Review

Beyond Functional Connectivity: Investigating Networks of Multivariate Representations

Stefano Anzellotti[1,3] and Marc N. Coutanche[2,3,*]

For over two decades, interactions between brain regions have been measured in humans by asking how the univariate responses in different regions co-vary ('Functional Connectivity'). Thousands of Functional Connectivity studies have been published investigating the healthy brain and how it is affected by neural disorders. The advent of multivariate fMRI analyses showed that patterns of responses within regions encode information that is lost by averaging. Despite this, connectivity methods predominantly continue to focus on univariate responses. In this review, we discuss the recent emergence of multivariate and nonlinear methods for studying interactions between brain regions. These new developments bring sensitivity to fluctuations in multivariate information, and offer the possibility to ask not only whether brain regions interact, but how they do so.

Highlights

A family of novel methods study the interactions between brain regions taking advantage of the information encoded in multivariate patterns of responses.

Some of these methods additionally capture nonlinear interactions, offering insights into how representations are transformed from brain region to brain region.

These advances enable researchers to ask questions not only about whether, but also about how multiple brain regions interact.

Incorporating Multivariate Information into Our Understanding of Brain Networks

Cognitive tasks engage multiple, interacting brain regions. Different regions have distinct functional roles, and information is transformed from region to region, implementing the computations that make behavior possible. The network of brain regions engaged in face recognition is a particularly well-studied example [1–3]. Visual information is first processed in early visual cortex, encoding basic perceptual dimensions, such as line orientation [4]. This is passed on to the occipital face area (OFA), which shows sensitivity to the presence of face parts, but not to their configuration [5]. In turn, OFA drives responses in the fusiform face area (FFA) and posterior superior temporal sulcus [6]. Responses in FFA influence activity in the anterior temporal lobe (ATL) [7], which encodes even more abstract representations of face identity [8–10], and may function as a hub to integrate knowledge about a person [11,12].

However, even for a relatively well-understood system such as the face network, standard analysis techniques leave open a key question: what is the relationship between the information encoded in one brain region and the information encoded in another? Over the past 5 years, new analysis approaches have been developed that go beyond standard **Functional Connectivity** (see Glossary) to tap into the rich multivariate structure of relationships between the responses in different brain regions, bringing these questions within reach for the fMRI community. In this review, we discuss the state of this emerging field, offering an overview of the available tools and examples of their applications. First, we outline the advances that led to multivariate information encoded by response patterns in different brain regions being taken into account outside the field of connectivity, and how this information is often neglected in connectivity analyses. Second, we discuss recent techniques that fill this gap, covering the approaches used to model multivariate responses in each region and their multivariate interactions. Finally, we discuss the significance of the change of perspective brought about by multivariate connectivity, and point to key directions for future research.

[1] Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA
[2] Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA
[3] These authors contributed equally

*Correspondence:
marc.coutanche@pitt.edu
(M.N. Coutanche).

https://doi.org/10.1016/j.tics.2017.12.002

From Univariate to Multivariate Signal

Traditional approaches to analyzing the **blood oxygenation level-dependent (BOLD) signal** ask whether the magnitude of the responses of a region is modulated by stimuli or tasks (e.g., low versus high working memory), as assessed through univariate statistical tests (e.g., ANOVA) [13,14]. Stronger responses to a stimulus or task are interpreted as reflecting the greater 'involvement' of a region in that stimulus or task. However, univariate differences can reflect a wide variety of neural computations. For example, greater activation to objects might reflect any combination of object detection ('object present'), shape processing, synthesizing viewpoints into invariant object identity, among others.

Multivoxel Pattern Analysis

Developments in analytical techniques over the past 15 years have brought new ways to probe neural activity. **Multivoxel pattern analysis** (MVPA) leverages machine learning and multivariate statistics to measure the information contained within distributed patterns of activity [15]. In this framework, each voxel in a brain region is considered a dimension, and the response in the entire region is viewed as a multidimensional vector (Box 1). Multivariate techniques can capture the unique distributed pattern (a 'fingerprint') corresponding to a condition or task [16]. Often, different conditions can be classified based on patterns of activity, despite being indistinguishable in their univariate signal [17,18]. For example, although some regions respond more to color than to gray-scale images, it is difficult to decode what color a participant is seeing from the univariate response of a region. By contrast, color can be decoded relatively accurately from fMRI responses using MVPA [19].

The ability to interrogate multivariate responses is a significant advance for understanding the neural bases of cognition. The level of specificity that can be extracted from multivoxel patterns is central to many perceptual and cognitive domains. For example, to recognize objects, it is not enough to identify that an object has a shape and a color (reflected in univariate responses). Instead, it is necessary to distinguish specific objects (e.g., 'apple' versus 'baseball'), shapes, or colors [20].

Moving beyond Univariate Approaches to Connectivity

Brain regions are organized into functional networks [21] in which different regions have distinct functional roles, and information is transformed from brain region to brain region to implement the computations that support cognition. Therefore, to fully understand information processing in the brain, it is necessary to study regional interactions. Functional Connectivity has been a dominant approach to studying interactions between brain regions, based on quantifying the

Box 1. Modeling the Multivariate Responses of a Region

We can consider a pattern of responses in a brain region as a point in a high-dimensional space: for fMRI, the coordinates on each dimension correspond to the response intensity in a corresponding voxel. Some analysis techniques that we discuss (Structural Covariance and Transfer Entropy) operate directly on the activity patterns of the voxels. Other techniques (Informational Connectivity) classify the patterns of responses to different conditions for each trial. These patterns can be considered as two different clouds of points (one for each condition) in high-dimensional space. We can look for a surface in the space that separates data points based on their condition, and characterizes each data point by its distance from this separating surface. A different approach (the one used in MVPD) looks for a small set of dimensions that capture most of the variance between response patterns. The response patterns in an experiment (which can be thought of as points in the high-dimensional space of voxels) do not uniformly cover all parts of the high-dimensional voxel space. Instead, usually some parts of the space have a high concentration of points, while other parts are almost empty. Techniques such as PCA (or nonlinear variants such as autoencoders) find a subspace (or submanifold) that approximates the part of the high-dimensional space that contains most of the datapoints. A response pattern can now be described by a smaller number of coordinates in the lower-dimensional space, while still capturing most of the variation between data points.

Glossary

Blood oxygen level-dependent signal (BOLD) signal: an indirect measure of neural activity, measured by fMRI, which results from local changes in magnetic susceptibility in the brain due to alterations in the concentration of oxyhemoglobin and deoxyhemoglobin in blood.

Classifier: an algorithm that, given a set of 'training' data samples $\{(x_i,c_i)\}$ comprising inputs x_i and discrete outputs ('classes') c_i, learns to infer the class c^{new} corresponding to new inputs x^{new}.

Functional Connectivity: the Functional Connectivity between two voxels or brain regions is given by the correlation between their responses over time.

Informational Connectivity: the Informational Connectivity between two brain regions is given by the correlation between fluctuations in multivoxel pattern information over time. Whereas Functional Connectivity examines interactions between regions based on fluctuations in their univariate response, Informational Connectivity examines regional interactions based on shared changes in pattern discriminability over time. 'Discriminability' can be distance from a classifier decision plane (where larger values indicate greater discriminability), or similarity to a 'prototype' pattern (generated from an independent part of the data). The specificity of the conditions helps determine the specificity of the informational link between regions. For example, a subtle distinction, such as between man-made objects, can identify a network that interacts based on fluctuations in object-specific information, such as identity, shape, color, and so on. A broader distinction, such as objects versus faces, will identify a network based on distinctions such as animacy, ability to physically manipulate, and so on.

Multivariate integration: a technique used to study statistical dependence between two brain regions by: (i) calculating the multivariate dissimilarity (i.e., Euclidean distance) between the responses at each pair of time points ('pattern-shift matrix') for each region; and (ii) computing the synchrony of responses between regions (or voxels) over time [22]. Some connectivity studies have applied the measure to resting-state data [23], while others have examined how regional connectivity is modulated by different tasks or stimuli (i.e., psychophysiological interactions [22]). Just as traditional approaches to examining the BOLD response can be blind to information contained in multivariate activity patterns, traditional Functional Connectivity does not measure fluctuations in information that is only represented in multivoxel patterns (Figure 1) [24]. With the importance of brain networks and multivoxel representations in perceptual and cognitive functions, understanding how regional multivoxel information interacts at the network level has great potential for shedding further light on how the brain implements cognition.

The approaches we discuss here are positioned at the intersection of connectivity [25,26] and MVPA [27,28]. Specifically, they take advantage of the information encoded across multiple voxels (as in MVPA), but instead of asking whether the activity patterns of a region can predict perceptual or cognitive characteristics (e.g., the category of an object), they examine the relationship between patterns of multiple brain regions, such as using the patterns from one region to predict patterns in another region. Here, we focus on investigations of multivariate representations in the brain, rather than on studies that examine the univariate response with machine-learning techniques (e.g., [29]). Although such approaches can be informative, once a multivoxel signal is removed (by taking a regional average) or ignored (by examining individual voxels separately), it is impossible to recover, even if multivariate techniques are used to characterize the (univariate) Functional Connectivity. Thus, we focus on approaches that maintain multivoxel information.

Different multivariate connectivity techniques vary in how they describe the multivariate signal of a region (e.g., output of a **classifier** or points in a multidimensional space), and in how they model regional interactions (e.g., calculating correlations, linear regressions, or nonlinear models). We now examine how methods differ in each of these key factors.

Modeling Regions' Multivariate Responses

Informational Connectivity

One approach to modeling multivariate responses across regions consists of training a classifier on multivoxel patterns from different conditions, and asking how regions co-vary (over time) in their representation of this information (Figure 2A). This approach, **Informational Connectivity** [24,30], draws on cross-validation (such as leave-one-run-out) to train a classifier on part of the data set, and then correlate the time courses of classifier performance in a set of regions, using the remaining data. For each time point, classifier performance (a discrete 0/1 value) can be extracted, although a more-refined variant uses a continuous metric reflecting classifier 'confidence' in its guess for a given trial, such as distance from the classification hyperplane, or correlations with held-out prototypical (mean) patterns for each condition [24]. The distance from the separating hyperplane is a measure of the information encoded by the region on a given trial. This continuous dimension of discriminability is a more precise and reliable metric than classification accuracy alone [31]. Drawing on fluctuations in a continuous metric, Informational Connectivity tracks the flow of multivariate information across time. Importantly, applications of the method have shown that two regions with similarly high classification performance are not guaranteed to have significant Informational Connectivity [32] (Box 2). Thus, the approach provides insights into regional interactions, beyond their MVPA performance alone.

In a recent study, Informational Connectivity was used to map a network of object regions in data collected as participants viewed man-made objects [24]. Different types of objects (e.g.,

Spearman correlation between the pattern-shift matrices of the two regions.

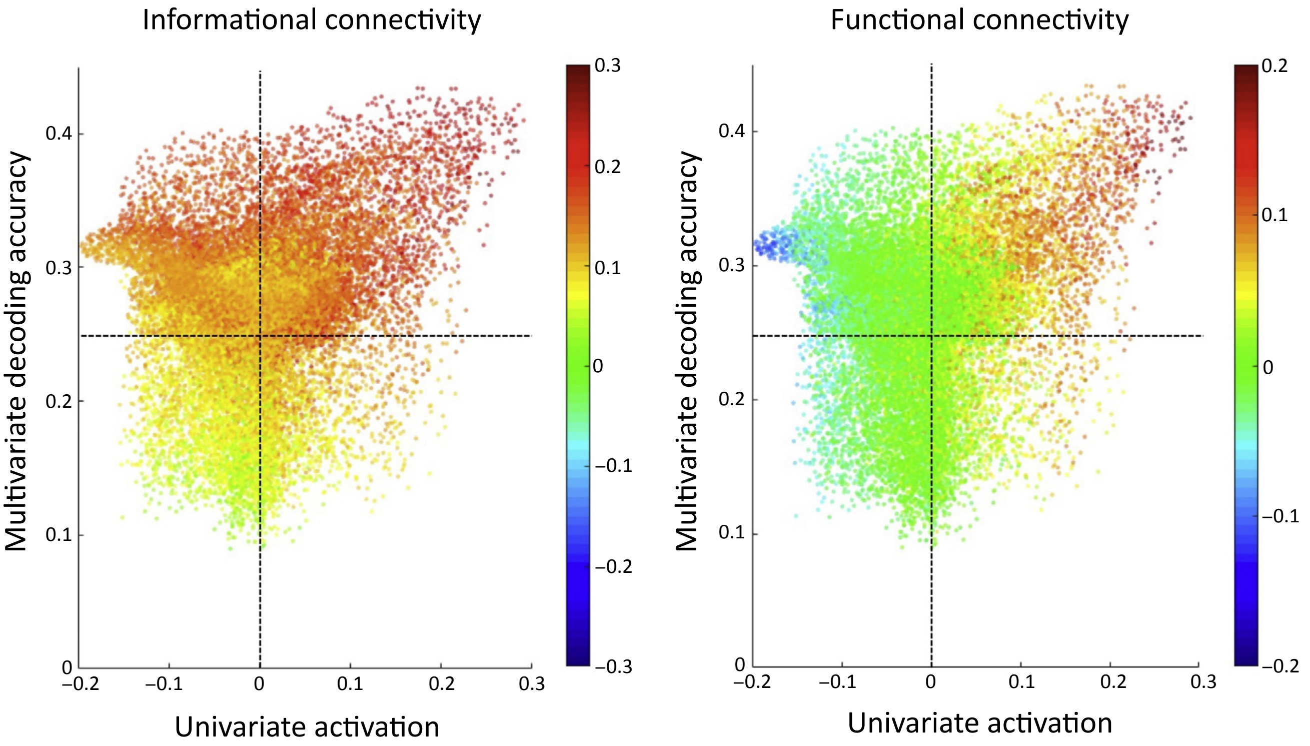
Multivariate pattern dependence (MVPD): a technique to study the statistical relationship between the multivariate responses in multiple brain regions. In MVPD, a data set is divided into two independent parts. The first part of the data is used to estimate a function that predicts multivariate responses in one region as a function of multivariate responses in another region. The remaining part of the data is used to calculate how well the function that was estimated captures the interactions between the two regions, thus reducing the risk of overfitting.

Multivoxel pattern analysis (MVPA): an analysis technique in which a brain region is studied by taking into account distributed responses across its many voxels, instead of computing the average response for a region, or analyzing single voxels separately.

Statistical dependence: the statistical dependence between two variables reflects any statistical relationship between them. Formally, two random variables X,Y show statistical dependence if they do not satisfy the conditions for probabilistic independence [that is if they violate $p(X,Y) = p(X)p(Y)$].

Structural covariance: a multivariate extension of correlation, in which statistical dependence between two regions is calculated by measuring the distance correlation between the multivariate time courses of the regions.

Transfer entropy: a measure of directed transfer of information between two regions, calculated as the reduction in the uncertainty of future responses in the first region due to knowing past responses in the second region (in addition to knowing past responses in the first region itself).



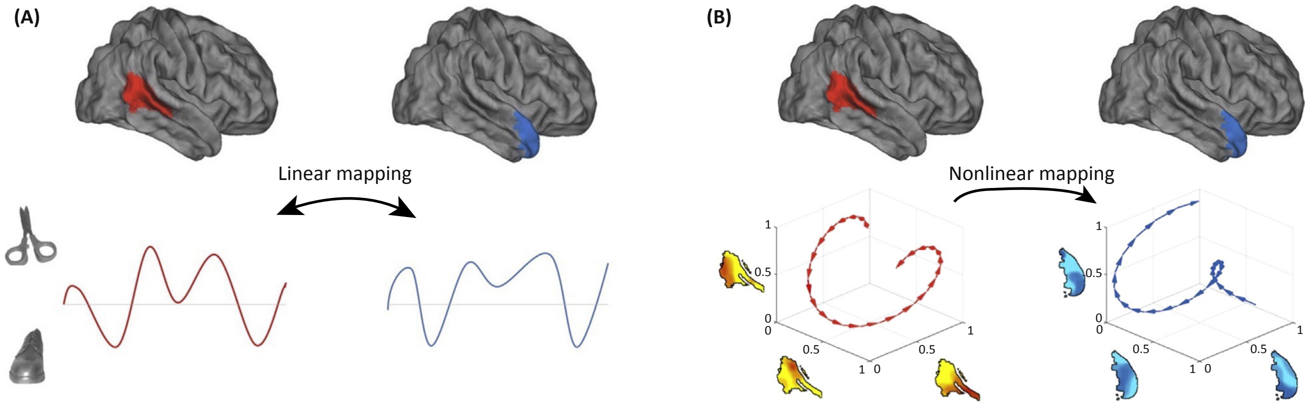
Trends in Cognitive Sciences

Figure 1. Multivariate Information in Connectivity Analyses. The strengths of informational (left) and functional (right) connectivity between regions (searchlights) across the brain and a seed in left fusiform gyrus, while participants viewed four types of man-made object. Connectivity for every brain searchlight is displayed according to a color scale, plotted according to its overall (univariate) mean activation (X-axis), and decoding accuracy (Y-axis) of object type. The broken lines indicate zero univariate activation and at-chance classification performance. A key difference in the graphs is revealed in their top-left quadrant, where Functional Connectivity (right) has low values, while Informational Connectivity (left) has high values. This top-left quadrant shows those areas with high decoding accuracy (i.e., multivariate information) but low univariate activation, the type of region that is not detected through Functional Connectivity. Modified from [24].

chairs versus scissors) can be distinguished in multivoxel patterns, but not univariate activation [28]. In turn, Informational Connectivity revealed a network of regions processing the different types of objects that was not apparent from Functional Connectivity. The strong informational connection found between the ventral and dorsal streams in this study is in line with recent work suggesting that object-identity information is shared, and interacts, between the two streams [33].

Since the method was introduced, another application has revealed that parahippocampal cortex and retrosplenial cortex (RSC) are informationally connected when processing scenes, but not non-scenes, suggesting stimulus-dependent modulation of information synchrony in this network [32]. An additional investigation, in the field of memory, found that successfully remembered items were accompanied by stronger Informational Connectivity (based on attentional state) during an incidental encoding task involving rooms and art, between the hippocampus and RSC [34]. Thus, informational coupling between the hippocampus and a category-selective region during encoding is linked to better subsequent memory.

A distinctive advantage of Informational Connectivity is that it enables researchers to test whether different brain regions interact in terms of their information content along specific, experimenter-defined dimensions. By choosing which classes to use, the experimenter



Trends in Cognitive Sciences

Figure 2. Informational Connectivity and Multivariate Pattern Dependence (MVPD). (A) Informational Connectivity. The spatial patterns of responses across the voxels of a region can represent specific percepts or cognitions. The response pattern at each time point is compared with the prototypical patterns for conditions of interest (through a correlation or classifier), which can comprise two conditions (such as scissors and shoes here) or more (e.g., scissors and other objects). The resulting continuous time course of multivariate information is then compared across regions to quantify their 'Informational Connectivity' [24]. (B) Nonlinear MVPD (NL-MVPD). Each brain region is characterized by a set of dimensions corresponding to spatial patterns of response across voxels. The multivariate time course in each region is modeled as a trajectory in the corresponding multidimensional space, and the statistical dependence between multiple regions is characterized by nonlinear functions (e.g., artificial neural networks) [35].

determines the particular information content used to assess the **statistical dependence** between regions. At the same time, because of the choice of one specific classification dimension, interesting variance in the responses of a region can remain unused and so the method might not provide a complete characterization of the representational content in each region. However, Informational Connectivity might still capture most of the variance in the event that: (i) the classification dimension is similar to the first principal component in both regions; and (ii) the first principal component accounts for most of the variance in the regions. Furthermore, testing interactions along specific, experimenter-defined dimensions renders Informational Connectivity complementary to other methods that study multivariate interactions between brain regions, offering the possibility of testing whether regional interactions are specific to particular representational content.

Multivariate Pattern Dependence

A complementary approach [**multivariate pattern dependence** (MVPD)] does not focus on one experimenter-defined dimension, but aims to capture a larger portion of the variance of each region [35]. To do so, MVPD models the responses in each region using more than one dimension, and these dimensions are chosen to maximize the amount of variance they explain

Box 2. MVPA Does Not Imply Multivariate Dependence

What is the difference between calculating MVPA in two different regions, and studying Informational Connectivity between those two regions? Suppose we found that two regions both classify chairs versus hammers accurately in 60% of the trials. The 60% of the trials in which region A is accurate could be the same 60% of the trials in which region B is accurate. In this case, Informational Connectivity between region A and region B would be high. Alternatively, the 60% of trials might have little overlap between the regions, with only 20% being classified correctly in both regions (and 40% of trials being classified correctly in region A but not B, and a different 40% being classified correctly in region B but not A). In this case, Informational Connectivity between the two regions would be low. MVPA classification in two regions does not always imply high Informational Connectivity between them. Analogous reasoning applies to MVPD and other techniques for studying multivariate interactions: modeling multivariate structure in each region separately does not inform about the temporal relationship between the multivariate structure in one region and the multivariate structure in another.

(e.g., using principal component analysis; PCA). This approach follows a similar logic to PCA models of representational spaces [36]. MVPD favors the use of the term 'statistical dependence' over 'connectivity' because functional 'connectivity' between regions does not imply the existence of direct structural connections [37].

Different regions within a network may represent information along different dimensions. For example, two dimensions describing the position of a stimulus in the visual field may be more suitable to model responses in primary visual cortex (V1) than in inferotemporal cortex (IT), while animacy may be more suitable to model IT responses than responses in V1. MVPD estimates PCA dimensions separately for different regions, so that the dimensions are tailored to the representational contents of each individual region (Figure 2B).

The data-driven nature of MVPD makes it suitable for application to resting-state data, whereas approaches based on the classification of conditions over time are not (there are no stimuli or conditions to be classified). Thus, MVPD can be used to increase the sensitivity, and further our understanding, of the many resting-state applications for which Functional Connectivity is currently used. A similar approach was recently used to show that activity flow through resting-state connections is related to how task-rule information is represented across distributed regions [38].

A hybrid between Informational Connectivity and MVPD is to adopt multiple dimensions for each region (as in MVPD) that, instead of being determined with PCA, are based on classifiers (as in Informational Connectivity). In such an approach, it will be important to take into account potential nonorthogonality between the different dimensions. A relevant recent study investigated how distinct sensory features (color and shape) are integrated to form representations of object identity for known objects [39]. A hybrid between information connectivity and MVPD could draw on classification along multiple dimensions within each region, and use multivariate functions to estimate the mapping between regions. One of the analyses from this study goes in this direction by modeling shape and color decoding in a right V4 region. Blocks with successful shape and color decoding in this same region were more likely to show successful object identity decoding in the left anterior temporal lobe [39].

Multivariate Integration

Another approach (Multivariate Integration) adopts a different strategy [40]. Instead of modeling directly the statistical dependence between multivariate patterns of responses, Multivariate Integration computes the structure of temporal shifts in similarity within each region. The multivariate time course in each brain region is converted into a 'pattern-shift matrix', in which each entry *i*,*j* contains the Euclidean distance between the pattern of responses in the same region at times *i* and *j*. Then, the Spearman correlation between the matrices is calculated. One way to think about Multivariate Integration is that it measures whether two brain regions are similar in terms of when they transition between different states over time, rather than modeling directly how responses in one region are mapped onto responses in another region. A unique asset of Multivariate Integration draws on the similarity between the response patterns in each region at multiple time lags. Multivariate Integration has been used to study task-dependent changes in connectivity for a task in which participants performed simulated driving while listening to either navigation instructions or the radio [40]. In the former 'integrated' task, stronger Multivariate Integration was observed between regions responding selectively during listening and regions responding selectively during driving.

In sum, a variety of different approaches exist to characterize the multivariate responses of different brain regions over time. Many of these approaches share a common feature: they reduce highly complex patterns of response across tens or hundreds of voxels to a small number of meaningful dimensions, chosen because they either help to discriminate between certain stimuli or tasks, or summarize accurately a large proportion of the overall responses.

Modeling Interactions between Regions

Informational Connectivity

Methods used to model interactions between regions differ on whether the ensuing models are trained and tested in independent data, and on whether the models are linear or nonlinear (Box 3). Perhaps the most popular technique to model the statistical dependence between two variables is correlation. Informational Connectivity takes advantage of correlation, but instead of applying it directly to the average time courses of responses in two regions (as with standard Functional Connectivity), it correlates the time courses of classification accuracy or pattern similarity in two or more regions. By virtue of the intervening classification step, Informational Connectivity obtains a multivariate measure of the statistical dependence between brain regions.

Structural Covariance and Transfer Entropy

An alternative approach, **Structural Covariance** [41], directly extends univariate correlation using distance correlation [42], a multivariate measure of statistical dependence. Distance correlation is closely related to a correlation between centered representational dissimilarity matrices. Structural Covariance is particularly appealing for its intuitive nature and, similar to MVPD [35], is data driven: it does not predetermine specific classification dimensions along which statistical dependence is computed. Structural Covariance has been shown to yield more-reliable estimates of the statistical dependence between brain regions compared with Functional Connectivity [41]. However, because of its use of distance correlation, Structural Covariance does not generate predictions about independent data.

In this respect, Structural Covariance is similar to **Transfer Entropy** [43]. In Transfer Entropy, information theoretic techniques are used to obtain a multivariate measure of statistical dependence. An important asset of Transfer Entropy, which sets it apart from Informational Connectivity and Structural Covariance, is its ability to capture nonlinear relationships. Given the limited number of observations in a typical experiment, scaling up to the large numbers of voxels contained in typical regions is a challenge for Transfer Entropy. A clever bootstrapping solution to this problem comprises sampling a small subset of voxels in each region, and averaging Transfer Entropy across multiple subsets of voxels [43]. One possible limitation of this solution is that it could overlook multivariate dependence that jointly involves large numbers of voxels.

Multivariate Pattern Dependence

Adopting an approach inspired by MVPA, MVPD uses independent data to train and test models of the dependence between brain regions. In the linear variant of MVPD, the relationship

Box 3. Modeling Regional Multivariate Interactions

Different techniques to model inter-regional interactions differ along several dimensions. First, models of the interactions between brain regions differ in whether they estimate and test the interactions between regions using independent subsets of the data. For example, correlation and distance correlation use all of the data to assess whether changes in the response in one region correspond to changes in the response in another region. By contrast, other approaches (i. e., regression and artificial neural networks as used in MVPD) use part of the data to model the mapping between the response in one region and the response in another, and use left-out data to test whether the mapping holds up. This latter approach requires more data, but also ensures that the relationships between the responses in different regions are reliable. Second, some models estimate linear relationships between regions (i.e., Informational Connectivity, Structural Covariance, and linear MVPD), while others (e.g., NL-MVPD and Transfer Entropy) estimate nonlinear relationships.

between the responses along the multiple dimensions in one region, and the responses along the dimensions in another region, are modeled using multiple linear regression. This approach is closely related to correlation [44], with the important difference that, given a new, independent data set, multiple regression can use the responses in one region to generate predictions for the responses in the other region. Linear MVPD can detect statistical dependence between brain regions that is not captured by standard Functional Connectivity, and explains more variance in left-out data than does univariate connectivity [45]. In a recent study, linear MVPD was used to show that different dimensions in the FFA show differential statistical dependence with posterior and anterior regions of the ventral stream, respectively [45]. This finding offers an example of a functionally relevant structure that can be observed with measures of multivariate statistical dependence but not with standard Functional Connectivity.

Nonlinear Multivariate Pattern Dependence

Interactions between brain regions are likely nonlinear (Box 4). Synaptic connections between regions can be integrated nonlinearly by the dendritic trees of postsynaptic neurons [46]. Computational models of cognition (e.g., deep neural networks for object recognition) also process their inputs through multiple layers of nonlinear filters [47]. Recent studies have shown that these models provide a good characterization of single neuron responses in visual regions in macaques [48] and of BOLD responses in humans [49,50].

A nonlinear variant of MVPD (NL-MVPD) was recently introduced to capture nonlinear and multivariate statistical dependence between brain regions [35]. In NL-MVPD (Figure 2B), multiple regression is replaced with nonlinear function estimation, for example with artificial neural networks [35]. While neural computations are nonlinear (and thus, in principle, nonlinear models would be needed to capture accurately the statistical dependence between brain regions), it is an open question whether the quantity and quality of fMRI data are adequate to estimate nonlinear models of statistical dependence. A recent study showed that NL-MVPD explains more variance in leave-one-run-out cross-validation than does linear MVPD across a variety of task domains, from person identity recognition to language comprehension [35]. In a network of brain regions engaged in the recognition of person identity, NL-MVPD individuated the expected organization by stimulus modality, differentiating between visual, auditory, and

Box 4. Linear versus Nonlinear Models: MVPA and Multivariate Dependence

Linear classifiers remain dominant in examinations of individual regions (i.e., MVPA). This is because MVPA uses brain data to predict stimulus or task properties, and the goal is to characterize the information content of a brain region. If nonlinear classifiers were used, then it would be unclear whether the information decoded is computed by the brain region analyzed or whether there is an additional contribution of the nonlinear classifier itself. To illustrate this point with an example, if we collected single neuron data from all of primary visual cortex at high temporal resolution, we could generate, using a suitably complex nonlinear classifier, invariant representations of objects and faces (after all, this is what the brain does). However, this would not mean that the representations in primary visual cortex are themselves invariant. Since the goal of MVPA is to characterize the representations in a region, linear classifiers are an ideal choice.

MVPD is fundamentally different from MVPA in this respect; instead of using brain data to predict stimulus properties, MVPD uses brain data to predict other brain data. The goal of MVPD is not to characterize the representations of the predictor region, but to model the transformation of representations across brain regions. Drawing a parallel with the example used for MVPA, MVPD could be applied to model the relationship between responses in primary visual cortex and the responses in downstream regions encoding invariant representations of object identity. A nonlinear transformation can be used to compute invariant object representations (as in deep network models of invariant object recognition [48]), including to capture the statistical dependence between primary visual cortex and downstream regions encoding invariant object representations. The interpretability issues of nonlinear classifiers in MVPA do not affect the logic of multivariate statistical dependence; instead, nonlinearity can be used to capture interactions between brain regions.

multimodal regions even after the univariate signal was removed from the analysis. Within the language network [51], NL-MVPD individuated four subnetworks reliably across two different tasks performed by the participants. Unlike linear MVPD and Functional Connectivity, differences in NL-MVPD between the two tasks could be used to distinguish between the tasks performed by participants [35], suggesting that NL-MVPD is sensitive to task-dependent changes in information processing.

Despite being in the early days of multivariate connectivity, the existing literature already offers a variety of alternatives for modeling multivariate interactions between regions. While we have mostly discussed statistical dependence between pairs of regions, many of the approaches described in this article can, and have been, applied to networks with more than two regions. For example, information within ATL activity patterns have been modeled as a function of patterns in both V4 and LOC (three regions) [39], and regions showing visual and auditory responses have been modeled as a function of regions responding only to visual or only to auditory stimuli [35].

From Studying the Presence of Interactions to Studying Their Kind

A fundamental motivation behind multivariate connectivity is the conviction that asking whether and how much two brain regions interact is far from sufficient to characterize the functioning of a network. It is also necessary to understand the form of their interaction.

One example in this direction is found in a recent investigation of how our knowledge of the color and shape of an object are combined to form its identity [39]. Some theories propose that object knowledge is supported by the same neural systems that underlie perception and action [52–54]. Alternatively, one or more integration areas might bind the properties of an object to form its identity elsewhere in cortex [55,56]. In this study, participants searched for a particular fruit or vegetable inside visual noise during an fMRI scan [39]. Analyzing data from the time period before an object was present in the noise targeted top-down retrieval (rather than bottom-up perception) of object knowledge. Two orthogonal visual dimensions of the searched-for items were decoded from specialized visual areas: shape (spherical versus elongated) in the lateral occipital complex, and color (orange versus green) in V4. To test a hypothesis that the ATL is a semantic hub, or 'convergence zone', the authors tested for contingencies between the distinct types of information in the regions. A logistic regression model used the decoding performance of each block for the features in their respective regions as predictors of object-identity decoding in the left ATL, revealing that the simultaneous presence of shape decoding in LOC, and color decoding in right V4, predicted successful object decoding in left ATL. Interestingly, participants with stronger contingencies between their ATL and LOC/V4, had ATL 'top-down' patterns that were more similar to visually driven 'bottom-up' patterns (recorded as participants passively viewed object exemplars), suggesting that information synchrony might lead to a fuller, more vivid reactivation in the ATL.

With NL-MVPD, multiple alternative models of the interactions between regions can be compared systematically. For example, a recent investigation assessed different accounts of multimodal integration, comparing the variance explained by different neural network architectures [35]. In one architecture, responses in visual regions and auditory regions were integrated nonlinearly within modality, and integration between the modalities was linear. By contrast, in a second architecture, nonlinear integration occurred both within and between modalities. The model with nonlinear integration between modalities was found to explain more variance in leave-one-out cross-validation in the left superior temporal sulcus, previously implicated in the integration between visual and auditory information [57]. With this type of analysis, fMRI data can be used to directly investigate questions about the type of computations implemented by brain networks, fostering increased dialog between neural evidence and theories of cognition [58,59].

Concluding Remarks and Future Perspectives

The past 5 years have seen the development of new techniques to model the interactions between brain regions by taking into account their rich, multivariate structure. Each technique has its own unique advantages and limitations, making it particularly suited to address a subset of research questions (see Outstanding Questions).

Future work might further examine the implications of two regions having greater synchrony in their multivoxel patterns, compared with univariate synchrony, and vice versa. The direction of any such difference might give insights into the specificity with which a brain network is processing a perceptual or cognitive target (e.g., detection of 'an object' versus detection of scissors), in the same manner that comparing MVPA and univariate analyses can shed light on this question [17,60,61].

An important direction for future work will center on extending the potential of current techniques by developing new measures that go beyond an assessment of the strength of interactions, and tap into how information is transformed from brain region to brain region (Box 5). For example, novel measures could characterize the complexity of an interaction between two regions, or the optimal family of functions that can be used to model it. A recent investigation of variance explained as a function of the number of hidden nodes in a neural network offers an early step in this direction [35]. Additional future research could comprise combining multivariate models of connectivity with directed models of interactions between regions, such as Dynamic Causal Modeling [62], Granger Causality [63], or Dynamic Network Modeling [64].

A significant challenge for multivariate connectivity, and for other connectivity methods, lies in their susceptibility to correlated noise that can affect multiple brain regions [65]. This kind of noise leads to the risk of reporting false positives. For this reason, careful denoising is essential for all connectivity analyses. Developing strategies to reduce and ideally eliminate the impact of correlated noise is a key direction for future work.

A promising area of application for multivariate models of statistical dependence will be the study, and possibly diagnosis, of psychiatric disorders. Prior work has shown that MVPA can more-sensitively track individual differences in clinical symptom severity than can univariate methods [66]. Analogously, the increased sensitivity of multivariate models to the interactions between regions could offer an important advantage over standard Functional Connectivity, and could serve to inform personalized medicine.

Box 5. Towards a Unified Framework

To fully realize the potential of multivariate analyses of interactions between regions, future research might seek to integrate the data-driven approaches (such as MVPD [35] and Transfer Entropy [43]) with forward models based on stimulus properties. This integration will enable researchers to interpret nonlinear multivariate mappings in terms of the informational content about the stimuli that they encode and transform.

In parallel, approaches that rely on dimensions defined by the experimenter (such as Informational Connectivity [24,30]) can be extended to multiple dimensions to better capture the information encoded in the responses of each region, and model nonlinear transformations between multiple regions. Ideally, this process will lead to a unified theoretical framework incorporating both multivariate connectivity based on classification along experimenter-defined dimensions, and multivariate connectivity based on data-driven dimensions.

Outstanding Questions

To what extent can we make inferences from imaging data about the underlying computations? Which alternative models of neural computation can be distinguished on the basis of fMRI data, and which distinctions are instead beyond reach?

Can these multivariate connectivity measures be applied successfully to other imaging modalities, such as MEG or Light Sheet Fluorescence Microscopy in nonhuman animals?

How can we continue to improve the ways we minimize the impact of noise on connectivity analyses?

Multivariate connectivity offers a window into transformations of information across brain regions, establishing a 'common language' between the analysis of neuroimaging data and computational models of behavior. How can we take advantage of this common language to formulate unified theories of neural and cognitive computation and test them with a combination of neural and behavioral data?

Acknowledgments

We would like to thank Dae S. Houlihan, Griffin Koch, and Rebecca Saxe for comments on an earlier draft of this article. S. A. was supported by a fellowship from the Simons Center for the Social Brain.

References

1. Fairhall, S.L. and Ishai, A. (2007) Effective connectivity within the distributed cortical network for face perception. *Cereb. Cortex* 17, 2400–2406
2. McGugin, R.W. *et al.* (2012) High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17063–17068
3. Rezlescu, C. *et al.* (2014) Normal acquisition of expertise with greebles in two cases of acquired prosopagnosia. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5123–5128
4. Kamitani, Y. and Tong, F. (2005) Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685
5. Liu, J. *et al.* (2010) Perception of face parts and face configurations: an FMRI study. *J. Cogn. Neurosci.* 22, 203–211
6. Pitcher, D. *et al.* (2014) Combined TMS and FMRI reveal dissociable cortical pathways for dynamic and static face perception. *Curr. Biol.* 24, 2066–2070
7. Collins, J.A. and Olson, I.R. (2014) Beyond the FFA: the role of the ventral anterior temporal lobes in face processing. *Neuropsychologia* 61, 65–79
8. Anzellotti, S. and Caramazza, A. (2016) From parts to identity: invariance and sensitivity of face representations to different face halves. *Cereb. Cortex* 26, 1900–1909
9. Anzellotti, S. *et al.* (2014) Decoding representations of face identity that are tolerant to rotation. *Cereb. Cortex* 24, 1988–1995
10. Nestor, A. *et al.* (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9998–10003
11. Anzellotti, S. (2017) Anterior temporal lobe and the representation of knowledge about people. *Proc. Natl. Acad. Sci. U. S. A.* 114, 4042–4044
12. Wang, H. *et al.* (2017) Different underlying mechanisms for face emotion and gender processing during feature-selective attention: evidence from event-related potential studies. *Neuropsychologia* 99, 306–313
13. DeYoe, E.A. *et al.* (1994) Functional magnetic resonance imaging (FMRI) of the human brain. *J. Neurosci. Methods* 54, 171–187
14. Friston, K.J. *et al.* (1994) Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210
15. Haynes, J.-D. and Rees, G. (2006) Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534
16. Norman, K.A. *et al.* (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430
17. Coutanche, M.N. (2013) Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cogn. Affect. Behav. Neurosci.* 13, 667–673
18. Coutanche, M.N. *et al.* (2016) A meta-analysis of fMRI decoding: quantifying influences on human visual population codes. *Neuropsychologia* 82, 134–141
19. Seymour, K. *et al.* (2010) Coding and binding of color and form in visual cortex. *Cereb. Cortex* 20, 1946–1954
20. Tong, F. and Pratte, M.S. (2012) Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63, 483–509
21. Fox, M.D. *et al.* (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9673–9678
22. Friston, K.J. *et al.* (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6, 218–229
23. van den Heuvel, M.P. and Hulshoff Pol, H.E. (2010) Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534
24. Coutanche, M.N. and Thompson-Schill, S.L. (2013) Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. *Front. Hum. Neurosci.* 7, 1–14
25. Biswal, B.B. *et al.* (2010) Task-dependent individual differences in prefrontal connectivity. *Cereb. Cortex* 20, 2188–2197
26. Biswal, B. *et al.* (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541
27. Diedrichsen, J. and Kriegeskorte, N. (2017) Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Comput. Biol.* 13, e1005508
28. Haxby, J.V. *et al.* (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430
29. Finn, E.S. *et al.* (2015) Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671
30. Coutanche, M.N. and Thompson-Schill, S.L. (2014) Using informational connectivity to measure the synchronous emergence of fMRI multi-voxel information across time. *J. Vis. Exp.* 89, 51226
31. Walther, A. *et al.* (2016) Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* 137, 188–200
32. Huffman, D.J. and Stark, C.E.L. (2014) Multivariate pattern analysis of the human medial temporal lobe revealed representationally categorical cortex and representationally agnostic hippocampus. *Hippocampus* 24, 1394–1403
33. Bracci, S. and Op de Beeck, H. (2016) Dissociations and associations between shape and category representations in the two visual pathways. *J. Neurosci.* 36, 432–444
34. Aly, M. and Turk-Browne, N.B. (2016) Attention promotes episodic encoding by stabilizing hippocampal representations. *Proc. Natl. Acad. Sci. U. S. A.* 113, E420–E429
35. Anzellotti, S. *et al.* (2017) Measuring and modeling nonlinear interactions between brain regions with fMRI. *bioRxiv* Published online September 12, 2017. http://dx.doi.org/10.1101/074856
36. Tamir, D.I. *et al.* (2016) Neural evidence that three dimensions organize mental state representation: rationality, social impact, and valence. *Proc. Natl. Acad. Sci. U. S. A.* 113, 194–199
37. Honey, C.J. *et al.* (2009) Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2035–2040
38. Ito, T. *et al.* (2017) Cognitive task information is transferred between brain regions via resting-state network topology. *Nat. Commun.* 8, 1027
39. Coutanche, M.N. and Thompson-Schill, S.L. (2015) Creating concepts from converging features in human cortex. *Cereb. Cortex* 25, 2584–2593
40. Sasai, S. *et al.* (2016) Functional split brain in a driving/listening paradigm. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14444–14449
41. Geerligs, L. *et al.* (2016) Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. *Neuroimage* 135, 16–31
42. Székely, G.J. *et al.* (2007) Measuring and testing dependence by correlation of distances. *Ann. Stat.* 35, 2769–2794
43. Lizier, J.T. *et al.* (2011) Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *J. Comput. Neurosci.* 30, 85–107
44. Rodgers, J.L. and Nicewander, W.A. (1988) Thirteen ways to look at the correlation coefficient. *Am. Stat.* 42, 59–66

45. Anzellotti, S. *et al.* (2016) Multivariate pattern connectivity. *bioRxiv* 2016, 046151

46. Xu, N. *et al.* (2012) Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature* 492, 247–251

47. Krizhevsky, A. *et al.* (2012) ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105

48. Yamins, D.L.K. *et al.* (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8619–8624

49. Khaligh-Razavi, S.-M. and Kriegeskorte, N. (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* 10, e1003915

50. Kriegeskorte, N. (2015) Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* 1, 417–446

51. Fedorenko, E. *et al.* (2011) Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16428–16433

52. Binder, J.R. and Desai, R.H. (2011) The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536

53. Martin, A. and Chao, L.L. (2001) Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201

54. Martin, A. (2007) The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45

55. Damasio, A.R. (1989) The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* 1, 123–132

56. Lambon Ralph, M.A. *et al.* (2010) Coherent concepts are computed in the anterior temporal lobes. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2717–2722

57. Beauchamp, M.S. *et al.* (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823

58. Coltheart, M. (2006) What has functional neuroimaging told us about the mind (so far)? *Cortex* 42, 323–331

59. Mather, M. *et al.* (2013) How fMRI can inform cognitive theories. *Perspect. Psychol. Sci.* 8, 108–113

60. Brants, M. *et al.* (2011) Multiple scales of organization for object selectivity in ventral visual cortex. *Neuroimage* 56, 1372–1381

61. Davis, B. *et al.* (2014) Functional and developmental significance of amplitude variance asymmetry in the BOLD resting-state signal. *Cereb. Cortex* 24, 1332–1350

62. Friston, K.J. *et al.* (2003) Dynamic causal modelling. *Neuroimage* 19, 1273–1302

63. Roebroeck, A. *et al.* (2005) Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* 25, 230–242

64. Anzellotti, S. *et al.* (2017) Directed network discovery with dynamic network modelling. *Neuropsychologia* 99, 1–11

65. Henriksson, L. *et al.* (2015) Visual representations are dominated by intrinsic fluctuations correlated between areas. *Neuroimage* 114, 275–286

66. Coutanche, M.N. *et al.* (2011) Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity. *Neuroimage* 57, 113–123