value per pair per session). Repeating the same statistical tests across the five sessions (i.e., averaging the results across the 8 stimuli for each session) also returned significant results ($p < 0.05$) for all tests, with the exception of Figure 2B, where we can no longer reject the null hypothesis that the average predictive performance is the same when predicting target V1 and V2.

DATA AND SOFTWARE AVAILABILITY

The MATLAB analysis code with sample data is available at https://github.com/joao-semedo/communication-subspace. V1-V2 data are available at the CRCNS data sharing web site, at https://doi.org/10.6080/K0B27SHN.

Neuron, Volume *102*

Supplemental Information

Cortical Areas Interact through a Communication Subspace

João D. Semedo, Amin Zandvakili, Christian K. Machens, Byron M. Yu, and Adam Kohn

Supplementary Information | Cortical areas interact through a communication subspace

João D. Semedo, Amin Zandvakili,
Christian K. Machens, Byron M. Yu, Adam Kohn

Supplementary Figures

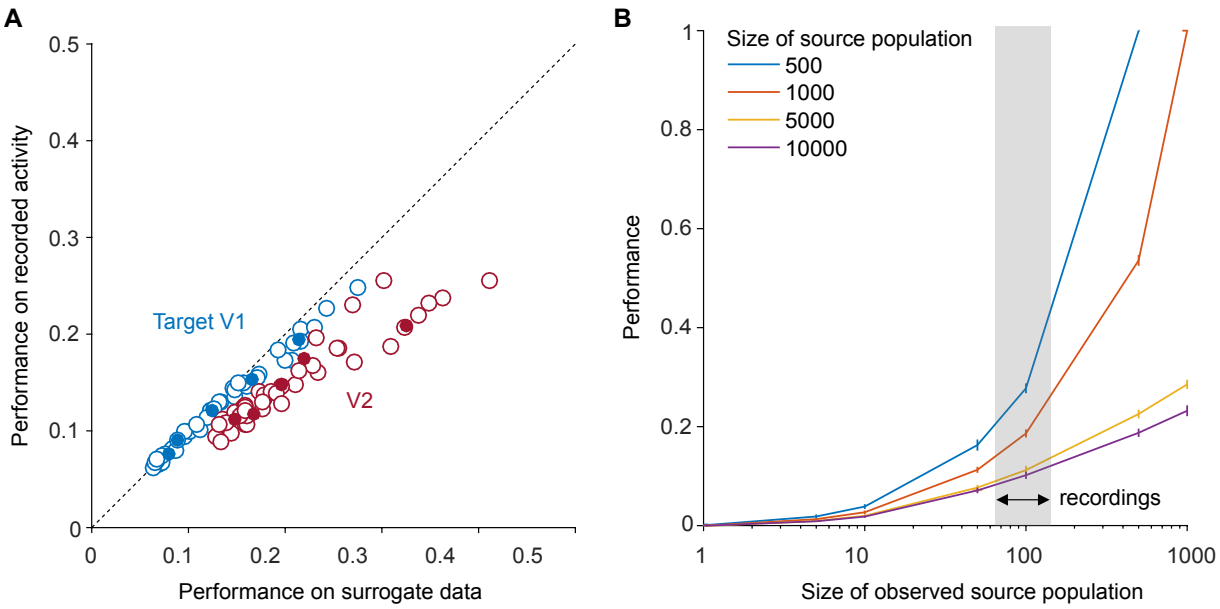

Figure S1, related to Figure 2. Predictive performance can be explained by Poisson variability in the target population or subsampling from a large source population.

Although we focus on relative comparisons of predictive performance and dimensionality, we wondered whether the absolute performance of our regression models was reasonable. In particular, the predictive performance might be limited because of variability in the target population that is not related to the recorded source population. This could occur because of stochasticity in spike generation in the target population (Faisal et al., 2008; Goris et al., 2014), or because the recorded source neurons are only a fraction of the relevant input neurons. In these cases, only part of the variability in the target population is predictable from the source population, regardless of the model used. Here we perform simulations to quantify the effect of stochasticity in spike generation in the target population, as well as of subsampling from a large source population. Our simulations indicate that the absolute predictive performance reported in Figure 2B is consistent with either spike generation stochasticity in the target population or subsampling of the input

population. The predictive performance for the recorded activity was higher than expected if both of these effects contribute.

(A) To assess the contribution of stochasticity, or noise, in the target population's spike generation process, we constructed surrogate target population responses (target V1 and V2) using linear combinations of the recorded source V1 activity and Poisson noise. We first applied reduced-rank regression (RRR) to the recorded activity to identify a (number of source neurons by number of target neurons) matrix B, which relates the activity between the source and target populations. In these simulations, B represents the ground truth for generating the surrogate target population responses. While other choices of B are possible, we used RRR to ensure the surrogate "residuals" both qualitatively match the recorded residuals and exhibit low-dimensional structure. The surrogate target residuals were obtained using $Y = XB$, where X is the recorded source residuals (a number of data points by number of source neurons matrix). We added the PSTH of the appropriate neuron and stimulus orientation to obtain a firing rate for each target neuron for each time bin and trial. We then drew spike counts from a Poisson distribution with the specified underlying rate. Note that we assumed Poisson spiking statistics and did not determine the level of noise from the data (Tolhurst et al., 1983). Using this approach, we generated surrogate target activity for both the V2 and target V1 populations for each of the recorded data sets, and then applied ridge regression to the residuals obtained from these surrogate data (to mimic the way in which we measured overall predictive performance in the main text). In these simulations, the predictive performance can only be limited by the Poisson variability, as the interaction between populations is linear by design.

We found that the predictive performance of the linear regression model on the surrogate data was similar to the performance on the recorded activity (ratio of recorded activity performance to surrogate data performance: 0.7 ± 0.01 for V2 and 0.95 ± 0.01 for target V1; open circles show average predictive performance for each data set; filled circles show average for each recording session). This outcome suggests that Poisson-like variability in the target population can largely account for the performance levels observed. That is, under the assumption that Poisson variability is "noise" and therefore not predictable, the linear model accounts for most of the predictable trial-to-trial variability in the target population.

(B) To mimic our recording from a fraction of the full neuronal source population, we conducted an additional simulation in which we subsampled from a large source population. Following the procedure of Smith and Kohn (2008), we generated the source activity (500 to 10,000 neurons) by drawing 4000 samples (matching the experimental data set size: 400 trials with 10 time bins each) from a multi-dimensional Gaussian distribution. The mean of the distribution was drawn from a uniform distribution between 0 and 100 for each neuron independently. To determine the covariance of the distribution, we first constructed a noise correlation matrix by assigning each neuron a "preferred orientation", which was drawn from a uniform distribution between $0°$ and $180°$. We then determined the correlation between each pair of neurons using the difference between preferred

orientations. Correlations varied from 0 to 0.3; these values were chosen so that the eigenspectrum of the resulting covariance matrix qualitatively matched that of the real data when matching the number of neurons. We then obtained the covariance matrix from the correlation matrix using the mean rates and assuming a Fano factor of 1. To generate activity in the target population (30 neurons), we defined the response of each target neuron using a weighted sum of the activity of the source population, with the weights chosen randomly from a standard Gaussian distribution. There is no added noise in this simulation. We then subsampled different proportions of the source population and used ridge regression to predict activity in the target population, as in the main text. We repeated the process of subsampling from the large source population and drawing a set of weights 25 times. All presented results are the average across the 25 repeats, with the shading indicating the standard deviation.

As expected, predictive performance depended strongly on the size of the observed source population. When the source population is completely observed, predictive performance is perfect, since there is no noise added to the target population in this simulation. As the size of the observed population decreases, so does predictive performance, reaching a value close to 0 when a single source neuron is observed. Notably, when the size of the observed population roughly matched the size of the source V1 populations (64 to 135 neurons, indicated by the shaded area), the predictive performance was similar to that observed in the recorded data (average predictive performance across all recordings: 0.14; predictive performance on the surrogate data: 0.28 for 100 observed neurons out of a source population of 500, 0.19 for 100 out of 1000, 0.11 for 100 out of 5000, 0.10 for 100 out of 10000). These results show that recording from a subset of the V1 neurons projecting to V2 limits predictive performance, leaving a large proportion of V2 variability unexplainable by the recorded activity.

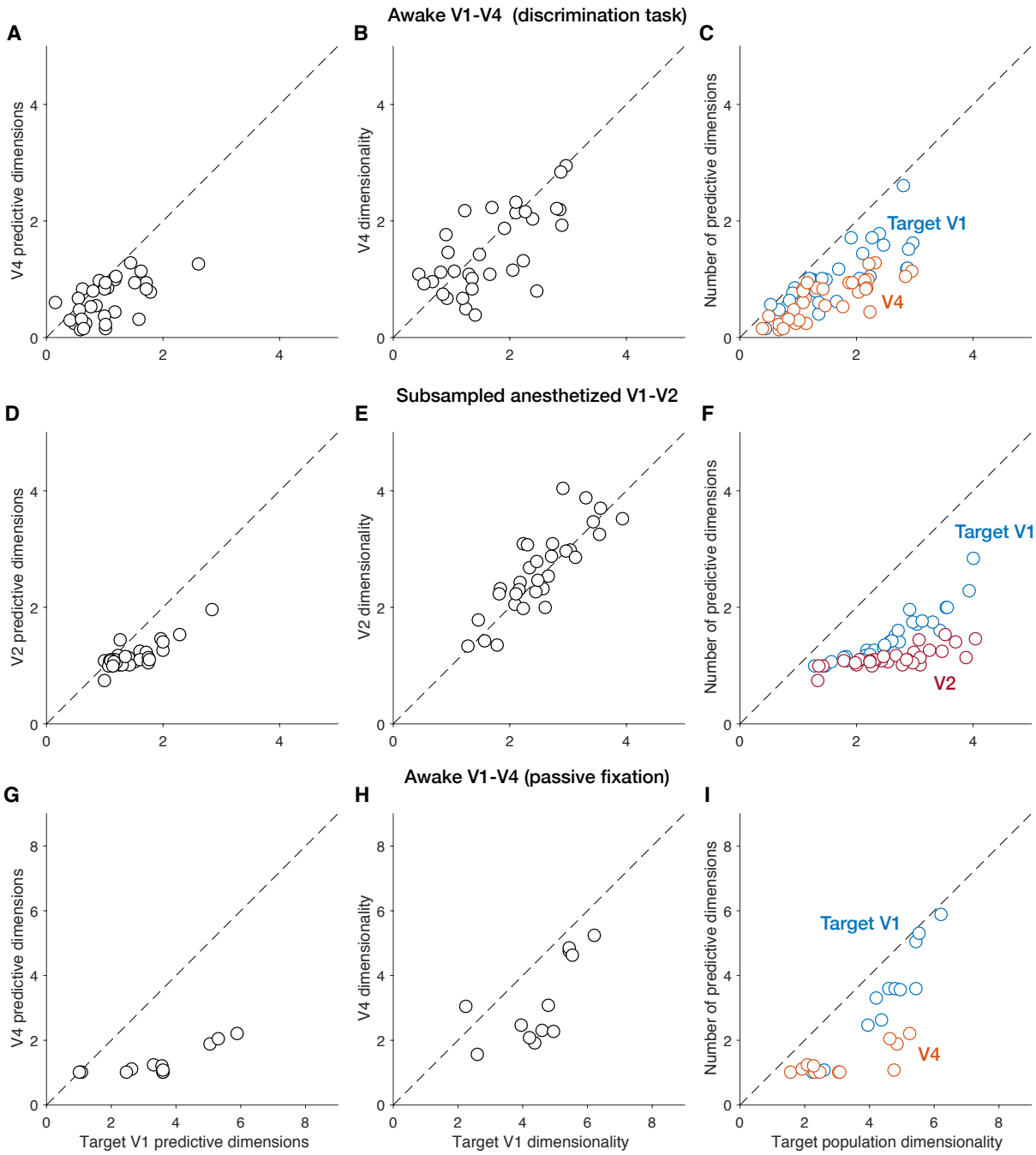
Awake V1-V4 (discrimination task)
A
B
C
V4 predictive dimensions
V4 dimensionality
Number of predictive dimensions
Target V1
V4
Subsampled anesthetized V1-V2
D
E
F
V2 predictive dimensions
V2 dimensionality
Number of predictive dimensions
Target V1
V2
Awake V1-V4 (passive fixation)
G
H
I
V4 predictive dimensions
V4 dimensionality
Number of predictive dimensions
Target V1
V4
Target V1 predictive dimensions
Target V1 dimensionality
Target population dimensionality

Figure S2, related to Figures 4 and 5. Interactions between V1 and V4 in awake animals are also low-dimensional.

Our V1-V2 recordings were performed in sufentanil-anesthetized animals. To test whether interactions between cortical areas are also low dimensional in awake animals, we analyzed simultaneous recordings from neuronal populations in V1 and V4, in two monkeys performing a grating orientation discrimination task, and a third monkey performing a fixation task.

(A-C) The two first animals were required to fixate on a small target ($0.1°$; fixation window diameter $1.4°$). After a delay of 200 ms, a drifting grating was presented for 200 ms (2-4 cpd, 6 Hz, 100% contrast). Animals were required to maintain fixation for an additional 200 ms, before reporting their decisions by making saccades to two choice targets. Saccades to the vertical choice target (i.e., above the fixation point) were rewarded if the orientation was less than $45°$; saccades to the horizontal choice target (i.e., on the horizontal meridian) were rewarded when stimulus orientation was larger than $45°$. Trials in which a $45°$ grating was presented were rewarded randomly.

After a training period of roughly 6 months, we implanted 48 channel Utah arrays in V1 and V4. The spatial receptive fields of the two populations were overlapping, such that the V1 receptive fields lay entirely within the aggregate receptive fields of the V4 population. We then recorded from V1 and V4 simultaneously while animals performed the discrimination task on gratings which covered the aggregate receptive fields of the recorded populations (stimulus diameters were $1 - 4°$, centered at eccentricities of $0.5 - 2.5°$). During recording sessions, we presented 7 orientations centered at $45°$ and chosen to straddle the slope of the psychometric function ($37.5°$ to $52.5°$), plus the two extremes ($0°$ and $90°$). The probability of presenting a $45°$ orientation in most sessions was twice that of the other orientations, whose presentation was equally likely. Additional details of the recording approach are provided in Jia et al. (2013) and Arandia-Romero et al. (2016).

We analyzed 145 recording sessions in the same way we analyzed the anesthetized recordings in the main text (see STAR Methods). Due to the limited number of trials for each grating orientation, we jointly analyzed all stimulus conditions (after removing the corresponding stimulus PSTHs) and used a 200 ms bin size. We treated each V1-V4 recording session as an independent data set. To form source and target V1 populations, we randomly split the V1 populations in half, and then randomly selected neurons from the V4 population to match the size of the target V1 population. For data sets for which the number of V4 neurons was smaller than the target V1 population, all V4 neurons were used and the extra target V1 neurons were re-assigned to the source V1 population. The process of randomly selecting the source and target V1 and V4 populations was repeated 50 times for each data set.

Unlike for the V1-V1 and V1-V2 comparison in the main text, we found that the predictive performance and population dimensionality for V4 were lower than for the corresponding target V1 populations. To ensure that our results would not be explained by these differences, we matched the joint distributions for these quantities for both target populations. This was done by selecting a subset of recording sessions such that the joint distributions of predictive performance and target population dimensionality were matched for both target populations. Since the target populations were paired (i.e., for each recording session, we used a common source V1 population to predict a target V1 and a V4 population), we could not separately sub-select from the target V1 and V4 populations to construct matched sets of recordings, resulting in an overconstrained matching problem. We thus used an approximate matching approach: we first computed the common joint predictive performance and population dimensionality distribution between the two target populations (i.e., the minimum between the joint predictive performance and population dimensionality histograms for each target population). We then randomly chose a target population and histogram bin, and randomly selected a session within that bin. In parallel, we selected the corresponding paired target population. We repeated this procedure until, for each bin, and for each target population, we had at least as many sessions as in the common histogram. The approximate matching procedure was repeated 100 times. We then used the matching for which the sum of the total absolute errors between the resulting target population histograms and the common histogram was smallest. This procedure yielded 32 matched sessions, containing 5 to 38 V1 neurons (average: 16 ± 7 s.d.) and 4 to 42 V4 neurons (26 ± 10 s.d.), with 414 to 4500 data points per session (1899 ± 1166 s.d.).

Having ensured a fair comparison between V1-V1 and V1-V4 interactions, we proceeded with our main analyses.

(A) We found that fewer predictive dimensions were required to predict V4 activity than target V1 activity (0.66 ± 0.06 for V1-V4; 1.04 ± 0.09 for V1-V1; one-sided Monte Carlo permutation test, $p < 10^{-3}$), consistent with analyses of within and between area interactions in the main text. Each open circle corresponds to one of the 32 matched recording sessions.

(B) As expected, due to the matching procedure, population dimensionality was similar for target V1 and V4 (1.66 ± 0.13 for target V1; 1.48 ± 0.12 for V4; two-sided Monte Carlo permutation test, $p > 0.05$).

(C) For a given target population dimensionality, fewer predictive dimensions were necessary to predict activity in the V4 population, compared with the target V1 population. The higher the target population dimensionality, the clearer the difference we observed between V1-V1 and V1-V4 interactions. For example, considering only recording sessions for which the target population dimensionality was above 2 for both target populations, the average number of predictive dimensions was 1.04 ± 0.06 for V1-V4 and

1.56 ± 0.18 for V1-V1 interactions (one-sided Monte Carlo permutation test, $p < 10^{-2}$). Thus, recordings from larger populations would likely yield still larger differences between V1-V1 and V1-V4 interactions.

(D-F) Although our analysis revealed a communication subspace between V1 and V4 populations recorded in awake animals, the results were less striking than for the V1 and V2 populations recorded in anesthetized animals (see Figure 5B). One possible explanation for this is that the awake recordings involved much smaller neuronal populations, which reduces the dimensionality of the population activity (Williamson et al., 2016) and thus any observable differences between within and across area interactions.

For comparison, we thus re-analyzed the V1-V2 data sets used in the main text to match the data available from the awake recordings (in panels A-C). To do so, we used a bin size of 200 ms, and randomly sub-selected the source and target V1 and V2 populations to match the population sizes of the V1-V4 recording sessions. Specifically, for each of the 32 awake recording sessions shown above, we randomly selected one of the 40 V1-V2 data sets, and re-analyzed it after randomly sub-selecting source and target neurons to match the size of the populations in the corresponding awake recording session (this was repeated 50 times, for different sub-selections of the source and target neurons; each symbol shows the average across all 50 repetitions). Additionally, since we could not rate mean-match the target V1 and V4 populations, we also did not employ the rate mean-matching procedure to the sub-selected target V1 and V2 populations.

(D) In these 'awake-matched' V1-V2 data, the number of V2 predictive dimensions was smaller than for target V1 (1.16 ± 0.04 for V1-V2; 1.44 ± 0.06 for V1-V1; one-sided Monte Carlo permutation test, $p < 10^{-3}$), but the difference between the two cases was smaller than in the main text (Figure 4) and similar to the awake data (panel A, above).

(E) Target population dimensionality was higher for V2 than for target V1 (2.64 ± 0.12 for V2; 2.45 ± 0.10 for target V1; two-sided Monte Carlo paired permutation test, $p = 0.02$), but the difference was smaller than in the main text (Figure 5) and again similar to the awake data (panel B, above).

(F) Comparing the number of predictive dimensions with the dimensionality of the corresponding target populations revealed that, for a given target population dimensionality, fewer predictive dimensions were necessary to predict activity in the V2 population, compared with the target V1 population. The difference between V1-V2 and V1-V1 interactions were similar to those found for the awake recordings (panel C, above). Thus, the reason why the results were less striking for the awake V1-V4 recordings than the anesthetized V1-V2 recordings is likely that there were fewer neurons recorded in the V1-V4 recordings.

(G-I) To verify that the small effect size found for the awake V1-V4 recordings was indeed due to the small number of neurons available, we analyzed three recording sessions from a third awake animal. In this animal, we had access to larger populations in V1 and V4 (192 electrodes distributed across the two areas; number of recorded V1 units: 34 to 57, mean 46.0 ± 11.5 S.D.; number of recorded V4 units: 58 to 69, mean 65.0 ± 6.1 S.D.). The animal passively fixated during the presentation of four different grating orientations (each stimulus was presented 558 to 1029 times, mean 749 ± 199 S.D., for 200 ms). We observed evidence of cross-talk between a small proportion of electrode pairs ($\sim 2\%$). We addressed this by removing one of the two affected electrodes. The mean pairwise correlations in the remaining pairs were small and similar within V1 and between V1 and V4 (average pairwise correlation: 0.051 ± 0.074 S.D. for V1-V1; 0.044 ± 0.040 S.D. for V1-V4).

Given the sizable number of V4 neurons in these recordings, the rate mean-matching procedure used for the V1-V2 recordings in the main text may assign more than half of the V1 neurons to the target population, which means the interaction becomes potentially limited by the number of source neurons. When this happened, we set the minimum source population size to be half the number of recorded V1 neurons, leading the source and mean-matched target populations to be of the same size (number of source population neurons: 17 to 29, mean 23.3 ± 5.1 S.D.; number of target population neurons (target V1 and V4): 17 to 28, mean 22.7 ± 4.7 S.D.).

(G) As with the V1-V2 results in the main text (see Figure 4C), fewer dimensions were required to predict activity in V4, compared to the target V1 population (average number of predictive dimensions: 1.3 ± 0.1 for V4; 3.5 ± 0.4 for target V1).

(H) In contrast to the V1-V2 results in the main text (see Figure 5A), the V4 population dimensionality was lower than the target V1 population dimensionality (target population dimensionality: 3.1 ± 0.44 for V4; 4.7 ± 0.4 for target V1). Importantly, the V4 population dimensionality was still much larger than the average number of predictive dimensions identified for the V1-V4 interaction (see next panel), indicating the V4 population dimensionality did not limit the dimensionality of the interaction.

(I) The results from analyzing these V1-V4 recordings agree with the V1-V2 results in the main text (see Figure 5B) and the other awake V1-V4 recordings presented in panels A-C. Namely, for a given target population dimensionality, fewer predictive dimensions were necessary to predict activity in the V4 population, compared with the target V1 population.

Finally, as for the anesthetized data (see Figure 7), we found that the dominant dimensions of the source V1 activity were better aligned with the predictive dimensions of the target V1 activity than with the predictive dimensions of the V4 activity (relative difference in prediction performance between the top source dominant dimension and the top predictive dimension was $10 \pm 2\%$ for target V1; $60 \pm 10\%$ for V4).

These results indicate that a communication subspace also exists for interactions between V1 and V4 in awake primate cortex, and thus the results in the main text cannot be ascribed to peculiarities of V1-V2 interactions or to anesthesia.

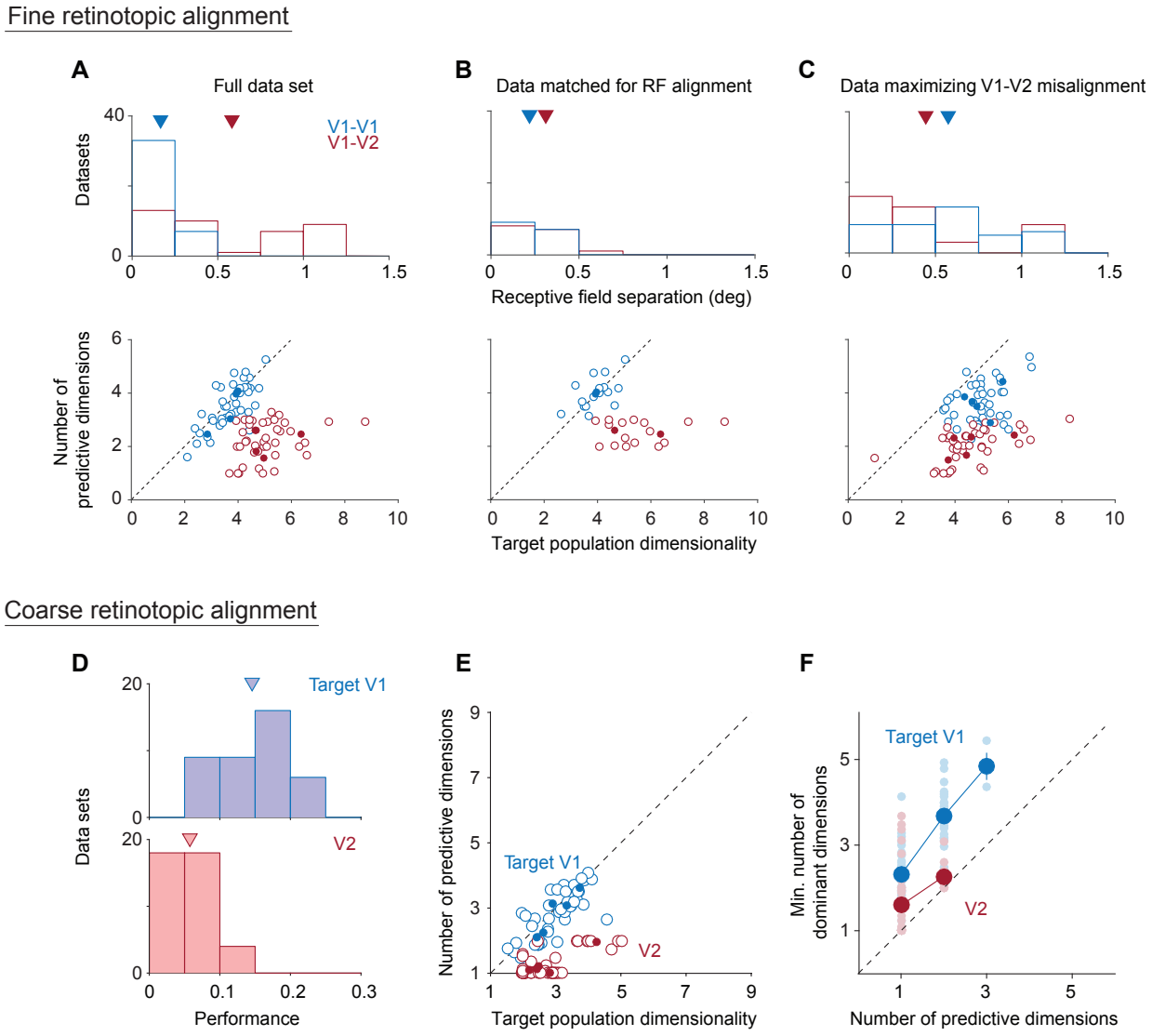

Figure S3, related to Figures 2, 5 and 7. Influence of V1-V2 retinotopic alignment on the communication subspace.

The V2 populations analyzed had spatial receptive fields that were closely aligned with the recorded source V1 populations. Because the target V1 population were selected from the same recording array as the source V1 population, these two populations also had well-aligned receptive fields. Nevertheless, one possible concern is that the difference in the number of V1 and V2 predictive dimensions was due to a subtle mismatch in receptive field alignment for V1-V1 compared to V1-V2.

(A) On average, the V1-V1 receptive field distance (center-to-center population spatial receptive field distance) was smaller than for V1-V2 (top; average V1-V1 receptive field difference: $0.17 \pm 0.02°$; average V1-V2 receptive field difference: $0.58 \pm 0.06°$; one-sided Monte Carlo permutation test, $p < 10^{-3}$).

To test whether the differences between V1-V1 and V1-V2 interactions (lower panel, replicating Figure 5B in the main text) could be explained by this difference in receptive field alignment, we repeated the analyses in the main text after matching the V1-V1 and V1-V2 alignment. We did this matching in two ways: (1) analyzing only the sessions for which V1-V2 alignment was as high as V1-V1 alignment; (2) selecting the target V1 populations so as to minimize their alignment with the source V1 populations.

(B) When we restricted our analysis to sessions for which V1-V2 alignment was as high as V1-V1 alignment (top; average V1-V1 receptive field difference: $0.21 \pm 0.04°$; average V1-V2 receptive field difference: $0.31 \pm 0.04°$; two-sided Monte Carlo permutation test, $p > 0.05$), we found V1-V1 and V1-V2 interactions were still strikingly distinct (bottom; compare with (a) bottom). Specifically, fewer predictive dimensions were necessary to predict the V2 activity than target V1 population activity (bottom; 2.6 ± 0.1 for V1-V2 vs 4.0 ± 0.2 for V1-V1; one-sided Monte Carlo paired permutation test, $p < 10^{-3}$), even though the V2 activity was higher dimensional than that of the target V1 population (bottom; 5.5 ± 0.3 for V2 vs 4.0 ± 0.2 for target V1; one-sided Monte Carlo paired permutation test, $p < 10^{-3}$). Likewise, V2 predictive dimensions were also not aligned with the source V1 dominant dimensions in these sessions (not shown).

(C) When we instead selected the target V1 population to minimize alignment with the source V1 populations (top; average V1-V1 receptive field difference: $0.57 \pm 0.05°$; average V1-V2 receptive field difference: $0.44 \pm 0.07°$; two-sided Monte Carlo permutation test, $p > 0.05$), we still found that fewer predictive dimensions were necessary to predict V2 than target V1 activity (bottom; 2.1 ± 0.1 for V1-V2 vs 3.7 ± 0.1 for V1-V1; one-sided Monte Carlo paired permutation test, $p < 10^{-3}$). In these data, both target populations had similar dimensionality (bottom; 4.6 ± 0.2 for V2 vs 4.9 ± 0.1 for target V1; two-sided Monte Carlo paired permutation test, $p > 0.05$). This similarity arises because the firing rate distributions for the target V1 and V2 populations were not mean-matched, as in the main text and above; the requirement to select a specific subset of V1 neurons with offset receptive fields (i.e., those in one corner of the array) precluded rate mean-matching. It is still case, however, that the smaller number of V2 than V1 predictive dimensions cannot be attributed to differences in target population dimensionality. In addition, it was still the case that V2 predictive dimensions were not well aligned with the source V1 dominant dimensions (not shown).

Together these results show that the differences between V1-V1 and V1-V2 interactions cannot be explained by subtle differences in retinotopic alignment with the two target populations. To test whether the V1-V2 communication subspace was at all sensitive to retinotopic alignment, we analyzed five additional sessions (each containing responses to gratings of 8 different orientations, for a total of 40 data sets), where there V1 and V2 receptive fields were clearly offset from one another average receptive field distance: $3.73°$ vs $0.58°$ for the datasets analyzed above).

(D) Our ability to predict responses in V2 was substantially lower when the source V1 population had retinotopically-offset receptive fields than when they are aligned (0.06 ± 0.01 for offset, red histogram vs. 0.15 ± 0.01 for aligned, shown in Figure 2B). The performance for the target V1 prediction, as expected, remained roughly the same (0.15 ± 0.01 for offset, blue histogram vs. 0.13 ± 0.01 for aligned, shown in Figure 2B).

(E) We then asked how many dimensions were necessary to predict the target populations in the data sets with retinotopically-offset receptive fields. When predicting target V1 activity, the number of predictive dimensions matched the dimensionality of the target V1 population (blue circles). When predicting the V2 population, on the other hand, the number of predictive dimensions was smaller than the dimensionality of the V2 population (red circles). Indeed, for most offset data sets only a single predictive dimension was needed to predict V2 (optimal dimensionality was one for 29 out of 40 data sets with retinotopically-offset receptive fields).

(F) To understand which source V1 population fluctuations were captured by the V2 predictive dimension(s), we assessed how similar the predictive dimensions were to the dominant dimensions in the source V1 population. Contrary to our findings for retinotopically-aligned recordings (cf. Figure 7C), we found that for the retinotopically-offset data sets predictive and dominant dimensions achieved similar performance when predicting the V2 population.

In summary, we found that V1-V2 interactions depend on retinotopical alignment: predictive performance is lower for recordings from populations with retinotopically-offset receptive fields, and these predictions frequently required only a single predictive dimension. The predictive dimensions for the offset data were similar to the dominant dimensions of activity in the source population.

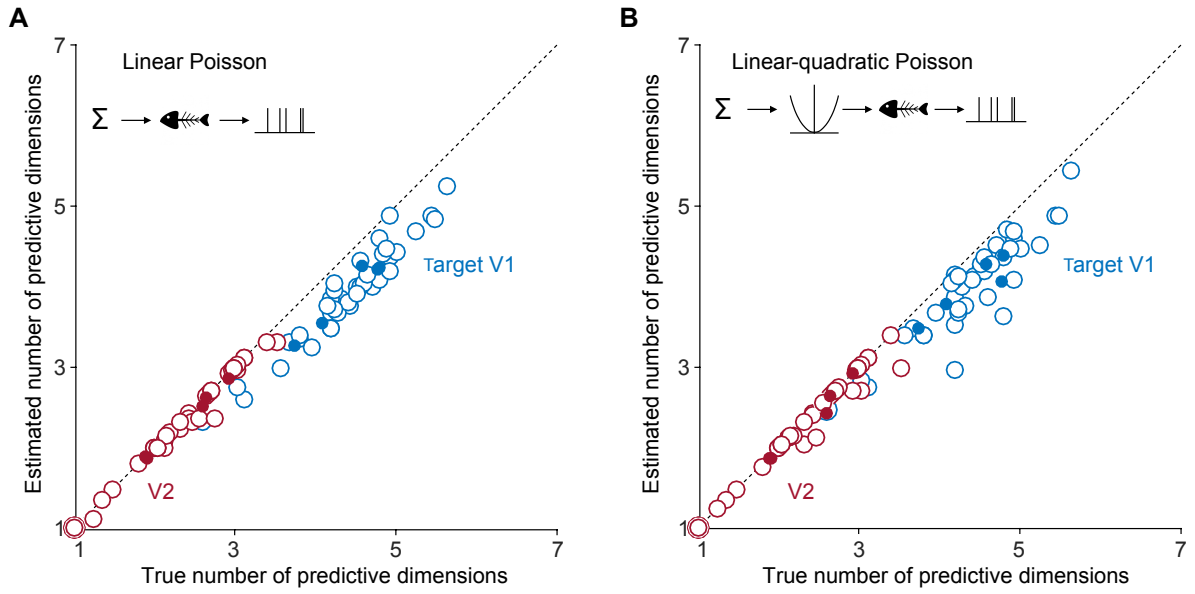

Figure S4, related to Figure 4. Reduced-rank regression recovers the correct number of predictive dimensions even when the mapping between populations is nonlinear.

Reduced-rank regression is a linear dimensionality reduction method which assumes additive Gaussian noise. We were concerned that our estimates of the dimensionality of V1-V1 and V1-V2 interactions might be inaccurate, either because interactions between populations are likely nonlinear or because neuronal variability is Poisson-like (not additive Gaussian). To test these possibilities, we applied reduced-rank regression to surrogate data sets in which target population variability was determined either by (1) a linear mapping from the source to target population followed by Poisson noise (linear-Poisson, or LP, model); or (2) by a linear-nonlinear mapping followed by Poisson noise (linear-nonlinear-Poisson, or LNP, model) (Schwartz et al., 2006).

(A) For the LP model, surrogate data were obtained using a similar procedure as in Figure S1A. Briefly, we defined the (number of source neurons by number of target neurons) mapping matrix B by applying reduced-rank regression, with the optimal number of dimensions, to the recorded source and target populations (without subtracting the PSTHs). We then generated the surrogate target rates $Y = XB$, where X is the recorded source activity (a number of datapoints by number of source neurons matrix). We obtained the target activity by generating Poisson spike counts based on the rates Y. To estimate the underlying number of predictive dimensions in the surrogate data, we applied reduced-rank regression to the residual activity in the source and target populations. We found that the estimated number of predictive dimensions matched the underlying model closely, especially for interactions involving a small number of predictive dimensions (open circles show average predictive performance for each data set; filled circles show average for each recording session).

(B) The LNP model was identical to the LP models, but the linear combination of the spike counts was passed through a quadratic nonlinearity before generating Poisson spike counts. As with the LP model, we found the number of predictive dimensions estimated with reduced-rank regression closely matched the dimensionality of the mapping matrix.

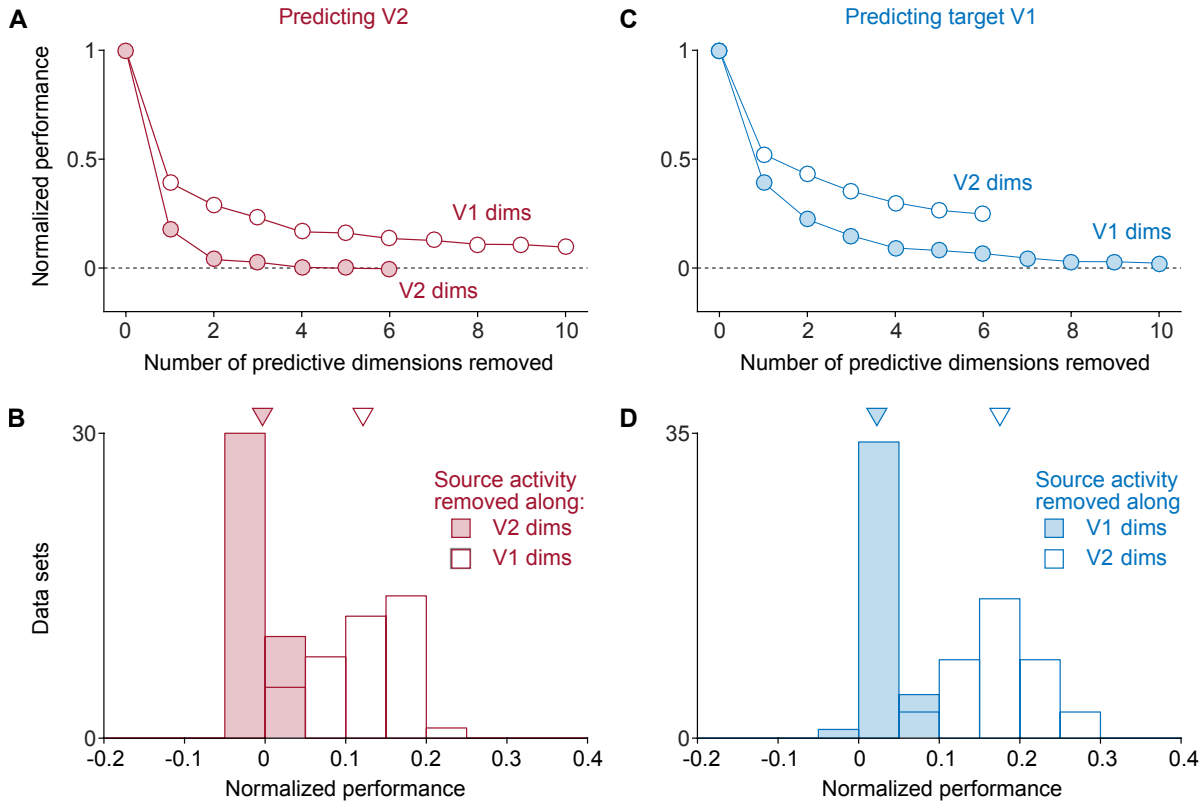

Figure S5, related to Figure 6. There exist dimensions of the source population activity that remain private regardless of the stimulus.

We showed that only a few dimensions of V1 activity are required to predict V2 responses. Consequently, a significant component of V1 population fluctuations remain private to V1 (from the perspective of V2). However, our analysis involved identifying predictive dimensions separately for each stimulus condition (i.e., data set). As a result, dimensions that are private for one stimulus condition may become predictive for another. Alternatively, some dimensions of the source V1 activity might remain private across all stimulus conditions. If so, these globally-private dimensions would constitute a subspace of V1 activity that is not related to V2 activity, perhaps representing processes internal to V1 which should not be relayed downstream.

(A) To determine whether there are globally-private dimensions of V1 activity, we first identified a joint predictive subspace – the subspace of the source V1 activity that is predictive of V2 activity across all stimulus conditions. We did so by simultaneously fitting the reduced-rank regression model to residual responses to all eight stimulus conditions in each recording session. This revealed that V1-V1 interactions involved as many predictive dimensions (9.7 ± 0.5) as the dimensionality of the target population (9.1 ± 0.5). V1-V2 interactions, on the other hand, involved fewer predictive dimensions (5.2 ± 0.3) than the target population dimensionality (9.6 ± 0.4). Thus, V1-V2 interactions

occurred through a communication subspace. We then applied the analysis of Figure 6 in the main text. Namely, we assessed how well we could predict V2 activity as we removed V1 activity that fell along the identified joint communication subspace.

As shown here for an example session, our ability to predict V2 activity quickly decreased as we removed the source V1 activity that fell along the V2 joint predictive subspace (filled circles). When we removed all activity in the joint communication subspace - in this case 6 dimensions - we were entirely unable to predict V2 responses. Removing the source V1 activity in the V1 joint predictive subspace had a smaller impact on V2 predictive performance (open circles). Predictive performance was normalized by the performance when no activity was removed. S.E.M. (across folds) is smaller than plotted circles.

(B) Since the joint predictive subspace is defined across all stimulus conditions, the activity outside the joint predictive subspace should not be predictive of V2 under any condition. Indeed, across data sets, average predictive performance vanished when all source activity aligned with the V1-V2 communication subspace had been removed (filled bars; mean fraction of original predictive performance across data sets: -0.002 ± 0.001 ; t-test $p > 0.05$). Removing the same number of V1 predictive dimensions had a smaller impact on performance (open bars; mean fraction of original predictive performance: 0.12 ± 0.01; one-sided Monte Carlo paired permutation test for difference between removing V1 and V2 predictive dimensions, $p < 10^{-3}$).

(C) We then performed a similar analysis, attempting to predict activity in the target V1 population. Specifically, we sought to determine whether the activity outside the V2 joint predictive subspace could be used to predict fluctuations in the target V1 population. The global private subspace would only be meaningful if there were substantial source activity that falls within it. As shown for an example session, target V1 predictive performance decreased more quickly when we removed source V1 activity along the target V1 predictive dimensions (filled circles), compared to removing source activity along the V2 predictive dimensions (open circles). Importantly, we retained some ability to predict activity in the target V1 populations, after removing all source V1 activity along the V2 joint predictive subspace. S.E.M. (across folds) is smaller than plotted circles.

(D) Across data sets, a substantial part of the target V1 activity could be predicted after removing the V2 joint predictive subspace (open bars; mean fraction of original predictive performance across conditions: 0.18 ± 0.01). Recall that when we identified the communication subspaces separately for each stimulus condition (Figure 6D), we retained 0.31 ± 0.01 of the original predictive performance. Thus, the predictive performance for the target V1 activity is largely preserved when going from a condition-specific V1-V2 private subspace to an across-condition V1-V2 private subspace. Removing the same number of target V1 predictive dimensions had a much larger effect on target V1 predictive performance (filled bars; mean fraction of original predictive performance

across conditions: 0.026 ± 0.003; one-sided Monte Carlo paired permutation test for difference between removing V1 and V2 predictive dimensions, $p < 10^{-3}$).

The difference between the effect of removing source activity along target V1 and V2 predictive dimensions stems from identifying distinct predictive dimensions for the target V1 and V2 populations. To check whether the differences we found between V1 and V2 predictive dimensions truly reflected differences in the way activity is related within versus across areas, we considered two separate target V1 populations, both mean-matched to the V2 population. We found that the predictive dimensions for both target V1 populations were significantly more similar to each other than they were to the V2 predictive dimensions, suggesting the results above reflect a difference between within and across area interactions.

Together, these results indicate that a substantial component of V1 activity is globally-private (across the stimuli tested), and not predictive of the V2 population. This is consistent with the results reported in Figure S6: while the predictive dimensions identified for the different stimuli are not identical, they differed only moderately, showing a high degree of overlap.

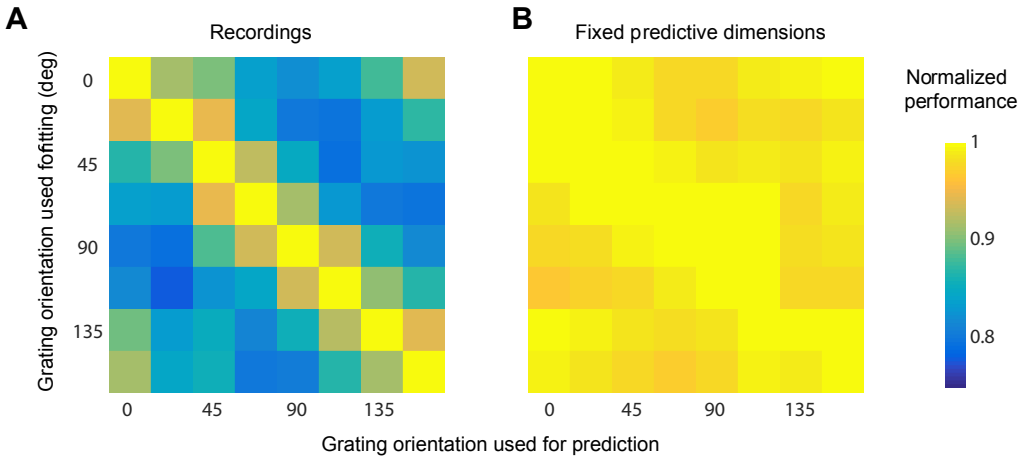

Figure S6, related to Figure 5. Predictive dimensions as local linear approximations of a globally nonlinear mapping.

In the main text, we identified predictive dimensions for each stimulus separately. An important motivation for this approach is that if the interactions between areas are nonlinear (Cowley et al., 2017), we can best apply linear methods to small perturbations in the population signals around a fixed operating point, which is precisely what the analysis of trial-to-trial fluctuations to a fixed stimulus accomplishes. However, a concern for this approach is that the predictive dimensions identified may change across stimulus conditions. Here, we investigate how the predictive dimensions identified from responses to different oriented gratings are related.

(A) To assess the similarity of the communication subspaces identified for different gratings, we determined the predictive dimensions for each stimulus (as in the main text) and then used these to predict responses evoked by the other stimuli. This was done by first projecting the source activity onto the identified communication subspace and then fitting a linear regression model between these projections and the target activity. Predictive performance was then normalized by the performance achieved by a model fit directly to the responses to that stimulus (i.e., without first projecting the responses onto the subspace defined for responses to another stimulus). Performance was measured using 10-fold cross-validation (i.e., for each fold, the model was fit to a training set pertaining to one stimulus and then used for prediction in test sets of all stimuli). If the communication subspaces were entirely distinct for different stimuli, performance should plummet when the regression model is fit using the subspace derived from responses to other gratings. Alternatively, if the communication subspaces are similar for different gratings, performance should be similar regardless of which responses are used to define the subspace.

We found that the communication subspaces are similar for different gratings. Each row corresponds to a different communication subspace and each column to the application of

those subspaces to a different stimulus condition. The diagonal elements thus indicate the normalized performance of identifying and applying the communication subspace to responses evoked by the same stimuli. The off-diagonal elements have normalized predictive performance values less than 1, indicating that the communication subspaces are not identical across the 8 stimulus conditions. However, the performance declined smoothly as the responses were evoked by stimuli of progressively different orientations. In addition, the drop in performance was modest. When the predictive subspace identified for one orientation was used to predict responses to an orthogonal grating - the most challenging scenario – performance was roughly 75% of that achieved for the subspace identified for the orthogonal grating responses. Performance is averaged across all sessions.

(B) For comparison, we applied the same analyses to data synthesized as in Figure S1. Briefly, we first defined a fixed linear mapping, using the residual responses to activity pooled across all conditions. We then generated target population activity for each stimulus by passing the corresponding source population activity through this fixed mapping, and added Poisson noise to each sample with mean given by the corresponding PSTH time bin. For these synthesized responses, we found that performance was uniformally high, when we identified the predictive dimensions using responses to one stimulus orientation and applied them to responses evoked by another. This analysis indicates that (1) the modest performance decrement in the physiological data (panel A) cannot be attributed to differences in the source population responses across orientations; (2) if the mapping between areas were strictly linear and fully identified by our analysis, we should not observe any decrease in performance. These results thus support the suggestion that the mapping between V1 and V2 is not strictly linear.

In summary, the communication subspaces identified for distinct stimuli were not identical. Furthermore, these differences could not be explain by changes in the statistics of the source and target populations. However, these subspaces changed smoothly and only moderately across all stimuli.

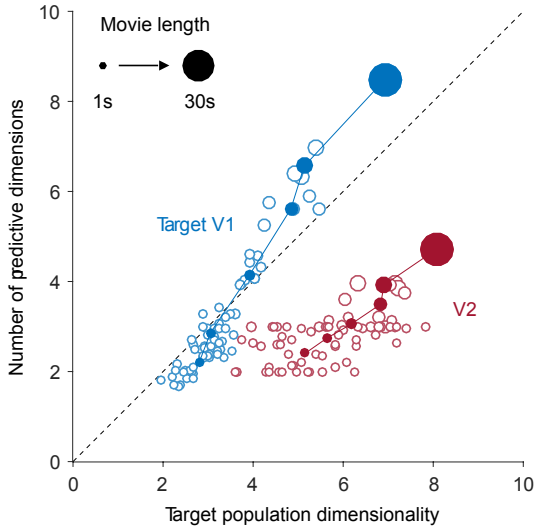

Figure S7, related to Figure 5. V1-V2 interactions driven by naturalistic stimuli also occur through a communication subspace.

The results in the main text are based on analyzing responses to oriented gratings. To understand the interaction between V1 and V2 for richer stimuli, we also analyzed V1 and V2 responses to repeated presentations of a natural movie.

We recorded from V1 (130 neurons) and V2 (18 units) in one anesthetized monkey, while presenting 300 repetitions of a 30s natural movie. Movies consisted of 750 unique frames, with each frame presented on 4 sequential monitor refreshes; given the refresh rate of 100 Hz, this yielded a video rate of 25 Hz. For our analysis, we divided the 30 s movie into shorter segments of either 1, 1.5, 3, 6, 10 or 30 s (the full movie), yielding 30, 20, 10, 5, 3 and 1 data sets, respectively. We then analyzed each data set independently. Note that different data sets (i.e., movie segments) correspond to activity evoked by distinct stimuli. As in the main text, we binned activity using a 100ms window, and subtracted the corresponding PSTH from each cell's response.

We found that the number of V1 predictive dimensions closely matched the dimensionality of the target V1 population (blue symbols, along the diagonal for all movie lengths, indicating by symbol size). V2 predictions, on the other hand, consistently required fewer predictive dimensions than the dimensionality of the V2 population (red symbols). Filled circles show the average across all movies of a given length. Open circles show the estimates for each movie segment (averaged across 25 rate mean-matched samples of the V1 and V2 populations).

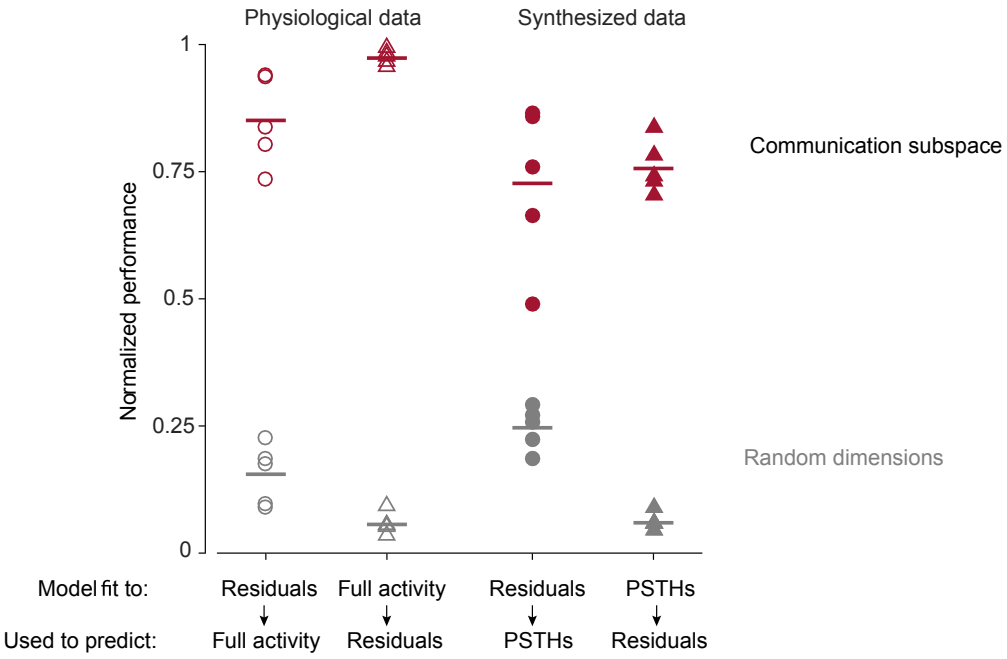

Figure S8, related to Figure 5. Predictive dimensions identified using residual activity can predict responses that include stimulus information.

In the main text, we identified predictive dimensions using residual activity, i.e., after subtracting from each trial the corresponding PSTH. While there are advantages to this approach, it is important to understand how the predictive dimensions identified in this way relate to stimulus processing. Have we lost critical components of the V1-V2 interaction by removing the stimulus component (or PSTH) from the activity? Or do the predictive dimensions identified for the residual activity largely capture these stimulus components as well?

To answer these questions, we directly compared the communication subspaces identified using residual activity with those identified when we did not subtract the PSTHs (termed here the "full activity"). We did so by estimating the predictive dimensions either from the residual activity (pooled across conditions, as in Figure S5) or from the full activity, and then used these subspaces to predict held-out sets of residual and full activity. For example, to predict full activity using the residual communication subspace, we first identified the residual predictive dimensions on a residual activity training set, then projected the full source V1 activity onto the subspace spanned by these dimensions and used these projections to predict the full V2 activity. The predictive performance was then normalized by the performance obtained from predicting the full activity using the subspace fit to the full activity training set. This process was repeated 10 times, using 10-fold cross-validation. There was no overlap between the data used to fit the models and the data used to quantify predictive performance.

When predicting the full activity, we found that the subspace identified using the residual activity retained 0.85 ± 0.04 of the predictive performance achieved by the predictive dimensions identified for the full activity itself (red open circles). For comparison, we also quantified the performance retained when using a random subspace with the same dimensionality as the residual subspace. We selected these random predictive subspaces by fitting a reduced rank regression (RRR) model to the same training residual activity, but after shuffling the sample order of the two areas independently, thereby destroying V1-V2 covariability. These random communication subspaces retained a much smaller fraction of the performance (black open circles; 0.16 ± 0.03).

Using the subspace identified for the full recordings to predict the residual activity resulted in an even higher fraction of performance retained (red open triangles; 0.97 ± 0.01 vs. 0.06 ± 0.01 for random dimensions). This was expected, given that the full activity is composed of the stimulus component and trial-to-trial variability. Thus predicting the full activity involves also predicting the residual activity. Consistent with this statement, more predictive dimensions were needed to predict the full activity than to predict residual activity (6.75 ± 0.3 for full activity vs 5.1 ± 0.3 for residual activity).

One might worry that the ability of a subspace identified using residual activity to predict the full activity is not because this subspace captures the stimulus component, but rather because the stimulus component is small compared to the trial-to-trial variability (i.e., the full activity is dominated by trial-to-trial fluctuations). If so, the residual subspace would retain most of the full activity predictability even if it failed to capture the stimulus component. To test whether this was the case, we used the residual subspace to predict the stimulus component. We did so by creating surrogate data with realistic levels of trial-to-trial variability for which all covariability, within and across areas, was solely due to the PSTHs. In other words, the only interactions between the two areas were those induced by the PSTHs. Specifically, for each stimulus, we generated 400 trials (the same number as with the real data) of responses by drawing from a Poisson distribution where the mean, for each time bin, was given by each neuron's PSTH. We used this synthesized data because directly fitting a RRR model to the PSTHs leads to overfitting, due to the small number of samples. This approach (fitting RRR to synthesized data created in this way) can be thought of as a form of regularization, similar to that employed by ridge regression but using Poisson noise. For these synthesized data, the residual subspace retained most of the predictability achieved by identifying predictive dimensions using the synthesized data itself (red filled circles; 0.73 ± 0.07 vs. 0.25 ± 0.02 for random dimensions, in black filled circles). The subspace identified using the synthesized data was also able to predict residual activity (red filled triangles; 0.76 ± 0.02 vs. 0.06 ± 0.01 for random dimensions in black filled triangles).

These results show that there is high overlap between the dimensions that are most predictive of residual activity and the dimensions that are most predictive of the stimulus components of the activity. Thus, one can use covariations of trial-to-trial fluctuations to

identify the relationship between mean activity of neurons in one area and those in another, at least for the limited stimulus ensemble used here.

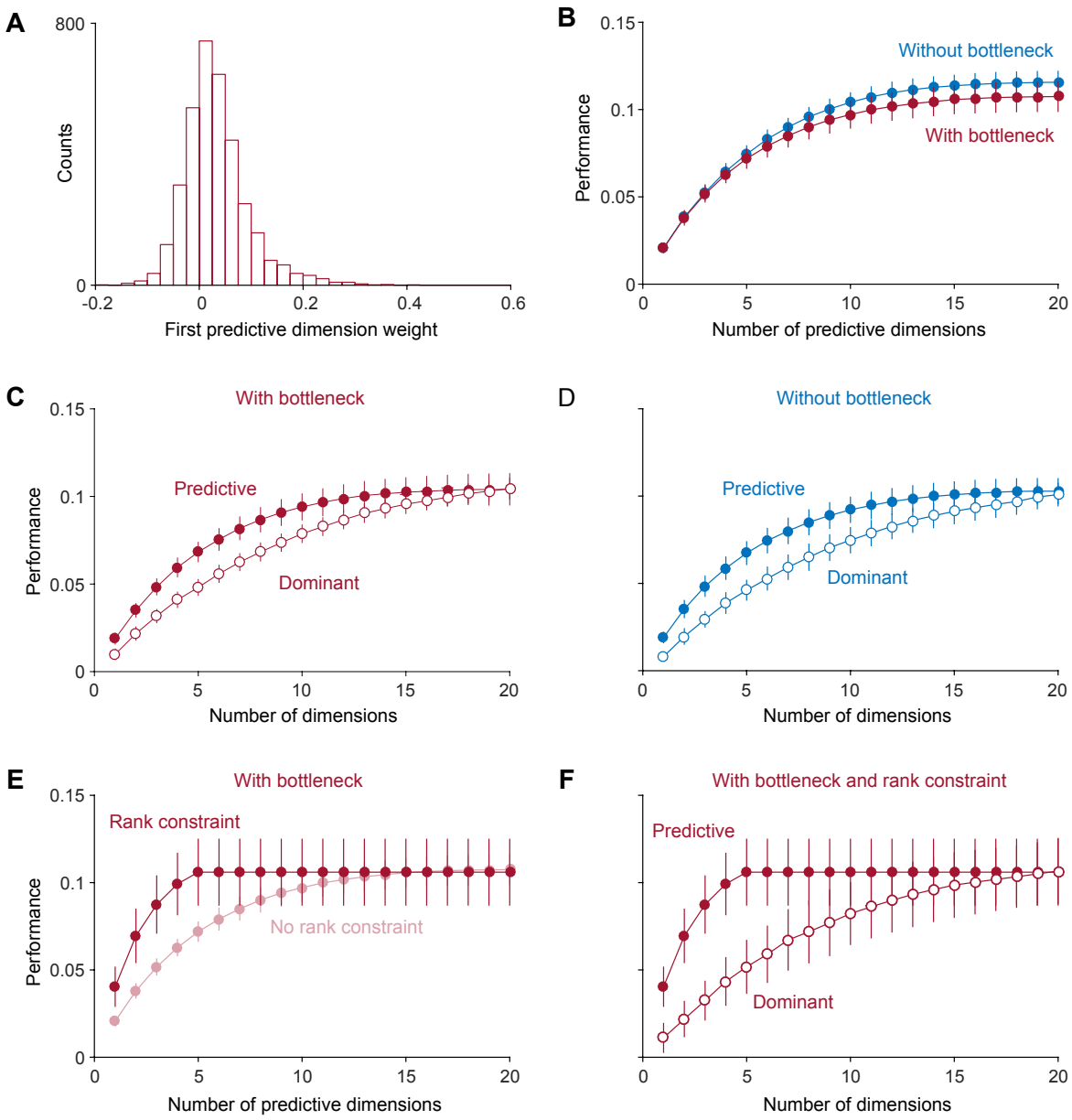

Figure S9, related to Figures 5 and 7. The small number of predictive dimensions cannot be explained by a physical bottleneck.

Only a fraction of the neurons in the output layers of V1 directly project to V2 (Sincich and Horton, 2005; Sincich et al., 2010). Thus, it is likely that not all of the recorded V1 neurons project to the recorded V2 population. We wondered whether the small number of V2 predictive dimensions could be a consequence of the linear regression model relying on those few projecting V1 neurons. If this were case, the V1-V2 communication subspace would be a straightforward consequence of a "physical bottleneck" between these areas. To test this possible explanation for our results, we performed two additional analyses.

(A) First, we studied the structure of the predictive dimensions, after fitting the reduced-rank regression model to the (z-scored) recorded data. If the predictive dimensions rely on only a few V1 neurons, we would expect the magnitude of a few regression weights to be large and those for the remaining source neurons to be low. We found no evidence of this: the weights for the V1 neurons varied broadly and showed no signs of bimodality. Plotted is the histogram of weights associated with the first predictive dimension across all data sets. Note that since the sign of the predictive dimension is arbitrary, we defined it for each data set as the sign for which most weights are positive.

(B) Second, we conducted a simulation in which a large source population influenced a target population via a physical bottleneck. In this simulation, not all source neurons project to the target population, and only a fraction of source and target neurons are observed, mimicking the situation in our recordings. We generated the source activity (10,000 neurons) in the same way as in Figure S1B. This was done to ensure that the correlations in the synthesized population were similar to those in the data, both in magnitude and structure. The results presented here were largely insensitive to the precise structure of correlations within V1. To generate the activity in the target population (30 neurons), we defined the response of each target neuron using a weighted sum of the activity of the source population, with the weights chosen from a standard Gaussian distribution. Importantly, the weights were chosen such that 95% of the source neurons did not project to the target population (i.e., 95% of the source neurons had their weights set to 0 for all target neurons) – a physical bottleneck in which only 5% of source neurons provide input to the target population. For comparison, we also conducted an additional simulation in which all neurons in the source population project downstream (i.e., all weights were chosen from a standard Gaussian distribution and none were set to zero so there is no physical bottleneck). The weights for this model were adjusted so that the covariance structure of the activity of the two target populations was matched. We did this using the singular value decomposition of the target population data matrix of the model without the physical bottleneck $Y_u = USV^T$, and then correcting it using:

$$\bar{Y}_u = Y_u V D V^T$$

where $\bar{Y}_u$ is the corrected target population data matrix of the model without the physical bottleneck. D is a diagonal matrix with entries σ_i^b/σ_i^u where σ_i^b and σ_i^u ($i = 1, ...30$) are the ordered singular values of the target populations with and without a physical bottleneck, respectively. This ensures both target populations have the same eigenvalues, and therefore any difference in the number of estimated predictive dimensions are not due to differences in the second order statistics of the target populations (recall that we controlled for this in the recorded data by estimating the dimensionality of the target populations, Figure 5). To estimate the mapping dimensionality, we applied reduced-rank regression to a random subset of 100 source neurons and all of the target neurons. Note that under this linear model generating 30 target responses is equivalent to randomly selecting 30 neurons from a larger target population. This entire process was repeated 25 times, using different random instantiations of the weighting matrices described above and different random selections of the observed source neurons.

We found that the physical bottleneck in the surrogate data had little influence on the number of estimated predictive dimensions. As shown here, the optimal number of predictive dimensions was roughly the same whether or not a physical bottleneck was present (15.6 ± 0.3 with physical bottleneck and 15.2 ± 0.2 without physical bottleneck; two-sided Monte Carlo paired permutation test, $p > 0.05$; error bars indicate standard deviation across random source subsets and choices of the mapping matrices). It might seem counterintuitive that the estimated mapping dimensionality was the same whether or not all observed neurons projected to the target population. Furthermore, although only 5 observed neurons directly influence the target population on average in the simulation with a physical bottleneck, the estimated dimensionality of the interaction is three times that number. Both of these findings arise due to the covariance structure of the source population activity, which contains dimensions of covariability that are shared across many neurons. As a result, even if a neuron does not project to the target population, it can still be predictive of the target population activity if it covaries with neurons that do project.

In these simulations, the activity of the target populations was completely determined by the activity of the source population (i.e., no noise was added). Since we had 500 (or 10,000 when no physical bottleneck was present) source neurons projecting to 30 target neurons, the true dimensionality of the interaction was 30. However, due to the limited sampling of the source population and finite data, the estimated dimensionality is smaller than the true value.

(C, D) We also tested whether the presence of a physical bottleneck could explain our finding that the dimensions that are most predictive of the V2 population are not well aligned with the dominant dimensions of the source V1 population (Figure 7). We found that the difference in predictive performance between predictive and dominant dimensions was small in these simulations, whether or not a physical bottleneck was present (panel C: with physical bottleneck, panel D: without physical bottleneck; error bars indicate standard deviation across random source subsets and choices of the mapping matrices).

(E) In order to qualitatively reproduce the results in the main text we had to restrict the linear mapping between the source and target populations to be low-dimensional. That is, the sets of weights relating each V2 neuron to the source V1 population were chosen as a linear combination of a small set of basis weights (in this example, a 5 dimensional basis set). Imposing this structure reduced the estimated optimal number of predictive dimensions to 4.6 ± 0.1.

(F) Furthermore, due to the way in which the low-dimensional mapping was chosen, the predictive dimensions now lay in a subspace of the source population activity that was randomly oriented with respect to the dominant dimensions of this population. As a result, using the source dominant dimensions to predict activity in the target population

led to worse performance, when compared with the performance afforded by the identified predictive dimensions.

We conclude that that existence of a physical bottleneck cannot explain: (1) the communication subspace; or (2) the misalignment of the predictive and dominant dimensions observed in the recorded activity. Only by enforcing a low-dimensional mapping were we able to reproduce these findings.

Finally, we note that there may be other scenarios, none of which apply to our data, in which a physical bottleneck can result in the identification of a communication subspace. For instance, if all of the neurons in the source and target areas are recorded, the true weights of the connections between areas can be recovered. If most of these weights are zero (i.e., most source neurons do not project), then the mapping between areas will be low-dimensional (i.e., there will be a communication subspace). Alternatively, a low-dimensional mapping can arise when the activity of the projection neurons is independent of the rest of the source population, even if only a subset of the source population is recorded. This is unlikely to be the case in our data given the well-established correlations among V1 neurons. Because neither of these scenarios likely applies to the current analyses, we believe that the most likely explanation for our results is that the mapping between V1 and V2 is low-dimensional.