Multivariate Pattern Connectivity

Stefano Anzellotti[1], Alfonso Caramazza[2,3], and Rebecca Saxe[1]

[1]Massachusetts Institute of Technology
[2]Harvard University
[3]Università degli Studi di Trento

March 28, 2016

Abstract

Whenever we engage in a cognitive task, multiple brain regions are engaged. Understanding how these regions interact is a fundamental step to uncover the neural mechanisms that make behavior possible. The majority of the investigations of interactions between brain regions have focused on the overall univariate responses in the regions. However, in the context of 'static' analyses, drastic advantages have derived from the application of multivariate techniques considering the fine-grained spatial structure of responses within each region (multivariate pattern analysis - MVPA). In the present article, we introduce and apply a technique to study connectivity in terms of the relations between multivariate patterns of responses within brain regions: multivariate pattern connectivity (MVPC). Considering the fusiform face area (FFA) as a seed region, we show that MVPC provides novel information about the interactions between regions that goes beyond univariate functional connectivity analyses.

1 Introduction

When we engage in a task - from recognizing the identity of a face to attributing mental states to others or understanding a sentence - multiple brain regions are engaged (Ishai [2008], Anzellotti and Caramazza [2015], Gallagher and Frith [2003], Fedorenko and Thompson-Schill [2014]). How do these regions work together to generate behavior? The investigation of large-scale interactions between multiple brain regions has attracted increasing interest in neuroscience, and a variety of methods have been developed to study connectivity both in terms of the anatomical structure of the brain (Le Bihan et al. [2001]), and of the relations between timecourses of responses during rest (Biswal et al. [1995]) and during specific experimental tasks (Friston et al. [2003], Roebroeck et al. [2005]). Functional Magnetic Resonance Imaging (fMRI) has proven to be a valuable instrument in this enterprise, offering noninvasive recording with good spatial resolution and whole-brain coverage.

In parallel with the development of methods for the study of connectivity, the fMRI literature has seen the introduction and diffusion of multivariate analysis techniques (multivariate pattern analysis or MVPA; Haxby et al. [2001]) that exploit the fine-grained spatial patterns of response within individual brain regions to achieve a better characterization of those regions' functional roles. MVPA has drastically increased the potential of fMRI for 'static' analyses of representational content, making it possible to detect information at a level of specificity that was unthinkable with previous univariate analyses (Kriegeskorte et al. [2007], Nestor et al. [2011], Anzellotti et al. [2013], Soon et al. [2008], Koster-Hale et al. [2013]). 'Static' analyses here are intended as analyses that average across time and do not exploit the temporal dimension of the data. Despite the success of MVPA for 'static' analyses of neural representations in individual brain regions, relatively few attempts have been made to transport the potential of multivariate analyses to the domain of dynamics and connectivity.

A recent study (Coutanche and Thompson-Schill [2014]) used successful vs unsuccessful classification of color and shape in area V4 and in the lateral occipital complex (LOC) to predict successful vs unsuccessful object classification in the anterior temporal lobe (ATL), providing evidence that classification accuracy in ATL in a given experimental block can be predicted from classification in V4 and the fusiform face area (FFA; Kanwisher et al. [1997]) in the same block. This study is an important first step towards exploiting multivariate voxel patterns to study interactions between brain regions, but it is limited by the use of a single discrete measure of a region's representational content (1 for successful classification, 0 for unsuccessful classification). An additional property of this method is that it uses classification along experimenter defined categories, which can be useful to probe a specific hypothesis about the type of interactions between a set of regions, but does not guarantee to capture properties of the stimuli and tasks that play a predominant role in driving the regions' responses. A large portion of the variance in the regions' responses might be determined by properties that are orthogonal to the categories determined by the experimenter.

Another innovative study (Henriksson et al. [2015]) investigated the relations between brain regions in terms of fine-grained patterns of response measuring correlation between representational dissimilarity matrices in different regions. Pairs of dissimilarity matrices calculated with the same set of trials showed higher correlations than pairs of dissimilarity matrices each of which was calculated with separate sets of data, showing that trial-specific fluctuations in the regions' responses can affect the information they encode in a manner that is trial-dependent but related across regions. This approach provides a richer characterization of each region's representational structure by comparing similarity matrices instead of successful vs unsuccessful classification. However, to make the dissimilarity matrices comparable across regions, the same set of conditions need to be used to generate the dissimilarity matrices in the different regions. This can be very effective when the conditions correspond to individual stimuli as in Henriksson et al. [2015], but if stimulus categories were used (e.g. faces vs.

animals vs. artifacts) it would raise the question of whether it is appropriate to characterize the representational spaces of brain regions as disparate as early visual cortex, inferior temporal cortex and the anterior temporal lobes in terms of the dissimilarities between the same set of categories. Adopting the same set of experimenter defined categories across all regions may be effective to test a specific hypothesis about the interactions between regions concerning a specific representational subspace, but it would be less suitable to evaluate the overall extent of the interaction between two regions.

Starting from these considerations, we developed a method (multivariate pattern connectivity - MVPC) to investigate the relations between brain regions in terms of fine-grained spatial patterns of responses. The method is composed of three main stages. In the first stage, the representational space in each brain region is modelled extracting a set of dimensions that correspond to spatial response patterns that 'best' characterize a region's responses over time. In the second stage, the multivariate timecourses of responses in each region are reparametrized as trajectories in the representational spaces defined by these dimensions. In the third stage, the relations between the trajectories in the representational spaces of different regions are modelled. In a procedure analogous to MVPA, independent data are used to train and test the models. Parameters modelling the relationship between two regions are estimated with all runs but one, and then used to model the relation between those regions in the remaining run.

In this article, we discuss the details of MVPC, and we compare the results obtained with MVPC to standard functional connectivity analyses using as a test case the connectivity of the fusiform face area and the rest of the brain (with a searchlight approach). The results indicate that MVPC provides complementary information about the connectivity of FFA and can enrich our understanding of the interactions between brain regions.

2 Multivariate Pattern Connectivity

Multivariate Pattern Connectivity (MVPC) is based on the idea of modelling the representational space in a brain region and reparametrizing the patterns of response in the region over time as trajectories in the representational space. The trajectories in the representational space in one region can then be used to predict trajectories in the representational space of other regions yielding a multivariate measure of connectivity.

2.1 Preprocessing

Let S_1 be a $n_1 \times T$ matrix and S_2 a $n_2 \times T$ matrix of Blood-Oxygen Level Dependent (BOLD) signal, where n_1 is the number of voxels in a region r_1, n_2 is the number of voxels in a region r_2 and T is the number of acquired functional volumes. Before modelling the representational spaces in regions r_1, r_2, we need to apply some preprocessing to the BOLD timecourses. Removal of

multiple sources of noise can be achieved with compcor (Behzadi et al. [2007]). A control region of interest (ROI) that does not contain gray matter such as the ventricles is defined, and principal components are extracted. Since the control ROI does not contain gray matter, its responses are thought to reflect noise. The timecourses of the components extracted from the control ROI are regressed out from the timecourses of every voxel in regions r_1, r_2 obtaining matrices of residuals E_1, E_2 of size $n_1 \times T$ and $n_2 \times T$ respectively. Preprocessing is applied to m functional runs yielding multivariate timecourses $E_1^1, \ldots, E_1^m$ and $E_2^1, \ldots, E_2^m$.

If the goal is to obtain a measure of multivariate connectivity that reflects interactions beyond what could be detected with functional connectivity, the mean response in the ROI needs to be subtracted from each voxel in the ROI at each timepoint. After this step, the average timecourse in the ROI is constantly zero and standard functional connectivity calculated after this stage would also be zero. Alternatively, if the aim is to investigate with the same measure the interactions attributable to both the mean timecourse and to finer-grained patterns, the mean can be kept. These alternatives are similar to the choice between using correlation distance vs. euclidean distance in MVPA.

2.2 Modelling Representational Spaces

Since we aim to generate a model of representational space that is not specific to an individual run but is common to the entire experiment, the multivariate timecourses $E_1^1, \ldots, E_1^m$ and $E_2^1, \ldots, E_2^m$ are first concatenated in time, yielding new timecourses $\tilde{E}_1$, $\tilde{E}_2$ of size $n_1 \times \tilde{T}$ and $n_2 \times \tilde{T}$ respectively, where $\tilde{T} = \sum_{i=1}^{m} T_i$ and T_i is the number of functional volumes acquired in run i. Singular value decomposition is then used to perform principal component analysis (PCA) in each region, obtaining:

$$\tilde{E}_1 = U_1 \Sigma_1 V_1^T \tag{1}$$

$$\tilde{E}_2 = U_2 \Sigma_2 V_2^T \tag{2}$$

The first k_1 and k_2 components are extracted, with $k_1 \leq n_1$ and $k_2 \leq n_2$. To obtain the new, dimensionality-reduced timecourses, it suffices to calculate

$$\tilde{E}_{1,red} = U_1 \Sigma_1 \tag{3}$$

$$\tilde{E}_{2,red} = U_2 \Sigma_2 \tag{4}$$

Many techniques can be used to decide how many components to select, for example the scree test (Cattell [1966], Zoski and Jurs [1996]) or cross validation methods (Josse and Husson [2012]). A recent study has shown that in the case of resting state, 5 components provide a good characterization of the data (Diez et al. [2015]), therefore in the applications in this article we use 5 components, leaving it for future work to test the effectiveness of different methods to determine the number of components. This procedure leaves us with multivariate timecourses $\tilde{E}_{1,red}$, $\tilde{E}_{2,red}$ of size $k_1 \times \tilde{T}$ and $k_2 \times \tilde{T}$ respectively, and for the applications in the present article we have $k_1 = k_2 = 5$.

We can take a moment to reflect on the interpretation of the procedure we just completed. For each region, each dimension obtained with PCA is a linear combination of the voxels in the region, whose weights define a multivariate pattern of response over voxels. Considering as an example region 1, the loadings of a dimension i are encoded in the i-th row of $\tilde{E}_{1,red}$, and represent the intensity with which the multivariate pattern corresponding to dimension i is activated over time.

2.3 Modelling Connectivity

In the previous processing stages, the timeseries of BOLD signal have been corrected for brain-wide noise with compcorr and the representational space has been modeled with PCA. We are now ready to model connectivity. As a first step, we split the multivariate timeseries $\tilde{E}_{1,red}$, $\tilde{E}_{2,red}$ into different runs (in order to estimate connectivity parameters and then test them with independent data), obtaining dimensionality reduced timeseries $E^i_{1,red}$ and $E^i_{2,red}$ for each run i, with $i = 1, \ldots, m$. We can then predict the timeseries $E^i_{2,red}$ using as predictors the timeseries $E^i_{1,red}$:

$$E^i_{2,red} = B^i E^i_{1,red} + F^i \tag{5}$$

where F^i are the residuals of the model. This model can be estimated using ordinary least squares (OLS). Fitting this model for all runs yields matrices of parameters $B^1, \ldots, B^m$.

2.4 Predicting Multivariate Timecourses

After having estimated multivariate connectivity parameters for each run, a leave-one-run-out cross validation procedure can be used to predict for each run the multivariate timecourses in region 2 as a function of the multivariate timecourses in region 1 using the average of the connectivity parameters estimated using data from all the other left-out runs. More formally, for each run i, we obtain mean parameters

$$B^{\setminus i} = \frac{\left(\sum_{j=1}^{i-1} B^j + \sum_{j=i+1}^{m} B^j\right)}{m-1} \tag{6}$$

and using these parameters we calculate predicted timecourses in region 2 for run i

$$\hat{E}^i_{2,red} = B^{\setminus i} E^i_{1,red}. \tag{7}$$

We can now calculate the proportion of variance explained by this model with zero free parameters in run i for each dimension k of the representational space of region 2:

$$R^2_k(i) = \frac{\text{var}\left(\hat{E}^i_{2,red}(k,:)\right)}{\text{var}\left(E^i_{2,red}(k,:)\right)}. \tag{8}$$

In order to facilitate comparison with functional connectivity (Biswal et al. [1995]), which uses correlation, we can also calculate the multiple correlation coefficient

$$R_k(i) = \sqrt{R_k^2(i)} \tag{9}$$

for each component k. Note that unlike the case of functional connectivity, we have k_2 values $R_k(i)$, and they are calculated on the basis of the proportion of variance explained in *independent* data. We can obtain a summary measure across all components computing the mean of the multiple correlation coefficients:

$$\bar{R}(i) = \frac{\sum_{k=1}^{k_2} R_k(i)}{k_2}. \tag{10}$$

2.5 Searchlight

For each run i, with $i = 1, \ldots, m$ the procedure described thus far can be applied to a fixed seed region of interest and spherical ROIs centered in all voxels of gray matter, following a searchlight procedure (see Kriegeskorte et al. [2006], Kriegeskorte and Bandettini [2007] for a description of searchlight in a MVPA context). The searchlight produces for each run i and for each voxel v a corresponding mean multiple correlation value $\bar{R}(i)(v)$ (Figure 1). In the empirical applications described in this paper, searchlight analysis was performed within spheres of radius 6mm.

2.6 Statistical Inference

Since the multivariate connectivity map R_{s_j} is always positive, before proceeding with statistical testing the whole-brain mean $\bar{R}_{s_j}$ is subtracted from R_{s_j} obtaining centered maps $\tilde{R}_{s_j}$, to make a permutation test possible.Subsequently, a moderate smoothing can be applied to the searchlight MVPC maps $\bar{R}(i)$ for the individual runs, which are then averaged into a single map R_{s_j} for each participant s_j. Statistical testing will individuate brain regions that show above-average multivariate connectivity with the seed region. The significance of multivariate connectivity can finally be tested with statistical nonparametric mapping (Nichols and Holmes [2002]).

3 Materials and Experimental Procedures

3.1 Participants

A total of ten volunteers (age range 18-50, mean 27.1) participated in the experiment. The volunteers' consent was obtained according to the Declaration of Helsinki (BMJ, 1991, pp. 302, 1194). The project was approved by the Human Subjects Committees at the University of Trento and Harvard University. Data

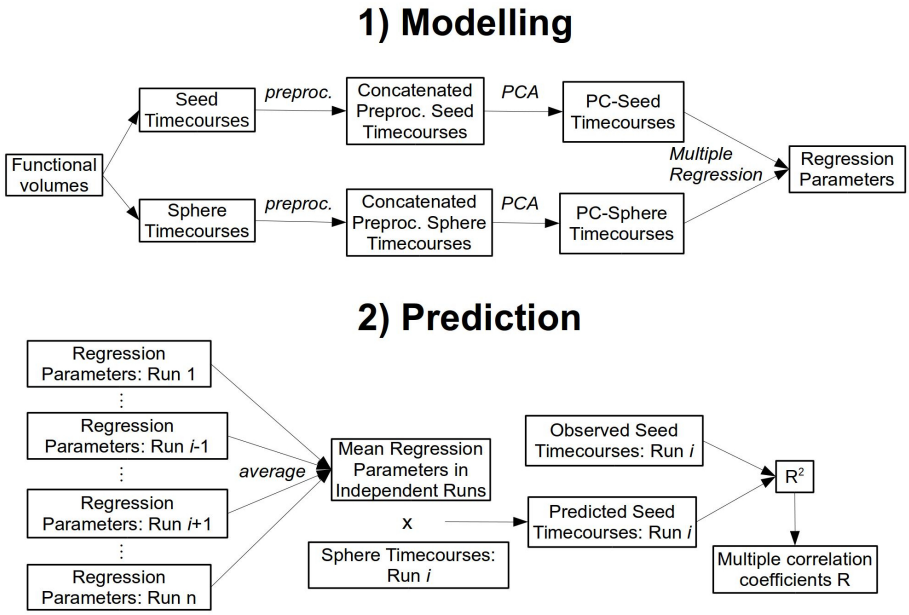

Figure 1: Analysis pipeline.

from one participant were discarded from the analysis because of poor performance during a behavioral training session administered on the day before the scanning.

3.2 Stimuli

Computer generated 3D models (using DAZ-3D) of 5 face identities were used to generate images at 5 different orientations for each identity (Figure 2). Stimuli were presented with Psychtoolbox (Brainard [1997], Pelli [1997]) running on MATLAB, with the add-on ASF (Schwarzbach [2011]), using an Epson EMP 9000 projector. Images were projected on a frosted screen at the top of the bore, viewed through a mirror attached to the head coil.

3.3 Experimental design

One of the five face identities was designated as the target, and participants were instructed to respond with the index finger of the right hand to the target face and with the middle finger to the other distractor faces (Figure 1A). Each trial consisted of the presentation of a face image (500ms) followed by a fixation cross (1500ms). The experiment was composed of three 12-minute runs, each consisting of approximately 320 trials. The order of presentation of the stimuli was generated with optseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/). A

Figure 2: Experimental stimuli.

block-design functional localizer with faces, houses and scrambled images was administered at the beginning of the fMRI session. None of the faces shown in the localizer were presented during the other parts of the experiment.

3.4 Data acquisition and analysis

3.4.1 MRI Scanning Parameters

The data were collected on a Bruker BioSpin MedSpec 4T at the Center for Mind/Brain Sciences (CIMeC) of the University of Trento using a USA Instruments eight-channel phased-array head coil. Before collecting functional data, a high-resolution ($1 \times 1 \times 1\ mm^3$) T1-weighted MPRAGE sequence was performed (sagittal slice orientation, centric phase encoding, image matrix = 256×224 [Read $\times$ Phase], field of view = $256 \times 224\ mm^2$ [Read$\times$Phase], 176 partitions with 1 mm thickness, GRAPPA acquisition with acceleration factor = 2, duration = 5.36 minutes, repetition time = 2700, echo time = 4.18, TI = 1020 $msec$, 7° flip angle). Functional data were collected using an echo-planar 2D imaging sequence with phase oversampling (image matrix = 7064, repetition time = 2000 $msec$, echo time = 21 $msec$, flip angle = 76°, slice thickness = 2 mm, gap = 0.30 mm, with 3×3 mm in plane resolution). Over three runs, 1095 volumes of 43 slices were acquired in the axial plane aligned along the long axis of the temporal lobe.

3.4.2 Preprocessing

Data were preprocessed with SPM8 (http://www.fil.ion.ucl.ac.uk/ spm/software/spm8/) and regions of interest were generated with MARSBAR (Brett et al. [2002]) running on MATLAB 2010a. Subsequent analyses were performed with custom MATLAB software. The first 4 volumes of each run were discarded and all images were corrected for head movement. Slice-acquisition delays were corrected using the middle slice as reference. Images were normalized to the standard SPM8 EPI template and resampled to a 3 mm isotropic voxel size. The BOLD signal was high pass filtered at $128s$ and prewhitened using an autoregressive model AR(1).

3.4.3 MVPC

The preprocessed volumes were then analyzed with MVPC as described above, using 5 principal components in the seed ROI and in the spheres, and smoothing with a 6mm FWHM gaussian kernel. Importantly, for the purposes of the present article, the univariate signal in each region was removed from the multivariate timecourses before performing PCA, subtracting from the multivariate pattern of response in a region the mean response across voxels for each timepoint. This was done by analogy with the removal of the mean in correlation-based MVPA, and enabled us to generate a measure of connectivity complementary to univariate functional connectivity(see Coutanche [2013] for a discussion). In other words, after mean removal, univariate functional connectivity measures would yield zero correlations between the regions: MVPC with mean removal measures the connectivity that can be detected between regions *above and beyond* univariate connectivity.

3.4.4 Functional connectivity

In addition to the MVPC analysis, standard functional connectivity analyses were performed for comparison. To facilitate comparison, functional connectivity was calculated with a searchlight procedure, calculating for the seed and for each sphere of 6mm radius (the same radius used for MVPC) the mean response at each timepoint, which was then low-pass filtered at 0.1Hz and correlated. As for MVPC, for each run the whole-brain mean functional connectivity was removed. The same amount of smoothing ($6mm$ FWHM) was applied to the functional connectivity maps, and statistical testing was performed with SnPM as for MVPC.

4 Results

Three brain regions were found to show significant multivariate connectivity with the FFA seed (Figure 3) even after removing the mean timecourse of responses. First, multivariate connectivity of the FFA with itself across different runs was significant (peak Pseudo-t = 8.89, FWE-corrected p = 0.002, MNI

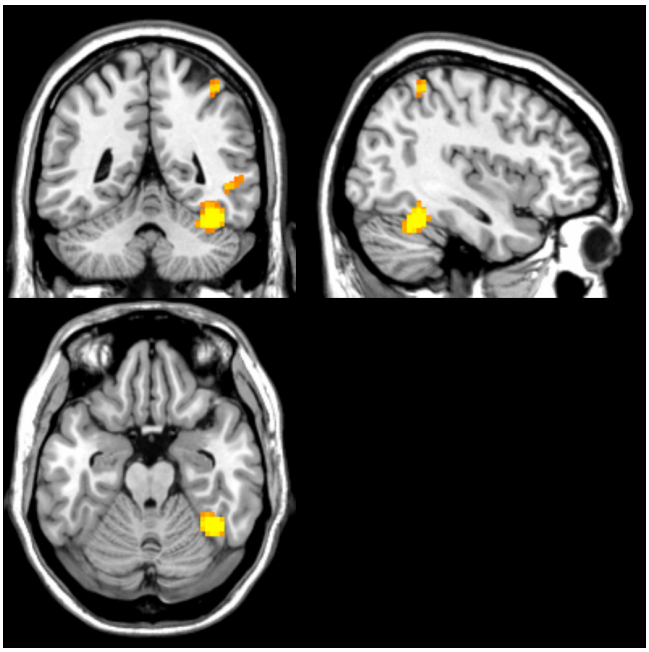

Figure 3: Brain regions individuated by multivariate pattern connectivity (MVPC) at a voxelwise FWE-corrected threshold $p < 0.05$. In addition to the fusiform face area, MVPC individuates the right posterior superior temporal sulcus (STS) and the intraparietal sulcus (IPS). The mean timecourses of responses were removed prior to performing MVPC, which therefore individuates connectivity that goes beyond standard univariate connectivity.

coordinates $[39, -49, -22]$). This is trivially the case because of the absence of time lag: the parameters for self-prediction are necessarily close to 1 across all runs. Second, the posterior superior temporal sulcus (STS) showed significant multivariate connectivity with the FFA (peak Pseudo-$t = 6.79$, FWE-corrected $p = 0.0215$, MNI coordinates $[48, -52, -4]$). Univariate functional connectivity between the FFA and this region was just subthreshold, with Pseudo-$t = 6.03$ (the threshold for FWE-corrected $p < 0.05$ in the functional connectivity data is Pseudo$-t > 6.87$).

Finally, significant multivariate connectivity was found between the FFA seed and the right intraparietal sulcus (IPS; peak Pseudo-$t = 6.84$, FWE-corrected $p = 0.0176$, MNI coordinates $[39, -46, 60]$). Univariate connectivity in the absence of connectivity between the finer grained spatial patterns of responses was found between FFA and the right anterior insula (4)(peak Pseudo-$t = 7.41$, FWE-corrected $p = 0.0254$, MNI coordinates $[33, 29, -2]$).

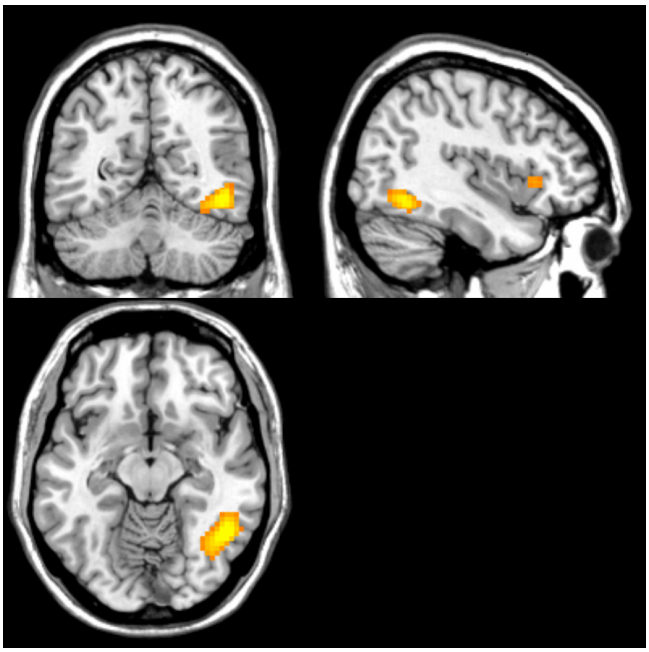

Figure 4: Brain regions individuated by standard univariate functional connectivity with a seed in the fusiform face area (FFA). For visualization purposes, data are thresholded at Pseudo-$t > 5$.

5 Discussion

This article introduces multivariate pattern connectivity (MVPC), a new method to investigate the interactions between brain regions in terms of their multivariate patterns of response. MVPC is characterized by several key properties. First, the BOLD signal in each brain region is modeled as a set of responses along multiple dimensions, with each dimension corresponding to a function of the voxels in that region. Second, MVPC investigates the interaction between two regions as the extent to which the responses in the dimensions characterizing one region can predict the responses in the dimensions characterizing the other region over time. Third, with an analogy to MVPA methods, MVPC uses a cross-validation procedure in which independent data are used for training and testing of the models. A subset of the runs are used as a training set to generate connectivity parameters which are then tested by measuring their ability to predict responses in a left-out independent run. This approach eliminates the bias that would otherwise be caused by sources of noise affecting simultaneously the whole brain.

In the examples described in the present article, dimensions are obtained with PCA as linear combinations of the voxels that tend to be jointly activated or deactivated over time. From a neuroscientific perspective, we can think of each region as consisting of multiple neural populations with selectivities for different properties of the stimuli that have different distributions over the course of the experiment.Each population has different spatial distributions over voxels. This leads different weighted combinations of voxels to having different timecourses of responses, whose dynamics can provide deeper insights into the interactions between regions than the investigation of average responses. Of course,while different populations withdifferent selectivities and different spatial distributions can lead to dimensionswithdifferent time courses,it is unlikely thatindividual dimensions obtainedwith PCA correspond in a one-to-one relationship to neural populations witha specific selectivity profile. For example, more than one neural population might be collapsed in a single principal component, or populations might not be assigned to dimensions in a one-to-one mapping because of the orthogonality constraints imposed by PCA.

In the empirical application of MVPC reported in this article, we observed three possible patterns of results. In some of the regions (FFA and in part STS) both univariate and multivariate connectivity was observed. This type of result indicates that in addition to a relationship between the overall amount of response in these regions, the regions are related in terms of the finer-grained information they encode over time. In the case of IPS, multivariate connectivity was found in the absence of univariate connectivity, indicating that MVPC can be more sensitive than univariate connectivity to detect certain interactions. This finding is of particular interest because the IPS has been shown to encode information about faces even though it does not show face selectivity (**?**).We conjecture that greater sensitivity of MVPC can be observed especially for regions that are engaged in processes that cut across multiple categories - in the case of IPS the attentional selection of individual objects (Xu and Chun [2006,

2009]). Third, the right insula was found to have univariate connectivity with the FFA seed, but not multivariate connectivity. While the overall amount of response in FFA and in the anterior insula is correlated the finer-grained representational structures in one of the regions cannot predict the finer-grained representational structure in the other.

MVPC differs in important respects from previous techniques aimed at studying the dynamic interactions between brain regions in terms of the information they encode. Unlike previous techniques (Coutanche and Thompson-Schill [2014], Henriksson et al. [2015]), MVPC does not rely on discrimination between categories determined by the experimenter, but on dimensions derived in a data-driven fashion. The data-driven dimensions can be related to properties of the stimuli or the task with a subsequent model (for instance regressing dimensions on conditions, or on stimulus properties using a forward model). Another difference between MVPC and the method introduced by Coutanche and Thompson-Schill (Coutanche and Thompson-Schill [2014]) is that the latter characterizes each region with one discrete measure (successful or unsuccessful classification), while MVPC adopts multiple continuous measures (the values along the multiple dimensions), which can provide a richer characterization of a region's representation at any given time. An important feature of MVPC that sets it apart from other methods relating representations across multiple regions is that MVPC uses independent data for the training and testing of the models. This feature helps to control for sources of noise that can affect the entire brain. For example, head motion can lead to different amounts of noise in different trials, which in turn could lead to correlations between regions in terms of whether categorization is successful or unsuccessful in a given trial, inflating estimates of the interaction between two regions obtained correlating classification performance within the same blocks. However, the most important asset of MVPC is probably its flexibility. The framework of 1) modelling representational spaces in individual regions, 2) considering multivariate time-courses as trajectories in these representational spaces, and 3) fitting models predicting the trajectory in the representational space of one region as a function of the trajectory in the representational space in another offers a wealth of possibilities to build increasingly refined models, both in terms of the characterization of representational spaces and in terms of the models of their interactions. For the characterization of representational spaces, in this article we adopted PCA as a simple example, but other methods such as independent component analysis (ICA) and nonlinear dimensionality reduction techniques can also be used. For modelling interactions between regions, we limited the current application to simultaneous, non-directed interactions following an approach similar to functional connectivity, but MVPC makes it possible to model nonlinear maps between representational spaces, and to use models that investigate the directionality of interactions using temporal precedence, along the lines of Granger Causality (Roebroeck et al. [2005]), Dynamic Causal Modelling (Friston et al. [2003]), and Dynamic Network Modelling ().

Acknowledgments

This study was supported by the Center for Mind/Brain Sciences (CIMeC) of the University of Trento and by NIH Grant 1R01 MH096914-01A1 to Prof. Rebecca Saxe. Stefano Anzellotti was supported by a Postdoctoral Fellowship from the Simons Center for the Social Brain.

References

Stefano Anzellotti and Alfonso Caramazza. From parts to identity: invariance and sensitivity of face representations to different face halves. *Cerebral Cortex*, page bhu337, 2015.

Stefano Anzellotti, Scott L Fairhall, and Alfonso Caramazza. Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex*, page bht046, 2013.

Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T Liu. A component based noise correction method (compcor) for bold and perfusion based fmri. *Neuroimage*, 37(1):90–101, 2007.

Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

David H Brainard. The psychophysics toolbox. *Spatial vision*, 10:433–436, 1997.

Matthew Brett, Jean-Luc Anton, Romain Valabregue, and Jean-Baptiste Poline. Region of interest analysis using the marsbar toolbox for spm 99. *Neuroimage*, 16(2):S497, 2002.

Raymond B Cattell. The scree test for the number of factors. *Multivariate behavioral research*, 1(2):245–276, 1966.

Marc N Coutanche. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cognitive, Affective, & Behavioral Neuroscience*, 13(3):667–673, 2013.

Marc N Coutanche and Sharon L Thompson-Schill. Creating concepts from converging features in human cortex. *Cerebral Cortex*, page bhu057, 2014.

Ibai Diez, Asier Erramuzpe, Inaki Escudero, Beatriz Mateos, Alberto Cabrera, Daniele Marinazzo, Ernesto J Sanz-Arigita, Sebastiano Stramaglia, and Jesus M Cortes. Information flow between resting state networks. *arXiv preprint arXiv:1505.03560*, 2015.

Evelina Fedorenko and Sharon L Thompson-Schill. Reworking the language network. *Trends in cognitive sciences*, 18(3):120–126, 2014.

Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

Helen L Gallagher and Christopher D Frith. Functional imaging of theory of mind. *Trends in cognitive sciences*, 7(2):77–83, 2003.

James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

Linda Henriksson, Seyed-Mahdi Khaligh-Razavi, Kendrick Kay, and Nikolaus Kriegeskorte. Visual representations are dominated by intrinsic fluctuations correlated between areas. *NeuroImage*, 114:275–286, 2015.

Alumit Ishai. Lets face it: itsa cortical network. *Neuroimage*, 40(2):415–419, 2008.

Julie Josse and François Husson. Selecting the number of components in principal component analysis using cross-validation approximations. *Computational Statistics & Data Analysis*, 56(6):1869–1879, 2012.

Nancy Kanwisher, Josh McDermott, and Marvin M Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11):4302–4311, 1997.

Jorie Koster-Hale, Rebecca Saxe, James Dungan, and Liane L Young. Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences*, 110(14):5648–5653, 2013.

Nikolaus Kriegeskorte and Peter Bandettini. Analyzing for information, not activation, to exploit high-resolution fmri. *Neuroimage*, 38(4):649–662, 2007.

Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3863–3868, 2006.

Nikolaus Kriegeskorte, Elia Formisano, Bettina Sorger, and Rainer Goebel. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, 104(51):20600–20605, 2007.

Denis Le Bihan, Jean-François Mangin, Cyril Poupon, Chris A Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. Diffusion tensor imaging: concepts and applications. *Journal of magnetic resonance imaging*, 13(4): 534–546, 2001.

Adrian Nestor, David C Plaut, and Marlene Behrmann. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences*, 108(24):9998–10003, 2011.

Thomas E Nichols and Andrew P Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15 (1):1–25, 2002.

Denis G Pelli. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4):437–442, 1997.

Alard Roebroeck, Elia Formisano, and Rainer Goebel. Mapping directed influence over the brain using granger causality and fmri. *Neuroimage*, 25(1): 230–242, 2005.

Jens Schwarzbach. A simple framework (asf) for behavioral and neuroimaging experiments based on the psychophysics toolbox for matlab. *Behavior Research Methods*, 43(4):1194–1201, 2011.

Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5):543–545, 2008.

Yaoda Xu and Marvin M Chun. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440(7080):91–95, 2006.

Yaoda Xu and Marvin M Chun. Selecting and perceiving multiple visual objects. *Trends in cognitive sciences*, 13(4):167–174, 2009.

Keith W Zoski and Stephen Jurs. An objective counterpart to the visual scree test for factor analysis: The standard error scree. *Educational and Psychological Measurement*, 56(3):443–451, 1996.