Title: Multi-Connection Pattern Analysis: Decoding the Representational Content of Neural Communication

Authors: Yuanning Li[1,2,3*] and Avniel Singh Ghuman[1,2,3]

Affiliations:
[1]Center for the Neural Basis of Cognition.
[2]Program in Neural Computation, Carnegie Mellon University and University of Pittsburgh.
[3]University of Pittsburgh, Department of Neurological Surgery.

***Correspondence to**:

Yuanning Li
S906 Scaife Hall
3550 Terrace Street
Pittsburgh, PA 15261
ynli@cmu.edu

Keywords: multi-connection pattern analysis, functional connectivity, multivariate statistical analysis, machine learning, decoding, representation similarity analysis, functional magnetic resonance imaging (fMRI)

Abstract

What information is represented in the interactions between neural populations is unknown due to the lack of multivariate methods for decoding the representational content of neural communication. Here we present Multi-Connection Pattern Analysis (MCPA), which probes the involvement of distributed computational processing and probe the representational structure of neural interactions. MCPA learns mappings between the activity patterns as a factor of the information being processed. These maps are used to predict the multivariate activity pattern from one neural population based on the activity pattern from another population. Simulations demonstrate the efficacy of MCPA in realistic circumstances. Applying MCPA to fMRI data shows that interactions between visual cortex regions are sensitive to information that distinguishes individual natural images. These results suggest that image individuation occurs through interactive computation across the visual processing network. Thus, MCPA can be used to assess the information processed in the coupled activity of interacting neural circuits.

Introduction

Since at least the seminal studies of Hubel and Wiesel (Hubel & Wiesel, 1959) the computational role that neurons and neural populations play in processing has defined, and has been defined by, how they are tuned to represent information. The classical approach to address this question has been to determine how the activity recorded from different neurons or neural populations varies in response to parametric changes of the information being processed. Single unit studies have revealed tuning curves for neurons from different areas in the visual system responsive to features ranging from the orientation of a line, shapes, and even high level properties such as properties of the face (Desimone, Albright, Gross, & Bruce, 1984; Hubel & Wiesel, 1959; Tsao, Freiwald, Tootell, & Livingstone, 2006). Multivariate methods, especially pattern classification methods from modern statistics and machine learning, such as multivariate pattern analysis (MVPA), have gained popularity in recent years and have been used to study neural population tuning and the information represented via population coding in neuroimaging and multiunit activity (Cox & Savoy, 2003; Ghuman et al., 2014; Haxby et al., 2001; Haynes & Rees, 2006; Hirshorn et al., 2016; Kamitani & Tong, 2005; Poldrack, 2011; Polyn, Natu, Cohen, & Norman, 2005). These methods allow one to go beyond examining involvement in a particular neural process by probing the nature of the representational space contained in the pattern of population activity (Edelman, Grill-Spector, Kushnir, & Malach, 1998; Haxby, Connolly, & Guntupalli, 2014; Kriegeskorte & Kievit, 2013).

Neural populations do not act in isolation, rather the brain is highly interconnected and cognitive processes occur through the interaction of multiple

populations. Indeed, many models of neural processing suggest that information is not represented solely in the activity of local neural populations, but rather at the level of recurrent interactions between regions (Grossberg, 1980; Kveraga, Ghuman, & Bar, 2007; Lee & Mumford, 2003). However previous studies only focused on the information representation within a specific population (Freiwald, Tsao, & Livingstone, 2009; Ghuman et al., 2014; Haxby et al., 2001; Hirshorn et al., 2016; Nestor, Plaut, & Behrmann, 2011; Tsao et al., 2006), as no current multivariate methods allow one to directly assess what information is represented in the pattern of functional connections between distinct and interacting neural populations. Such a method would allow one to assess the content and organization of the information represented in the neural interaction. Thus, it remains unknown whether functional connections passively transfer information between encapsulated modules (Fodor, 1983) or whether these interactions play an active computational role in processing.

Univariate methods exist for assessing active communication between neural regions, notably the psychophysiological interactions (PPI; (K. J. Friston et al., 1997)) method. However, when compared with univariate methods, it has been noted that multivariate methods allow for "more sensitive detection of cognitive states," "relating brain activity to behavior on a trial-by-trial basis," and "characterizing the structure of the neural code" (Norman, Polyn, Detre, & Haxby, 2006). Thus, a multivariate pattern analysis method for functional connectivity analysis is required to decode the representational content of interregional interactions.

In this paper, we introduce a multivariate analysis algorithm combining functional connectivity and pattern recognition analyses that we term Multi-Connection Pattern

Analysis (MCPA). MCPA works by learning the discriminant information represented in the shared activity between distinct neural populations by combining multivariate correlational methods with pattern classification techniques from machine learning in a novel way. As a rough analogy MCPA is to PPI as MVPA is to a t-test or ANOVA in that MVPA methods combine multivariate models of local neural activity with pattern classification techniques to go beyond univariate statistics in a similar way that MCPA goes beyond univariate PPI. Of note, unlike previous approaches that statistically test the absolute or relative connectivity between two populations, MCPA estimates the activity pattern in one neural population based on the activity in another neural population, allowing one to make predictions about new observations and assess classification accuracy. This is much like MVPA methods allow for prediction and classification accuracy at the local level, which t-tests and ANOVAs do not.

The MCPA method consists of an integrated process of learning connectivity maps based on the pattern of coupled activity between two populations A and B conditioned on the stimulus information and using these maps to classify the information representation in shared activity between A and B in test data. The rationale for MCPA is that if the activity in one area can be predicted based on the activity in the other area and the mapping that allows for this prediction is sensitive to the information being processed, then the areas are communicating with one another and the communication pattern encodes the information being processed. Thus, MCPA simultaneously asks two questions: 1) Are the multivariate patterns of activity from two neural populations correlated? (i.e. is there functional connectivity?) and 2) Does the connectivity pattern adaptively change based on the information being processed? This is operationalized by

learning a connectivity map that maximizes the multivariate correlation between the activities of the two populations in each condition. This map can be thought of like the regression weights that transform the activity pattern in area A to the activity pattern in area B (properly termed "canonical coefficients" because a canonical correlation analysis [CCA] is used to learn the map). These maps are then used to generate the predictions as part of the classification algorithm. Specifically, a prediction of the activity pattern in one region is generated for each condition based on the activity pattern in the other region projected through each mapping. Single trial classification is achieved by comparing these predicted activity patterns with the true activity pattern (see Figure 1 for illustration). With MCPA single trial classification based on multivariate functional connectivity patterns is achieved allowing the nature of the representational space of the interaction to be probed.

We present a number of simulations to validate MCPA for a realistic range of signal-to-noise ratios (SNR) and to show that MCPA is insensitive to local information processing. We apply MCPA to examine the inter-regional representation for categorical visual stimuli in visual cortex using functional magnetic resonance imaging (fMRI) data. Specifically, we show that the interactions between regions of the visual stream (V1, V2, V3, V4, and LO) are sensitive to information about individual natural images. These results demonstrate that MCPA can be used to probe the nature of representational space resulting from information processing distributed across neural regions.

Results

Simulations

We used simulations to test and verify the performance and properties of MCPA on synthetic data. Specifically, synthetic data representing neural activity of two distinct populations and the information represented in the interaction between those populations were manipulated to construct different testing conditions.

In the first simulation, we evaluated the ability of MCPA to detect information represented in the functional connectivity pattern when it was present as a factor of the SNR and the number of dimensions of the data. The mean and standard error of *d*' from 100 simulation runs for each particular setup (dimensionality and SNR) are shown in Figure 2a. The performance of the MCPA classifier increased when SNR or effective dimensionality increased. Classification accuracy saturated to the maximum when SNR and number of dimensions were high enough (SNR > 10 dB, dimensionality > 10). The performance of MCPA was significantly higher than chance ($p < 0.01$, permutation test) for SNRs above -5 dB for all cases where the dimensionality was higher than 2, when the pattern of the multivariate mapping between the activity was changed between conditions. It is notable that significant MCPA classification was seen despite there being no local information present in either of the two simulated populations ($p > 0.1$ for all cases using MVPA).

The first control simulation was designed to confirm that when two unconnected populations both carry local discriminant information, MCPA would not be sensitive to that piece of information. As shown in Figure 2b, MCPA did not show any significant classification accuracy above chance (*d*' = 0) as *k* changed. On the other hand, the

MVPA classifier that only took the data from local activity showed significant classification accuracy above chance level and the performance increased as local discriminant information increased.

The second control simulation was designed to test if MCPA would be sensitive to changes in local discriminant information when there was constant information coded in neural communication. Local discriminant information was injected into the populations by varying the ratio of the standard deviation (k) between the two conditions. When MVPA was applied to the local activity, increasing classification accuracy was seen as *k* became larger (Figure 2c). This result confirmed that discriminant information was indeed encoded in the local activity in the simulation. On the other hand, the performance of MCPA did not change with the level of local discriminant information, demonstrating that MCPA is only sensitive to changes in information contained in neural interactions.

The final control simulation tested whether MCPA is simply sensitive to the presence of functional connectivity between two populations *per se* or is only sensitive to the whether the functional connectivity contains discriminant information. Specifically, are local discriminant information in two populations, and a correlation between their activity, sufficient for MCPA decoding? It should not be considering that MCPA requires that the pattern of the mapping between the populations to change as a factor of the information being processed (see Figure 1). For example, the local activity in either or both populations could code for the information being processed, but the mapping between the activity in each region could be constant and insensitive to the changes in conditions, e.g. the CCA coefficients could be the same. This would be the case if each

population was an informationally encapsulated module where information transfer occurs in the same way regardless of the stimulus being processed or cognitive state. In this case, one would not want to infer that distributed processing was taking place because the nature of the interregional communication is not sensitive to the computation being performed (e.g. the information transfer is passive, rather than reflecting distributed computational processing) and all of the information processing is done locally in each population. The final control simulation was designed to assess whether MCPA is sensitive to the case where two populations communicate, but in a way that would not imply distributed computational processing. Specifically, neural activity in areas A and B were simulated such that local discrimination was possible in each population and the activity of the two populations was correlated, but the interaction between them was invariant to the information being processed. Figure 2d shows that in this case MCPA did not classify the activity above chance, despite significant correlation between the regions and significant local classification (MVPA). Thus, functional connectivity between the populations is a necessary, but not sufficient, condition for MCPA decoding. Therefore, MCPA is only sensitive to the case where the mapping itself changes with respect to the information being processed, which is a test of the presence of distributed neural computation.

fMRI Data

To assess its performance on real neural data, MCPA was applied to Blood-oxygen-level-dependent (BOLD) fMRI measurements of human occipital visual areas, in two subjects (S1 and S2) during passive viewing of 13 repetitions of 120 natural images (Kay, Naselaris, & Gallant, 2011; Kay, Naselaris, Prenger, & Gallant, 2008; Naselaris,

Prenger, Kay, Oliver, & Gallant, 2009). MCPA was used for single-trial classification of these images for the interactions between V1-V2, V2-V3, V3-V4, and V4-lateral occipital (LO) cortex (e.g. 4 total region pairs * 2 subjects; see Figure 5 of Naselaris et al. (Naselaris et al., 2009) for depictions of these regions in these subjects). Across the 8 pairs of regions the mean classification accuracy (sensitivity index *d'*) of the single trial classification was 0.405 (SD = 0.094), with all of the pairs showing significant classification at $p < 0.05$ corrected for multiple comparisons. In both subjects, MCPA classification accuracy declined going up the classic visual hierarchy. For S1, the classification accuracies (*d'*) were 0.451 (V1-V2), 0.432 (V2-V3), 0.403 (V3-V4), and 0.333 (V4-LO); for S2, the classification accuracies (*d'*) were 0.563 (V1-V2), 0.469 (V2-V3), 0.317 (V3-V4), and 0.274 (V4-LO).

We then performed a representational similarity analysis (RSA) to compare the representational structure of the interaction between two regions to the structure of the representation within each region. This was done by correlating the representational similarity matrix from MCPA to ones calculated using MVPA in each region. To increase our power we performed this RSA at the category level (animals, buildings, humans, natural scenes, and textures) rather than the single image level because the dataset contained many more repetitions per category than per image (Figure 3). This yielded a total of 16 correlations (8 MCPA-based matrices correlated with each of the two regions that contribute to each MCPA). 14 out of the 16 correlations were negative, many showing large negative correlation coefficients (See Table 1, mean Pearson's $\rho = -0.349$, SD = 0.299, range = -0.771 ~ 0.251). In other words, categories that were relatively easy to decode based on the activity within regions using MVPA were relatively more difficult

to decode based on the shared activity between that region and the other regions in the visual stream using MCPA and vice versa (Figure 3). This suggests that the communication between regions represents information that has not been explained aspects by local computational processes. This is consistent with coding for error propagation or prediction errors (K. J. Friston, 2005; Rumelhart, Hinton, & Williams, 1986).

Discussion

This paper presents a novel method to assess the information represented in the patterns of interactions between two neural populations. MCPA works by learning the mapping between the activity patterns from the populations from a training data set, and then classifying the neural communication pattern using these maps in a test data set. Simulated data demonstrated that MCPA was sensitive to information represented in neural interaction for realistic SNR ranges. Furthermore, MCPA is only sensitive to the discriminant information represented through different patterns of interactions irrespective of the information encoded in the local populations. Finally, we applied this method to fMRI data, and provide novel neuroscientific insights: that the multivariate connectivity pattern between areas along the visual stream represent information about individual natural images and that, at the category level, the representational structure of the interaction between regions is negatively correlated to the structure within each region.

It is worth noting that significant discrimination within each population and significant functional connectivity between them is not sufficient to produce MCPA and

indeed local classification within each population is not even necessary (Figures 2d and 2a respectively). MCPA requires the pattern of connectivity between the two populations to vary across the different conditions. As an example, if the two populations interact, but the interaction behaves like a passive filter, mapping the activity between the populations in a similar way in all conditions, MCPA would not be sensitive to the interaction because the mapping does not change (Figure 2d). Instead, MCPA is more akin to testing for adaptive filtering or distributed, interactive computation where the nature of the interaction changes depending on the information that is being processed. Recent studies demonstrate that neural populations in perceptual areas alter their response properties based on context, task demands, etc. (Gilbert & Li, 2013). These modulations of response properties suggest that lateral and long-distance interactions are adaptive and dynamic processes responsive to the type of information being processed and MCPA provides a platform for examining the role of interregional connectivity patterns in this adaptive process. Indeed, MCPA can be interpreted as testing whether distributed computational "work" is being done in the interaction between the two populations (K. J. Friston et al., 1997) and the interaction does not just reflect a passive relay of information between two encapsulated modules (Fodor, 1983).

In addition to allowing one to infer whether distributed computational work is being done in service of information processing, MCPA provides a platform for assessing its representational structure. Much as MVPA has been used in representational similarity analyses to measure the structure of the representational space at the level local neural populations (Edelman et al., 1998; Kriegeskorte, 2011; Kriegeskorte & Kievit, 2013), MCPA can be used to measure the structure of the representational space at the level of

network interactions. Specifically, the representational geometry of the interaction can be mapped in terms of the similarity among the multivariate functional connectivity patterns corresponding to the brain states associated with varying input information. It is notable that this type of single-trial or single-stimulus representational similarity analysis is not possible by directly applying a classifier to functional connectivity features and requires learning the mapping between neural activity patterns, as in MCPA (see next paragraph for further discussion of this distinction). The representational structure can be compared to behavioral measures of the structure to make brain-behavior inferences and assess what aspects of behavior a neural interaction contributes to. It can also be compared to models of the structure to test theoretical hypotheses regarding the computational role of the neural interaction (Kriegeskorte, 2011; Kriegeskorte & Kievit, 2013).

These two properties of MCPA, 1) being able to assess distributed computational processing rather than just whether or not areas are communicating and 2) being able to determine the representational structure of the information being processed, set MCPA apart from previously proposed functional connectivity methods. In these previous methods the functional connectivity calculation is performed separately from the classification calculation. Specifically, either functional connectivity is first calculated using standard means, then a model is built on the population of connectivity values and this model is tested using classification approaches (Finn et al., 2015; Richiardi, Eryilmaz, Schwartz, Vuilleumier, & Van De Ville, 2011; Rosenberg et al., 2016; Shirer, Ryali, Rykhlevskaia, Menon, & Greicius, 2012; Wang, Cohen, Li, & Turk-Browne, 2015) or the model is first built on the activity in each region and tested using classification approaches and the classification performance is correlated (Coutanche & Thompson-

Schill, 2013; Kriegeskorte & Kievit, 2013). These methods are very useful for assessing how differences large-scale patterns of connectivity relate to individual subject characteristics (e.g. connectome fingerprinting) in the first case and comparing the representational structure between regions in the second case. In contrast, in MCPA the model is the connectivity map and classification is done to directly test the information contained in these maps. The separation of the connectivity and classification calculations in other approaches precludes being able to assess distributed computational processes because these methods are sensitive to passive information exchange between encapsulated modules, as described above, and thus conflate passive and active information exchange. Critically, this separation also does not allow for single trial or single sample classification, as is required to perform the representational similarity analysis in a practical manner and decode how the information processed in the interaction is encoded and organized. As a concrete example, these previous methods would not be able to compare the representational structure of the interaction between regions to the structure seen locally within each region, as was done here with fMRI.

MCPA can be seen as a multivariate extension of PPI with an inherent classification framework. Compared to PPI, which is univariate, MCPA allows one to exploit the multivariate space of interaction patterns. As a result, MCPA is sensitive to aspects of information coded in interregional interactions that PPI cannot detect, for example in event-related fMRI designs where PPI is known to lack statistical power (O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012). In this sense, much the way MVPA allows one to go beyond ANOVAs/t-tests in a single area/population (e.g.

single trial classification, RSA, complex model testing), MCPA allows one to go beyond PPI and do these types of analyses at the level of the shared activity between regions.

Two examples of the types of studies that can be performed using MCPA help highlight the potential utility of this method. First, MCPA can be used to provide a strong test of the binding-by-synchrony hypothesis (Singer, 1999). This hypothesis asserts that information that results from computations arising from spatially distinct regions of the brain combine their information into a coherent representation through interregional synchronization. Thus far this hypothesis has primarily been tested by demonstrating increased interregional synchrony for conditions that show greater binding, such as showing greater synchrony when an image is perceived as a coherent Gestalt (Rodriguez et al., 1999). However, increased synchrony could reflect a number of different effects, not all of which necessarily imply binding (Reynolds & Desimone, 1999; Shadlen & Movshon, 1999). MCPA could provide a stronger test of the binding-by-synchrony hypothesis by allowing one to decode the representational content of the interaction and determine if it is consistent with a combination of what is represented in each area, as predicted by the hypothesis. Along similar lines, MCPA could be used to assess whether specific multivariate connectivity patterns bind together particular memory traces to examine hypotheses regarding memory reinstatement (Hermans et al., 2016).

A second potential use for MCPA is that a MCPA-based RSA may help inform models of how representations are transformed between neural populations along a processing pathway (Haxby et al., 2014). An increasingly successful approach for evaluating and developing computational models of neural processing is by assessing the similarity between the representations implied by the models to the one measured in the

brain (Kriegeskorte, Mur, & Bandettini, 2008), e.g. MVPA-based RSA. By comparing the fits of models to the neural representation, one can assess how well these models approximate the neural representation in both absolute and relative terms. Multi-layer computational models that approximate multiple, interacting neural regions (Riesenhuber & Poggio, 1999), as well as more data-driven models, such as deep neural network models (Yamins et al., 2014), have transfer functions that project information from one region to another. Much the way MVPA-based RSA analyses have been used to examine these models at the level of individual brain regions (Kriegeskorte et al., 2008), RSA analyses could be used to assess how well the representation inferred by these models' transfer functions fit the representation measured in the brain using MCPA. As an example, neural networks of visual processing have layers that are thought to approximate layers of the visual processing stream (Riesenhuber & Poggio, 1999; Yamins et al., 2014). Information is transferred between these regions through different variations of pooling, normalization, and convolution, or, alternatively, prediction and verification signals (Lee & Mumford, 2003). In the space of the image being processed, these operations have different implications for the similarity structure of the interaction between regions across a set of images. These similarity structures could be compared to the one found using MCPA to determine how well these models approximate the neural transfer function, in both an absolute and relative sense.

The specific instantiation of MCPA presented here treats connectivity as a bi-directional linear mapping between two populations. However, the MCPA framework could be easily generalized into more complicated cases. For example, instead of using correlation-based methods like CCA, other directed functional connectivity algorithms,

such as Granger causality based on an autoregressive framework, could be used to examine directional interactions. This would allow one to examine time-lagged multivariate connectivity patterns to infer directionality or alternatively exploit the asymmetric nature of regression to infer directionality. Additionally, kernel methods, such as kernel CCA, could be applied to account for non-linear interactions. A more general framework would be to use non-parametric functional regression method to build a functional mapping between the two multidimensional spaces in the two populations. MCPA can also be expanded to look at network-level representation by implementing the multiset canonical correlation analysis, wherein the cross-correlation among multiple sets of activity patterns from different brain areas is calculated (Kettenri.Jr, 1971). MCPA could be used with a dual searchlight approach to examine whole brain communication (Kriegeskorte, Goebel, & Bandettini, 2006). Also, MCPA could be adapted by optimizing the CCA to find the connectivity maps that uniquely describe, or at least best separate, the conditions of interest. Furthermore, both with and without these modification, the framework of MCPA may have a number of applications outside of assessing the representational content of functional interactions in the brain, such as detecting the presence of distributed processing on a computer network, or examining genetic or proteomic interactions. MCPA is used here with fMRI BOLD signals, but it can be applied to nearly any neural recording modality, including scalp or intracranial EEG, MEG, multiunit firing patterns, single unit firing patterns, spike-field coherence patterns, to assess the information processed by cross-frequency coupling, etc.

The MCPA results from visual cortex show that interactions between regions of the visual cortex are sensitive to the information contained in individual natural images.

The RSA analysis suggests an inverse relationship between the information processed at the local and distributed levels. This inverse relationship is consistent with Bayesian models that suggest that visual processing occurs through iterative prediction-verification processing(Lee & Mumford, 2003). In some implementations of this class of models, interactions between regions are thought to code for prediction errors (K. Friston, 2010). The negative correlation between local, MVPA-based representation and the distributed, MCPA-based representation suggests that information that cannot be resolved locally is represented in the interaction between regions. With the strong caveat that these results require replication in more subjects and assessment with paradigms designed to directly test these hypotheses, this negative correlation is consistent with the hypothesis that neural interactions code for information not resolved in local computational processes (e.g. prediction-verification errors). More broadly, the MCPA results suggest that the computational work done in service of visual processing occurs not only on the local level, but also at the level of distributed brain circuits.

Conclusion

Previously, multivariate pattern analysis have been used to analyze either the information processing within a certain area and functional connectivity methods have been used to assess whether or not brain networks participate in a particular process. With MCPA, the two perspectives are merged into one algorithm, which extends multivariate pattern analysis to enable the detailed examination information processing at the network level. Thus, the introduction of MCPA provides a platform for examining how computation is

carried out through the interactions between different brain areas, allowing us to directly test hypotheses regarding circuit-level information processing.

Materials and methods

Overview

The MCPA method consists of learning phase and a test phase (as in machine learning, where a model is first learned, then tested). In the learning phase, the connectivity maps for each condition that characterize the pattern of shared activity between two populations is learned. In the test phase, these maps are used to generate predictions of the activity in one population based on the activity in the other population as a factor of condition and these predictions are tested against the true activity in the two populations. Similar to linear regression where one can generate a prediction for the single variable A given the single variable B based on the line that correlates A and B, MCPA employs a canonical correlation model (a generalization of multivariate linear regression) and produces a mapping model for each condition as a hyperplane that correlates multidimensional spaces A and B. Thus one can generate a prediction of the observation in multivariate space A given the observation in multivariate space B on a single trials basis. In this sense, MCPA is more analogous to a machine learning classifier combined with a multivariate extension of PPI (K. J. Friston et al., 1997) rather than being analogous to correlation-based functional connectivity measures.

The general framework of MCPA is to learn the connectivity map between the populations for each task or stimulus condition separately based on training data. Specifically, given two neural populations (referred to as A and B), the neural activity of the two populations can be represented by feature vectors in multi-dimensional spaces (Haxby et al., 2014). The actual physical meaning of the vectors would vary depending

on modality, for example spike counts for a population of single unit recordings; time point features for event-related potentials (ERP) or event-related fields; time-frequency features for electroencephalography, electrocorticography (ECoG) or magnetoencephalography; or single voxel blood-oxygen-level dependent (BOLD) responses for functional magnetic resonance imaging. A mapping between A and B is calculated based on any shared information between them for each condition on the training subset of the data. This mapping can be any kind of linear transformation, such as any combination of projections, scalings, rotations, reflections, shears, or squeezes.

These mappings are then tested as to their sensitivity to the differential information being processed between cognitive conditions by determining if the neural activity can be classified based on the mappings. Specifically, for each new test data trial, the maps are used to predict the neural activity in one area based on the activity in the other area and these predictions are compared to the true condition of the data. The trained information-mapping model that fits the data better is selected and the trial is classified into the corresponding condition. This allows one to test whether the mappings were sensitive to the differential information being represented in the neural interaction in the two conditions.

Connectivity Map

The first phase of MCPA is to build the connectivity map between populations. The neural signal in each population can be decomposed into two parts: the part that encodes shared information, and the part that encodes non-shared local information (including any measurement noise). We assume that the parts of the neural activities that represent the shared information in the two populations are linearly correlated (though,

this can easily be extended by the introduction of a non-linear kernel). The model can be described as follows

$$\boldsymbol{C} \sim \mathcal{N}(0, \boldsymbol{I}_d), \min\{m_A, m_B\} \geq d \geq 1$$

$$\boldsymbol{A}|\boldsymbol{C} = \boldsymbol{W}_A\boldsymbol{C} + \boldsymbol{D}, \boldsymbol{D} \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Psi}_A), \boldsymbol{W}_A \in \mathbb{R}^{m_A \times d}, \boldsymbol{\Psi}_A \succcurlyeq 0$$

$$\boldsymbol{B}|\boldsymbol{C} = \boldsymbol{W}_B\boldsymbol{C} + \boldsymbol{E}, \boldsymbol{E} \sim \mathcal{N}(\boldsymbol{\mu}_B, \boldsymbol{\Psi}_B), \boldsymbol{W}_B \in \mathbb{R}^{m_B \times d}, \boldsymbol{\Psi}_B \succcurlyeq 0$$

where $\boldsymbol{C}$ is the common activity, $\boldsymbol{D}$ and $\boldsymbol{E}$ are local activities, m_A, m_B are the dimensionalities of activity vector in population A and B respectively. Without loss of generality, $\boldsymbol{\mu}_A = \boldsymbol{\mu}_B = 0$ can be assumed. The activity in population A can be decomposed into shared activity $\boldsymbol{W}_A\boldsymbol{C}$ and local activity $\boldsymbol{D}$, while activity in B can be decomposed into shared activity $\boldsymbol{W}_B\boldsymbol{C}$ and local activity $\boldsymbol{E}$. The shared discriminant information only lies in the mapping matrix $\boldsymbol{W}_A$ and $\boldsymbol{W}_B$ since $\boldsymbol{C}$ always follows the standard multivariate normal distribution (though correlation measures that do not assume normally distributed data can also be applied with minor modifications to the calculation).

In statistics, canonical correlation analysis (CCA) is optimally designed for such a model and estimate the linear mappings (Bach, 2005; Borga, 1998). In brief, let *S* be the covariance matrix

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{AA} & \boldsymbol{S}_{AB} \\ \boldsymbol{S}_{BA} & \boldsymbol{S}_{BB} \end{bmatrix} = \mathbb{E}\left[\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \quad \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix}^T\right]$$

Therefore $\boldsymbol{W}_A$ and $\boldsymbol{W}_B$ can be estimated by solving the following eigen problem

$$\begin{cases} \boldsymbol{S}_{AA}^{-1}\boldsymbol{S}_{AB}\boldsymbol{S}_{BB}^{-1}\boldsymbol{S}_{BA}\widehat{\boldsymbol{U}}_A = \rho^2\widehat{\boldsymbol{U}}_A \\ \boldsymbol{S}_{BB}^{-1}\boldsymbol{S}_{BA}\boldsymbol{S}_{AA}^{-1}\boldsymbol{S}_{AB}\widehat{\boldsymbol{U}}_B = \rho^2\widehat{\boldsymbol{U}}_B \end{cases}$$

and we have

$$\begin{cases} \widehat{\boldsymbol{W}}_A = \boldsymbol{S}_{AA}\widehat{\boldsymbol{U}}_{Ad}M_1 \\ \widehat{\boldsymbol{W}}_B = \boldsymbol{S}_{BB}\widehat{\boldsymbol{U}}_{Bd}M_2 \end{cases}$$

where $\widehat{\boldsymbol{U}}_{Ad}$ and $\widehat{\boldsymbol{U}}_{Bd}$ are the first *d* columns of canonical directions $\widehat{\boldsymbol{U}}_A$ and $\widehat{\boldsymbol{U}}_B$, and M_1, $M_2 \in \mathbb{R}^{d\times d}$ are arbitrary matrices such that $\boldsymbol{M}_1\boldsymbol{M}_2^T = \boldsymbol{P}_d$, $\boldsymbol{P}_d$ is the diagonal matrix with the first *d* elements of $\boldsymbol{P} = \boldsymbol{U}_B^T\boldsymbol{S}_{BA}\boldsymbol{U}_A$.

With $\boldsymbol{W}_A$ and $\boldsymbol{W}_B$, the shared information $\boldsymbol{C}$ can be estimated using its posterior mean $\mathbb{E}(\boldsymbol{C}|\boldsymbol{A})$ and $\mathbb{E}(\boldsymbol{C}|\boldsymbol{B})$, where $\mathbb{E}(\boldsymbol{C}|\boldsymbol{A}) = \boldsymbol{M}_1^T\boldsymbol{U}_A^T\boldsymbol{A}$ and $\mathbb{E}(\boldsymbol{C}|\boldsymbol{B}) = \boldsymbol{M}_2^T\boldsymbol{U}_B^T\boldsymbol{B}$. Let $\boldsymbol{M}_1 = \boldsymbol{M}_2$ and equate $\mathbb{E}(\boldsymbol{C}|\boldsymbol{A})$ and $\mathbb{E}(\boldsymbol{C}|\boldsymbol{B})$, this shared information can be used as a relay to build the bidirectional mapping between A and B. Specifically,

$\widehat{\boldsymbol{B}} = (\boldsymbol{M}_2^T\boldsymbol{U}_B^T)^\dagger\boldsymbol{M}_1^T\boldsymbol{U}_A^T\boldsymbol{A} = \boldsymbol{U}_B^{T\dagger}\boldsymbol{U}_A^T\boldsymbol{A} = \boldsymbol{R}\boldsymbol{A}$ and $\widehat{\boldsymbol{A}} = (\boldsymbol{M}_1^T\boldsymbol{U}_A^T)^\dagger\boldsymbol{M}_2^T\boldsymbol{U}_B^T\boldsymbol{B} = \boldsymbol{U}_A^{T\dagger}\boldsymbol{U}_B^T\boldsymbol{B} = \boldsymbol{R}^\dagger\boldsymbol{B}$, where $\boldsymbol{R} = \boldsymbol{U}_B^{T\dagger}\boldsymbol{U}_A^T\boldsymbol{A}$.

In the first step, the connectivity map is estimated for each condition separately. If we have n_1 trials in condition 1 and n_2 trials in condition 2 in the training set, the training data for the two conditions are represented in matrices as $\left[\boldsymbol{X}_A^{(1)}, \boldsymbol{X}_B^{(1)}\right]^T$ and $\left[\boldsymbol{X}_A^{(2)}, \boldsymbol{X}_B^{(2)}\right]^T$ respectively, where $\boldsymbol{X}_A^{(1)} \in \mathbb{R}^{m_A\times n_1}$, $\boldsymbol{X}_B^{(1)} \in \mathbb{R}^{m_B\times n_1}$ are the population activity for A and B under condition 1 respectively, and $\boldsymbol{X}_A^{(2)} \in \mathbb{R}^{m_A\times n_2}$, $\boldsymbol{X}_B^{(2)} \in \mathbb{R}^{m_B\times n_2}$ are the population activity for A and B under condition 2 respectively. The testing data vector is then represented as $[\boldsymbol{x}_A, \boldsymbol{x}_B]^T$, where $\boldsymbol{x}_{\square} \in \mathbb{R}^{m_A}$ and $\boldsymbol{x}_B \in \mathbb{R}^{m_B}$ are population activities in A and B respectively. Using CCA, the estimations of the mapping matrices with respect to different conditions are $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$.

To sum up, by building the connectivity map, a linear mapping function $\boldsymbol{R}$ is estimated from the data for each condition so that the activity of the two populations can

be directly linked through bidirectional functional connectivity that captures only the shared information.

Classification

The second phase of MCPA is a pattern classifier that takes in the activity from one population and predicts the activity in a second population based on the learned connectivity maps conditioned upon the stimulus condition or cognitive state. The testing data is classified into the condition to which the corresponding model most accurately predicts the true activity in the second population.

The activity from one population is projected to another using the learned CCA model, i.e. $\boldsymbol{x}_B^{(i)} = \boldsymbol{U}_B^{(i)\dagger}\boldsymbol{U}_A^{(i)}\boldsymbol{x}_{\boldsymbol{A}}$. The predicted projections $\boldsymbol{x}_B^{(i)}$ are compared to the real observation $\boldsymbol{x}_{\boldsymbol{B}}$, and then the testing trial is labeled to the condition where the predicted and real data match most closely. Any similarity metric could be used for this comparison; here cosine similarity (correlation) is used. The mapping is bidirectional, so A can be projected to B and vice versa. In practice, the similarities from the two directions are averaged in order to find the condition that gives maximum average correlation coefficient.

Simulated data

To test the performance of MCPA, we simulated shared and local activity in two populations and tested the performance of MCPA on synthetic data as a factor of the number of dimensions in each population and signal-to-noise ratio (SNR; Figure 2a). In addition to the MVPA control described above, we further evaluated the following three

control experiments to demonstrate that MCPA is insensitive to the presence or change in the local information. In the first control experiment (no functional connectivity, no shared information, varying local information), we simulated the case where two populations are totally independent under both conditions, but there is local discriminant information in each (Figure 2b). In the second control experiment (functional connectivity, constant shared information, varying local information), we introduced local discriminant information into population A without changing the amount of shared information between populations A and B (Figure 2c). In the third control experiment (functional connectivity, no shared information, varying local information), we eliminated the information represented in the pattern of interaction, but maintained the functional connectivity by keeping the correlation between populations invariant with regard to conditions.

For the first simulation (Figure 2a), shared activity for both conditions in population A was drawn independently from a *d*-dimensional normal distribution $\boldsymbol{Y}_A^{(i)} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_d)$, for $i = 1,2$. The shared activity in population B under two different conditions were generated by rotating $\boldsymbol{Y}_A$ with different rotation matrices separately, $\boldsymbol{Y}_B^{(i)} = \boldsymbol{R}^{(i)} \boldsymbol{Y}_A^{(i)}$, where $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ were two *d*-by-*d* random rotation matrices corresponding to the information mapping under condition 1 and 2 respectively, and $\boldsymbol{R}^{(i)^{\boldsymbol{T}}} \boldsymbol{R}^{(i)} = \boldsymbol{I}_d$.

The two important parameters here are the dimensionality *d* and the variance σ^2. SNR was used to characterize the ratio between the variance of shared activity and variance of local activity, and the logarithmic decibel scale $\mathrm{SNR}_{dB} = 10 \log_{10}(\sigma^2)$ was used. To cover the wide range of possible data recorded from different brain regions and

different measurement modalities, we tested the performance of MCPA with *d* ranging from 2 to 25 and SNR ranging from -20 dB to 20 dB (σ^2 ranged from 0.01 to 100). Note that each of the *d* dimensions contain independent information about the conditions though have the same SNR. Thus the overall SNR does not change, but the amount of pooled information does change with *d*. For each particular setup of parameters, the rotation matrices $\boldsymbol{R}^{(i)}$ were randomly generated first, then 200 trials were randomly sampled for each condition and evenly split into training set and testing set. MCPA was trained using the training set and tested on the testing set to estimate the corresponding true positive rate (TPR) and false positive rate (FPR) for the binary classification. The sensitivity index *d'* was then calculated as $d' = Z(TPR) - Z(FPR)$, where $Z(x)$ is the inverse function of the cdf of standard normal distribution. This process was repeated 100 times and the mean and standard errors across these 100 simulations were calculated. Note that the only discriminant information about the two conditions is the pattern of interactions between the two populations, and neither of the two populations contains local discriminant information about the two conditions in its own activity. We further tested and confirmed this by trying to classify the local activity in populations A and B (see below). To avoid an infinity *d'* value, with 100 testing trials, the maximum and minimum for TPR or FRP were set to be 0.99 and 0.01, which made the maximum possible *d'* to be 4.65.

The MCPA method captures the pattern of correlation between neural activities from populations and is invariant to the discriminant information encoded in local covariance. To see this, we first take the simulation data described above and apply MVPA (naïve Bayes) to each of the two populations separately. Note that in each of the

two populations, we set the two conditions to have the same mean and covariance. As a result there should be no local discriminant information within any of the two populations alone.

Control simulations

For the first control simulation (Figure 2b), for condition 1, $\boldsymbol{X}_A^{(1)}, \boldsymbol{X}_B^{(1)}$ were drawn independently from the same distribution $\mathcal{N}(0, \boldsymbol{I}_d)$; for condition 2, $\boldsymbol{X}_A^{(2)}, \boldsymbol{X}_B^{(2)}$ were drawn independently from the same distribution $\mathcal{N}(0, \boldsymbol{I}_d)$. Then we changed the local variance in one of the conditions. For the features in population A and B under condition 1, we used $\boldsymbol{X}_A^{(1)'} = k\boldsymbol{X}_A^{(1)}$ and $\boldsymbol{X}_B^{(1)'} = k\boldsymbol{X}_B^{(1)}$, where *k* ranged from 1 to 9. Thus, in both populations, the variance of condition 1 was different from the variance of condition 2, and such difference would increase as *k* became larger. Therefore, there was no information shared between the two populations under either condition, but each of the population had discriminant information about the conditions encoded in the variance for any $k \neq 1$.

For the second control simulation (Figure 2c), we fixed the dimensionality at 10 and SNR at 0 dB ($\sigma^2 = 1$) and kept the rotation matrices of different conditions different from each other. As a result, the amount of shared discriminant information represented in the patterns of interactions stayed constant. Then we changed the local variance in one of the conditions. For the features in population A under condition 1, we used $\boldsymbol{X}_A^{(1)'} = k\boldsymbol{X}_A^{(1)}$, where *k* ranged from 1 to 9. Thus, population A, the variance of condition 1 was different from the variance of condition 2, and such difference would increase as *k* became larger. According to our construction of MCPA, it should only pick up the

discriminant information contained in the interactions and should be insensitive to the changes in local discriminant information from any of the two populations.

For the third control simulation (Figure 2d), we introduced local discriminant information into the two populations to demonstrate that MCPA is insensitive to the presence of constantly correlated local information (Figure 2d). We fixed the dimensionality at 10 and SNR at 0 dB ($\sigma^2 = 1$) and kept the rotation matrices constant for different conditions. As a result, the amount of shared discriminant information represented in the patterns of interactions was 0. Then we changed the local variance in one of the conditions. Then we changed the local variance in one of the conditions. For the features in population A and B under condition 1, we used $\boldsymbol{X}_A^{(1)'} = k\boldsymbol{X}_A^{(1)}$ and $\boldsymbol{X}_B^{(1)'} = k\boldsymbol{X}_B^{(1)}$, where *k* ranged from 1 to 9. Thus, in both populations, the variance of condition 1 was different from the variance of condition 2, and such difference would increase as *k* became larger. Notably, such local information was actually correlated through interactions between the populations. However, since the pattern of interaction did not vary as the condition changed, there was no discriminant information about the conditions represented in the interactions. According to our construction of MCPA, it should not pick up any discriminant information in this control case.

Examining visual cortex coding for natural images using MCPA

The fMRI dataset was taken from CRCNS.org (Kay et al., 2011). See (Kay et al., 2008; Naselaris et al., 2009) for details regarding subjects, stimuli, MRI parameters, data collection, and data preprocessing. In the experiment, two subjects performed passive natural image viewing tasks while BOLD signals were recorded from the brain. The

experiment contains two stages: a training stage and a validation stage. In the training stage, two separate trials were recorded in each subject. In each trial, a total of 1750 images were presented to the subject, which yields a total of 3500 presentations of images (3500 = 1750 images * 2 repeats). In the validation stage, another 120 images were presented to the subject in 13 repeated trials, which yields a total of 1560 presentations (1560 = 120 images * 13 repeats). The single-trial beta weights for each voxel were estimated and used for the following analysis. The voxels were assigned to 5 visual areas (V1, V2, V3, V4, and LO) based on retinotopic mapping data from separate scans (Kay et al., 2008; Naselaris et al., 2009).

MCPA analysis

Categorical image classification

To control for repetition of each individual image and increase the image number being used, we used the data from the training stage for the categorical image classification. The 1750 images were manually sorted into 8 categories (animals, buildings, humans, natural scenes, textures, food, indoor scenes, and manmade objects). In order to maintain enough statistical power, only categories with more than 100 images were used in the analysis. As a result, 3 categories (food, indoor scenes, and manmade objects) were excluded.

For each pair of ROIs, namely V1-V2, V2-V3, V3-V4, and V4-LO, MCPA was applied to classify the functional connectivity patterns for each possible pair of image categories (total of 10 pairs). For each specific pair of categories, BOLD signal from all the voxels in the ROIs were used as features in MCPA. Principal Component Analysis

(PCA) was used to reduce the dimensionality to *P*, where *P* corresponds to the number of PCs that capture 90% of variation in the data, which yielded between 100-200 *Ps*. Leave-one-trial-out cross-validation was used in order to estimate the classification accuracy. This procedure was repeated for all 10 pairs. *d'* was used to quantify the performance of MCPA.

Single image classification

For single image classification the 13 repetitions of each individual image from the validation stage data was used.

For each pair of ROIs, namely V1-V2, V2-V3, V3-V4, and V4-LO, MCPA was applied to classify the functional connectivity patterns for each possible pair of images (total of 7140 pairs). For each specific pair of categories, BOLD signal from all the voxels in the ROIs were used as features in MCPA. Considering the limited number of trials in each condition, PCA was first used with the data from the training stage to reduce the representation dimensionality to 10. Because the top PCs that explain most variations may contain variance not related to the stimuli, the 10 PCs were selected from the top 50 PCs, based on maximizing the between-trial correlations for single simages. As a result, we reduce the dimensionality of the validation data from more than 1000 to 10 based on the training dataset. Leave-one-trial-out cross-validation was then used in order to estimate the classification accuracy. This procedure was repeated for all 7140 pairs. *d'* was used to quantify the performance of MCPA.

MVPA analysis

MVPA was applied to classify the neural activity within each ROI (V1, V2, V3, V4, and LO) for each possible pair of categories (total of 10 pairs). The same features extracted from all the voxels within the ROI, as described above, were used in MVPA analysis. Naïve Bayes classifier was used as the linear classifier and leave-one-trial-out cross-validation was used in order to estimate the classification accuracy. This procedure was repeated for all 10 pairs. *d'* was used to quantify the performance of MVPA.

Permutation test

Permutation testing was used to determine the significance of the classification accuracy *d'*. For each permutation, the condition labels of all the trials were randomly permuted and the same procedure as described above was used to calculate the *d'* for each permutation. The permutation was repeated for a total of 200 times. The *d'* of each permutation was used as the test statistic and the null distribution of the test statistic was estimated using the histogram of the permutation test.

Representational similarity analysis

Similar to previous example, based on the classification results, for each classification analysis, the dissimilarity matrix $\boldsymbol{M}$ was constructed such that the *j*th element in the *i*th row m_{ij} equals the dissimilarity (classification accuracy) between the category *i* and category *j* in the corresponding representational space defined by the analysis. Pearson's correlation was used to compare representational dissimilarity matrices.

Acknowledgements

We would like to thank Kendrick Kay for sharing the fMRI dataset. We thank Marc Coutanche and Julie Fiez for their insightful comments and feedback on this work. This work was supported by the National Institute on Drug Abuse under award NIH R90DA023420 (to YL) and the National Institute of Mental Health under award NIH R01MH107797 (to ASG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Reference:

Bach, F. R. J., Michael I. (2005). *A probabilistic interpretation of canonical correlation analysis* (688). Technical Report 688, Department of Statistics, University of California, Berkeley.

Borga, M. (1998). *Learning multidimensional signal processing.* Linköping University, Linköping, Sweden.

Coutanche, M. N., & Thompson-Schill, S. L. (2013). Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. *Front Hum Neurosci, 7*, 15. doi:10.3389/fnhum.2013.00015

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage, 19*(2 Pt 1), 261-270.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci, 4*(8), 2051-2062.

Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology, 26*(4), 309-321.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., . . . Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci, 18*(11), 1664-1671. doi:10.1038/nn.4135

Fodor, J. A. (1983). *The modularity of mind : an essay on faculty psychology*. Cambridge, Mass.: MIT Press.

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nat Neurosci, 12*(9), 1187-1196. doi:10.1038/nn.2363

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci, 11*(2), 127-138. doi:10.1038/nrn2787

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B-Biological Sciences, 360*(1456), 815-836. doi:10.1098/rstb.2005.1622

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6*(3), 218-229. doi:10.1006/nimg.1997.0291

Ghuman, A. S., Brunet, N. M., Li, Y., Konecky, R. O., Pyles, J. A., Walls, S. A., . . . Richardson, R. M. (2014). Dynamic encoding of face information in the human fusiform gyrus. *Nat Commun, 5*, 5672. doi:10.1038/ncomms6672

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nat Rev Neurosci, 14*(5), 350-363. doi:10.1038/nrn3476

Grossberg, S. (1980). How Does a Brain Build a Cognitive Code. *Psychological Review, 87*(1), 1-51.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience, Vol 37, 37*, 435-456. doi:10.1146/annurev-neuro-062012-170325

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. doi:DOI 10.1126/science.1063736

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523-534. doi:10.1038/nrn1931

Hermans, E. J., Kanen, J. W., Tambini, A., Fernandez, G., Davachi, L., & Phelps, E. A. (2016). Persistence of Amygdala-Hippocampal Connectivity and Multi-Voxel Correlation Structures During Awake Rest After Fear Learning Predicts Long-Term Expression of Fear. *Cerebral Cortex*. doi:10.1093/cercor/bhw145

Hirshorn, E. A., Li, Y., Ward, M. J., Richardson, R. M., Fiez, J. A., & Ghuman, A. S. (2016). Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc Natl Acad Sci U S A, 113*(29), 8162-8167. doi:10.1073/pnas.1604126113

Hubel, D. H., & Wiesel, T. N. (1959). Receptive Fields of Single Neurones in the Cats Striate Cortex. *Journal of Physiology-London, 148*(3), 574-591.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat Neurosci, 8*(5), 679-685. doi:10.1038/nn1444

Kay, K. N., Naselaris, T., & Gallant, J. L. (2011). *fMRI of human visual areas in response to natural images*.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*(7185), 352-355. doi:10.1038/nature06713

Kettenri.Jr. (1971). Canonical Analysis of Several Sets of Variables. *Biometrika, 58*(3), 433-&. doi:Doi 10.2307/2334380

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *Neuroimage, 56*(2), 411-421. doi:10.1016/j.neuroimage.2011.01.061

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3863-3868. doi:10.1073/pnas.0600244103

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences, 17*(8), 401-412. doi:10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci, 2*, 4. doi:10.3389/neuro.06.004.2008

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain Cogn, 65*(2), 145-168. doi:10.1016/j.bandc.2007.06.007

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America a-Optics Image Science and Vision, 20*(7), 1434-1448. doi:Doi 10.1364/Josaa.20.001434

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron, 63*(6), 902-915. doi:10.1016/j.neuron.2009.09.006

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences of the United States of America, 108*(24), 9998-10003. doi:10.1073/pnas.1102433108

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424-430. doi:10.1016/j.tics.2006.07.005

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Soc Cogn Affect Neurosci, 7*(5), 604-609. doi:10.1093/scan/nss055

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron, 72*(5), 692-697. doi:10.1016/j.neuron.2011.11.001

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310*(5756), 1963-1966. doi:10.1126/science.1117645

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron, 24*(1), 19-29, 111-125.

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., & Van De Ville, D. (2011). Decoding brain states from fMRI connectivity graphs. *Neuroimage, 56*(2), 616-626. doi:10.1016/j.neuroimage.2010.05.081

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci, 2*(11), 1019-1025. doi:10.1038/14819

Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature, 397*(6718), 430-433. doi:10.1038/17120

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nat Neurosci, 19*(1), 165-171. doi:10.1038/nn.4179

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Representations by Back-Propagating Errors. *Nature, 323*(6088), 533-536. doi:DOI 10.1038/323533a0

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron, 24*(1), 67-77, 111-125.

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex, 22*(1), 158-165. doi:10.1093/cercor/bhr099

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron, 24*(1), 49-65, 111-125.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311*(5761), 670-674. doi:10.1126/science.1119983

Wang, Y. D., Cohen, J. D., Li, K., & Turk-Browne, N. B. (2015). Full correlation matrix analysis (FCMA): An unbiased method for task-related functional connectivity. *Journal of Neuroscience Methods, 251*, 108-119. doi:10.1016/j.jneumeth.2015.05.012

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci U S A, 111*(23), 8619-8624. doi:10.1073/pnas.1403112111

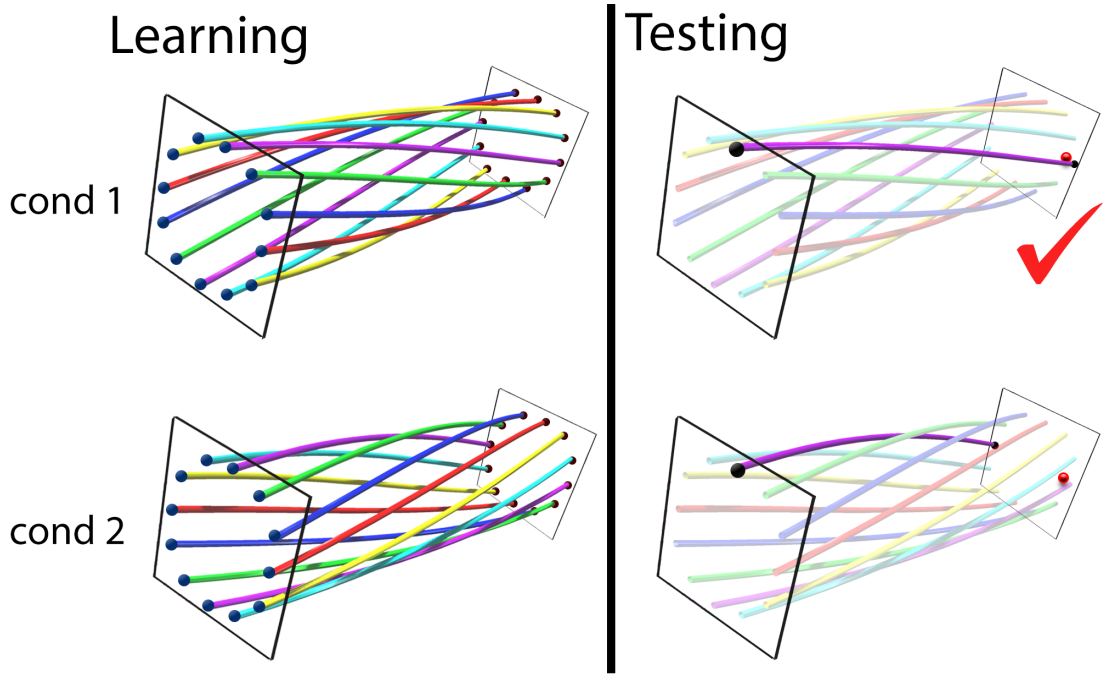

Figure 1. Illustration of the connectivity map and classifier of MCPA.

The MCPA framework is demonstrated as a two-phase process: learning and testing.

Top left: An illustration of the learned functional information mapping between two populations under condition 1. The representational state spaces of the two populations are shown as two planes and each pair of blue and red dots correspond to an observed data point from the populations. The functional information mapping is demonstrated as the colored pipes that project points from one space onto another (in this case, a 90 degree clockwise rotation).

Bottom left: An illustration of the learned functional information mapping between two populations under condition 2 (in this case, a 90 degree counterclockwise rotation).

Top right: An illustration of the predicted signal by mapping the observed neural activity from one population onto another using the mapping patterns learned from condition 1. The real signal in the second population is shown by the red dot.

Bottom right: An illustration of the predicted signal by mapping the observed neural activity from one population onto another using the mapping patterns learned from condition 2.

In this case, MCPA would classify the activity as arising from condition 1 because of the better match between the predicted and real signal.

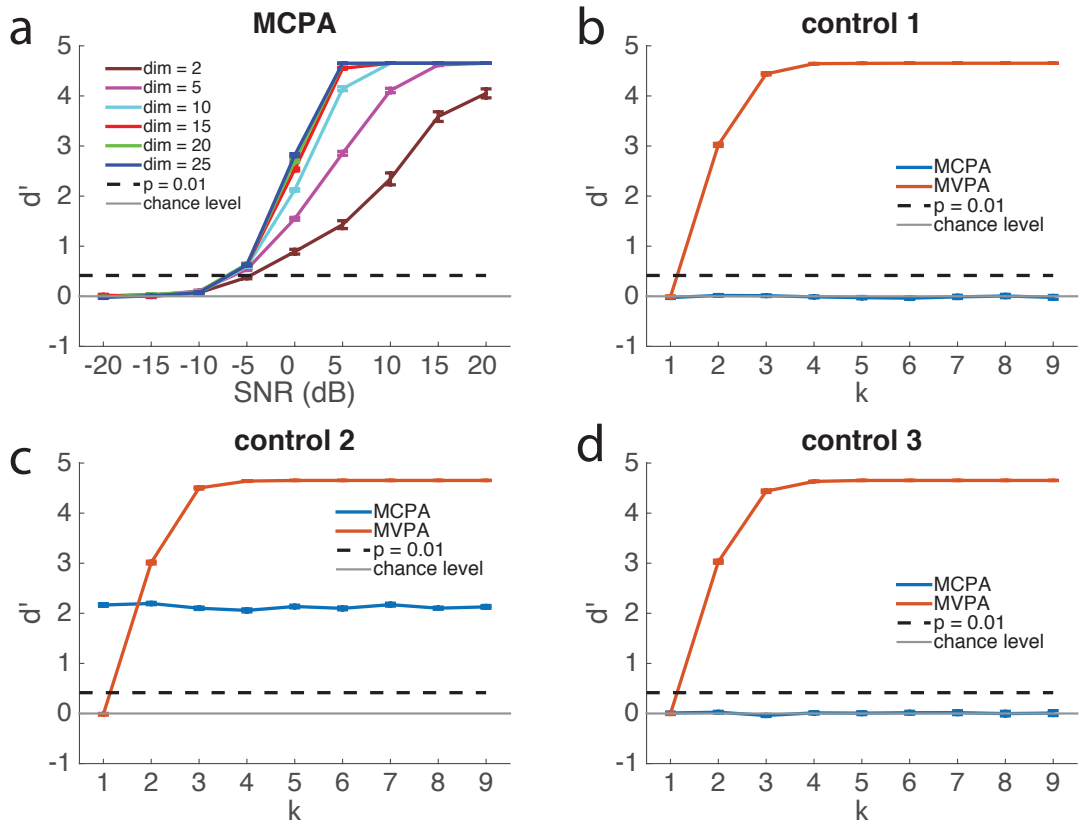

Figure 2. Synthetic data and control simulation experiments. The mean and standard error for 100 simulation runs are plotted. The horizontal gray line corresponds to chance level (*d*' = 0). The dashed line (*d*' = 0.42) corresponds to the chance threshold, p = 0.01, based on a permutation test. The maximum possible *d'* = 4.65 (equivalent to 99% accuracy because the d' for 100% accuracy is infinity).

a) The sensitivity of MCPA for connectivity between two populations as a factor of SNR and the number of effective dimensions in each population. MCPA was applied to synthetic data, where two conditions had different patterns of functional connectivity (measured by SNR and dimensionality). Performance of MCPA was significantly higher than chance level when SNR ≥ -5 dB and the number of dimensions ≥ 2. Performance of MCPA saturated to maximum when SNR > 5 dB and the number of dimensions > 10.

b) The insensitivity of MCPA when there is variable local discriminant information, but no circuit-level information (control case 1). MCPA and MVPA were applied to control case 1. The SNR was fixed at 0 dB and the number of dimensions is fixed at 10 for panels b, c, and d. *k* corresponds to the ratio of the standard deviations of the two conditions in panels b, c, and d.

c) The insensitivity of MCPA to changes in local discriminant information with fixed circuit-level information when there is both local and circuit-level information (control case 2).

d) The insensitivity of MCPA to variable local discriminant information when the circuit-level activity is correlated, but does not contain circuit-level information about what is being processed (control case 3).

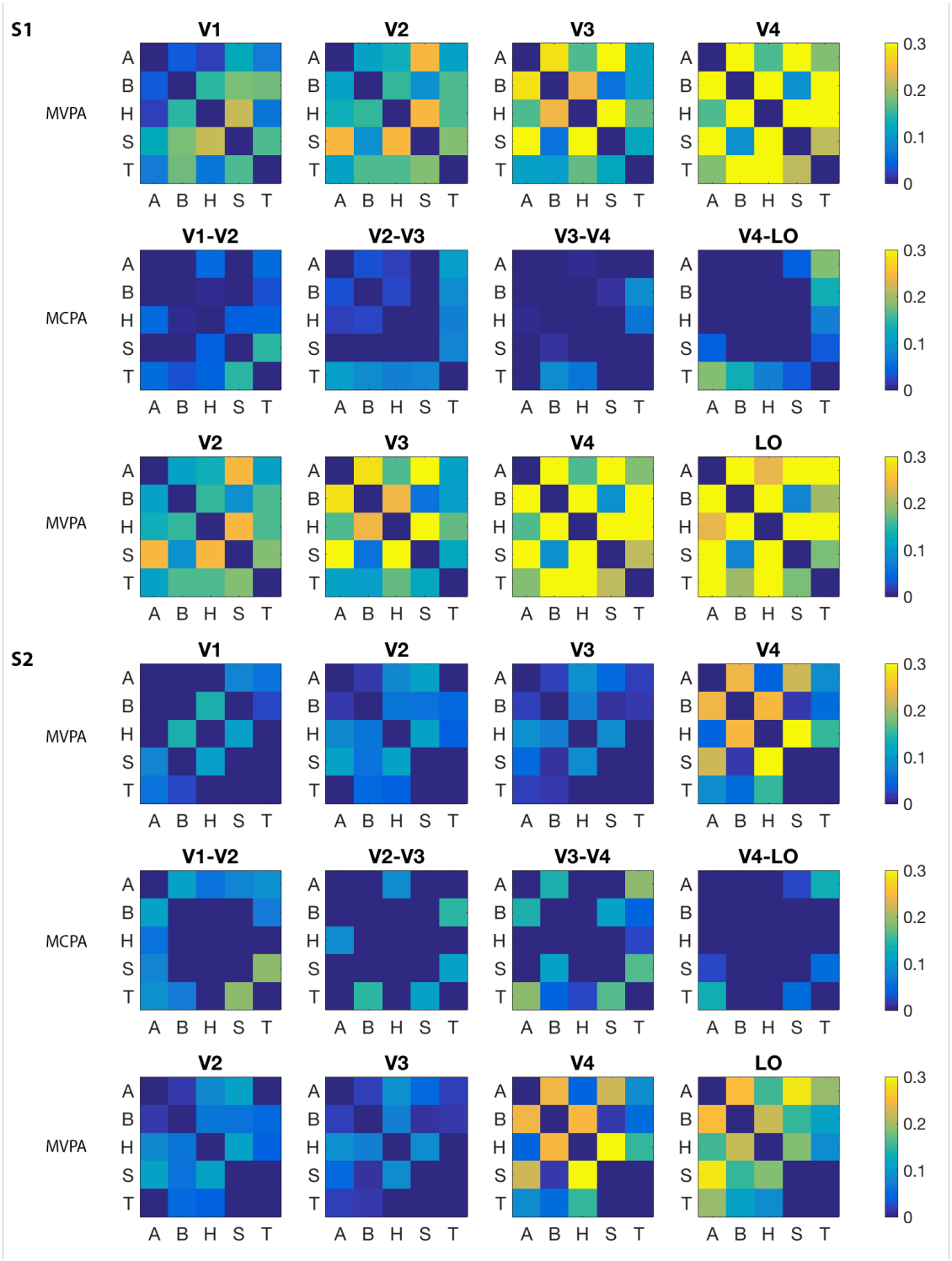

Figure 3. MCPA and MVPA results for fMRI data.

RSA results based on MCPA and MVPA for V1, V2, V3, V4, and LO from subjects S1 and S2. Categories: A-animals, B-buildings, H-humans, S-natural scenes, T-textures.

Row 1 & 3: RSA based on MVPA for V1, V2, V3, V4, and LO of S1, each entry represents the classification accuracy (*d*') between the corresponding categories;

Row 2: RSA based on MCPA for V1-V2, V2-V3, V3-V4, and V4-LO of S1, each entry represents the classification accuracy (*d*') between the corresponding categories;

Row 4 & 6: RSA based on MVPA for V1, V2, V3, V4, and LO of S2, each entry represents the classification accuracy (*d*') between the corresponding categories;

Row 5: RSA based on MCPA for V1-V2, V2-V3, V3-V4, and V4-LO of S2, each entry represents the classification accuracy (*d*') between the corresponding categories.

Table 1 Pearson's correlation coefficients between MCPA of ROI1-ROI2 and MVPA of ROI1 or ROI2

	S1				S2			
ROI1-ROI2	V1-V2	V2-V3	V3-V4	V4-LO	V1-V2	V2-V3	V3-V4	V4-LO
ROI1	0.251	-0.539	-0.768	-0.417	-0.420	-0.257	-0.770	-0.193
ROI2	0.110	-0.641	-0.555	-0.177	-0.575	-0.0720	-0.426	-0.131