Review TRENDS in Cognitive Sciences Vol.xxx No.x

Full text provided by www.sciencedirect.com ScienceDirect

Beyond mind-reading: multi-voxel pattern analysis of fMRI data

Kenneth A. Norman[1], Sean M. Polyn[2], Greg J. Detre[1] and James V. Haxby[1]

[1] Department of Psychology, Princeton University, Green Hall, Washington Road, Princeton, NJ 08540, USA
[2] Department of Psychology, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104, USA

A key challenge for cognitive neuroscience is determining how mental representations map onto patterns of neural activity. Recently, researchers have started to address this question by applying sophisticated pattern-classification algorithms to distributed (multi-voxel) patterns of functional MRI data, with the goal of decoding the information that is represented in the subject's brain at a particular point in time. This multi-voxel pattern analysis (MVPA) approach has led to several impressive feats of mind reading. More importantly, MVPA methods constitute a useful new tool for advancing our understanding of neural information processing. We review how researchers are using MVPA methods to characterize neural coding and information processing in domains ranging from visual perception to memory search.

Introduction

The most fundamental questions in cognitive neuroscience deal with the issue of *representation*: what information is represented in different brain structures; how is that information represented; and how is that information transformed at different stages of processing? Functional MRI (fMRI) constitutes a powerful tool for addressing these questions: While a subject performs a cognitive task, we can obtain estimates of local blood flow (a proxy for local neural processing) from tens of thousands of distinct neuroanatomical locations, within a matter of seconds. However, the large size of these datasets (up to several gigabytes) and the high levels of noise inherent in fMRI data pose a challenge to researchers interested in mining these datasets for information about cognitive processes.

Traditionally, fMRI analysis methods have focused on characterizing the relationship between cognitive variables and individual brain voxels (volumetric pixels). This approach has been tremendously productive. However, there are limits on what can be learned about cognitive states by examining voxels in isolation. The goal of this article is to describe a different approach to fMRI analysis, where — instead of focusing on individual voxels — researchers use powerful pattern-classification algorithms, applied to multi-voxel patterns of activity, to decode the information that is represented in that pattern of activity. We call this approach multi-voxel pattern analysis (MVPA).

The idea of applying multivariate methods to fMRI data (i.e. analyzing more than one voxel at once) is not new. For example, several researchers have used multivariate methods to characterize functional relationships between brain regions (e.g. [1–5]). A major development in the last few years is the realization that fMRI data analysis can be construed, at a high level, as a pattern-classification problem (i.e. how we can recognize a pattern of brain activity as being associated with one cognitive state versus another). As such, all of the techniques that have been developed for pattern classification and data mining in other domains (e.g. handwriting recognition) can be productively applied to fMRI data analysis. This realization has led to a dramatic increase in the number of researchers using pattern-classification techniques to analyze fMRI data. This trend in the fMRI literature is part of a broader trend towards the application of pattern-classification methods in neuroscience (for applications to EEG data, see [6–11]; for applications to neural recording data from animal studies, see [12–14]).

The first part of the article provides an overview of the main benefits of the MVPA approach, as well as a listing of some of the feats of 'mind reading' that have been accomplished with MVPA. The next part provides a more detailed overview of the methods that make this mind reading possible. The third part of the article discusses some case studies in how researchers can go beyond mind reading (for its own sake), and use MVPA to address meaningful questions about how information is represented and processed in the brain.

The benefits of MVPA

More sensitive detection of cognitive states

Given the goal of detecting the presence of a particular mental representation in the brain, the primary advantage of MVPA methods over individual-voxel-based methods is increased sensitivity. Conventional fMRI analysis methods try to find voxels that show a statistically significant response to the experimental conditions. To increase sensitivity to a particular condition, these methods spatially average across voxels that respond significantly to that condition. Although this approach reduces noise, it also reduces signal in two important ways: First, voxels with weaker (i.e. non-significant) responses to a particular condition might carry some information about the presence/absence of that condition. Second, spatial averaging blurs out fine-grained spatial patterns that might discriminate between experimental conditions [15].

Corresponding author: Norman, K.A. (knorman@princeton.edu).
Available online xxxxxx.

www.sciencedirect.com 1364-6613/$ – see front matter © 2006 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2006.07.005

Like conventional methods, the MVPA approach also seeks to boost sensitivity by looking at the contributions of multiple voxels. However, to avoid the signal-loss issues mentioned above, MVPA does not routinely involve spatial averaging of voxel responses. Instead, the MVPA approach uses pattern-classification techniques to extract the signal that is present in the pattern of response across multiple voxels, even if (considered individually) the voxels might not be significantly responsive to any of the conditions of interest. The multi-voxel pattern of response can be thought of as a combinatorial code with a very large capacity (not yet precisely quantified) for representing distinctions between cognitive states. Because MVPA analyses focus on high-spatial-frequency (and often idiosyncratic) patterns of response, MVPA analyses are typically conducted within individual subjects.

A study by Haxby *et al.* [16] illustrates how multi-voxel patterns of activity can be used to distinguish between cognitive states. Subjects viewed faces, houses, and a variety of object categories (e.g. chairs, shoes, bottles). The data were split in half, and the multi-voxel pattern of response to each category in ventral temporal (VT) cortex was characterized separately for each half. By correlating the first-half patterns with the second-half patterns (within a particular subject), Haxby *et al.* were able to show that each category was associated with a reliable, distinct pattern of activity in VT cortex (e.g. the first-half 'shoe' pattern matched the second-half 'shoe' pattern more than it matched the patterns associated with other categories); see [17–22] for similar results, and Section 3 for additional discussion of this work.

In addition to decoding the category of a viewed object, MVPA methods have been to used to decode (among other things) the orientation of a striped pattern being viewed by the subject [23,24]; the direction of movement of a viewed field of dots [25]; whether the subject is looking at a picture or a sentence; whether the subject is reading an ambiguous versus a nonambiguous sentence; and the semantic category of a viewed word (the last three examples are from [26]). All of those studies deal with decoding the properties of a perceived stimulus. Other MVPA studies have focused on decoding properties of the subject's cognitive state that cannot be inferred from simple inspection of the stimulus, for example: whether the subject is lying about the identity of a playing card [27]; which of two rival stimuli is being perceived at a particular moment in a binocular rivalry paradigm [28]; which of two overlapping striped patterns [23] or moving dot patterns [25] the subject is attending to during a particular trial; and which of three categories the subject is thinking about during a memory retrieval task [29]. Throughout this article, we will be using the term 'mind reading' inclusively, to refer to all of the types of decoding mentioned above (note also that some mind-reading studies have used conventional fMRI analysis methods instead of MVPA; see for example [30]).

Relating brain activity to behavior on a trial-by-trial basis

The increased sensitivity afforded by MVPA methods makes it feasible to measure the presence/absence of cognitive states based on only a few seconds' worth of brain activity. If the cognitive states in question are sufficiently distinct from one another, discrimination can be well above chance based on single brain scans (acquired over a period of ~2–4 s) [19,22,24,26,28,29,31–34]. This increase in temporal resolution makes it possible to create a temporal trace of the waxing and waning of a particular cognitive state over the course of the experiment, which (in turn) can be related to subjects' ongoing behavior. The ability to correlate classifier estimates with behavioral measures across trials (within individual subjects, over the course of a single experiment) is one of the most important benefits of the MVPA approach. Although the temporal resolution of MVPA is intrinsically limited by temporal dispersion in the hemodynamic response measured by fMRI, extant studies have used MVPA to successfully resolve cognitive changes that occur on the order of seconds. For example, MVPA has been used to predict the time course of recall behavior in a free-recall task [29], and it has also been used to predict second-by-second changes in perceived stimulus dominance during a binocular rivalry task [28].

Characterizing the structure of the neural code

In addition to allowing us to sensitively detect and track cognitive states, MVPA methods can be used to characterize how these cognitive states are represented in the brain. The MVPA approach assumes that cognitive states consist of multiple aspects ('dimensions'), and that different values along a particular dimension are represented by different patterns of neural firing. This implies that we can measure how strongly cognitive dimension x is represented in brain region y, by measuring how much the pattern of neural activity in region y changes, as a function of changes along dimension x. Here, we are using 'region y represents dimension x' to mean 'region y carries information about dimension x'; as with all other neuroimaging results, this is no guarantee that region y plays a causal role in enacting behavior based on dimension x.

One concrete way to test hypotheses about whether region y represents cognitive dimension x is to measure how well a pattern classifier, applied to voxels in region y, can discriminate between cognitive states that vary along dimension x (but see Section 2 for important caveats about how to interpret good classifier performance). A more powerful extension of this approach is to vary similarity in a graded fashion along dimension x and see if classifier performance decreases in a graded fashion as similarity increases; examples of this approach are discussed in Section 3 [22,23]. An alternative approach to studying neural coding is to compare the brain patterns without passing them through a classifier; for example, [35] used multidimensional scaling (applied to raw brain data) to show that multi-voxel patterns in lateral occipital cortex tend to cluster by object category. Finally, it is worth noting that MVPA is not the only way to measure the similarity of neural representations with fMRI (see, e.g. the fMRI-adaptation approach described by [36]).

MVPA methods

The basic MVPA method is a straightforward application of pattern classification techniques, where the patterns to

Please cite this article as: Kenneth A. Norman et al., Beyond mind-reading: multi-voxel pattern analysis of fMRI data, TRENDS in Cognitive Sciences (2006), doi:10.1016/j.tics.2006.07.005.

be classified are (typically) vectors of voxel activity values. Figure 1 illustrates the four basic steps in an MVPA analysis. The first step, *feature selection*, involves deciding which voxels will be included in the classification analysis (Figure 1a); Box 1 describes feature selection in more detail. The second step, *pattern assembly*, involves sorting the data into discrete 'brain patterns' corresponding to the pattern of activity across the selected voxels at a particular time in the experiment (Figure 1b). Brain patterns are labeled according to which experimental condition generated the pattern; this labeling procedure needs to account for the fact that the hemodynamic response measured by the scanner is delayed and smeared out in time, relative to the instigating neural event. The third step, *classifier training*, involves feeding a subset of these labeled patterns into a multivariate pattern classification algorithm. Based on these patterns, the classification algorithm learns a function that maps between voxel activity patterns and experimental conditions (Figure 1c). The fourth step is *generalization testing*: Given a new pattern of brain activity (not previously presented to the classifier), can the trained classifier correctly determine the experimental condition associated with that pattern (Figure 1d)?

Choosing a classifier

Machine learning researchers have developed an enormous range of classification algorithms that can potentially be used in MVPA studies (see [37] for details of the classification algorithms discussed below). Most MVPA studies have used linear classifiers, including correlation-based classifiers [16,17], neural networks without a hidden layer [29], linear discriminant analysis [19,22,24,28], linear support vector machines (SVMs) [20,23,26], and Gaussian Naive Bayes classifiers [26]. These classifiers all compute a weighted sum of voxel activity values; this weighted sum is then passed through

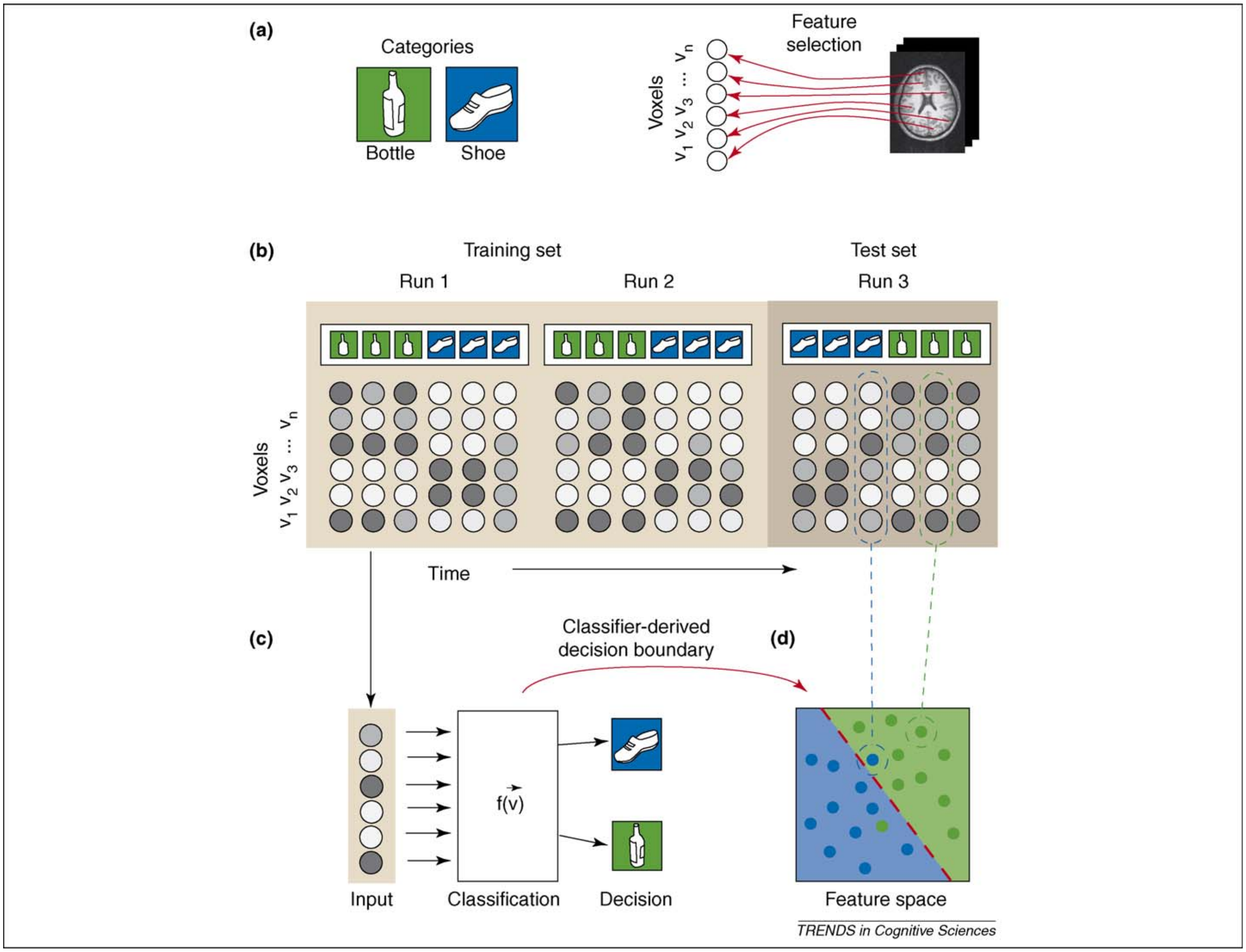

Figure 1. Illustration of a hypothetical experiment and how it could be analyzed using MVPA. **(a)** Subjects view stimuli from two object categories (bottles and shoes). A 'feature selection' procedure is used to determine which voxels will be included in the classification analysis (see Box 1). **(b)** The fMRI time series is decomposed into discrete brain patterns that correspond to the pattern of activity across the selected voxels at a particular point in time. Each brain pattern is labeled according to the corresponding experimental condition (bottle versus shoe). The patterns are divided into a training set and a testing set. **(c)** Patterns from the training set are used to train a classifier function that maps between brain patterns and experimental conditions. **(d)** The trained classifier function defines a decision boundary (red dashed line, right) in the high-dimensional space of voxel patterns (collapsed here to 2-D for illustrative purposes). Each dot corresponds to a pattern and the color of the dot indicates its category. The background color of the figure corresponds to the guess the classifier makes for patterns in that region. The trained classifier is used to predict category membership for patterns from the test set. The figure shows one example of the classifier correctly identifying a bottle pattern (green dot) as a bottle, and one example of the classifier misidentifying a shoe pattern (blue dot) as a bottle.

Please cite this article as: Kenneth A. Norman et al., Beyond mind-reading: multi-voxel pattern analysis of fMRI data, TRENDS in Cognitive Sciences (2006), doi:10.1016/j.tics.2006.07.005.

Box 1. Feature selection methods

As discussed in Section 1, one of the defining features of MVPA is that it factors in contributions from voxels that do not meet conventional criteria for statistical significance. However, there is a cost to being too inclusive: Voxels with especially high levels of noise (and low levels of signal) can sharply reduce classifier performance. This suggests that classifier performance will benefit from *feature selection* methods that can remove noisy and/or uninformative voxels before classification. At this point, we have enough experience with feature selection to know that it is valuable, but our understanding of how to best implement feature selection is still preliminary.

One way to select features is to limit the analysis to specific anatomical regions (e.g. Haxby *et al.* [16] focused on ventral temporal cortex in their study of visual object processing). Another approach to feature selection is to compute univariate (voxel-wise) statistics; for example, one can select out the voxels that, considered individually, do the best job of discriminating between the conditions of interest [16,29,26]. Indeed, any univariate statistic used in conventional fMRI analysis can be used for feature selection.

The main concern with univariate feature selection methods is that, even with a liberal threshold, it is possible that these methods are discarding voxels that (when taken in aggregate) would have provided useful information about the experimental conditions. We can avoid this problem if we replace univariate feature selection methods with multivariate feature selection methods that evaluate *sets* of voxels, based on the informativeness of patterns of activity expressed over those voxels. A challenge faced by this approach is that (because of combinatorial explosion issues) the space of voxel sets is much too large to search exhaustively. This issue can be addressed by constraining the search to sets of spatially adjacent voxels [15], or by adding voxels to the set one a time to maximize (at each step) the multivariate goodness of the current voxel set (see [47] and Bryan and Haxby, unpublished).

a decision function, which effectively creates a threshold for saying whether or not a category is present.

Other MVPA analyses have used nonlinear classifiers; examples include nonlinear support vector machines [20,27] and neural networks with hidden layers [21]. The key difference between nonlinear and linear classifiers is that nonlinear classifiers can respond to high-level feature conjunctions in a way that differs from their response to individual features. For example, a nonlinear classifier can learn that coactivity of voxels a and b signals the presence of a particular cognitive state, even if voxel a and voxel b (considered on their own) do not convey information about that state.

Although nonlinear classifiers are more powerful than linear classifiers (in terms of the types of mappings they can learn), extant MVPA studies have not found a clear performance benefit for nonlinear versus linear classifiers (for a direct comparison, see [20]). Furthermore, Kamitani and Tong [23] have argued that good performance in a nonlinear classifier is harder to interpret than good performance in a linear classifier (Box 2).

MVPA case studies: going beyond mind-reading

The previous two sections focused on describing the MVPA method and how it affords increased sensitivity in detecting cognitive states. In this section, we present two case studies that show how these methodological advances are being harnessed to test theories of how the brain processes visual information [23,24]. Box 3 presents a case study of how MVPA methods are being employed to study memory retrieval [29].

Box 2. Classifier-based brain mapping

If a classifier performs well, what inferences can we make about the properties of the voxels being classified?

Inferences about specific voxels

Given a trained classifier, several methods have been developed for reading out which voxels are contributing the most to classifier performance. For linear classifiers, one can discern the contribution of voxel i to detecting category j by looking at the weight between voxel i and category j [23,29,32]. For nonlinear classifiers (e.g. nonlinear support vector machines), the process of determining a voxel's importance is more complex, insofar as each voxel's contribution to recognizing a category is a function of multiple learned weights (see, e.g. [21,27] for ways of addressing this problem). Maps of 'important' voxels derived using the above methods provide insight into the basis for classification. However, these maps are not guaranteed to include all the voxels that are involved in representing the categories of interest. For example, classifiers tend to focus on discriminative features and ignore features that are shared across categories.

Inferences about voxel sets

The primary added value of the MVPA approach (with regard to brain mapping) is that we can characterize the coding properties of voxel *sets* (i.e. does region y code for cognitive dimension x) by looking at how well a classifier performs when applied to that voxel set (see Section 1). However, as pointed out by Kamitani and Tong [23], the inferences that one can make about neural coding depend on the type of classifier that is being used. Because linear classifiers integrate the evidence provided by each voxel (separately) about category membership (see Section 2), linear classifiers will show above-chance classification *only if* some voxels are individually sensitive to the dimension of interest. By contrast, nonlinear classifiers can show good classification even if none of the input voxels are individually sensitive to the dimension of interest. For example, multi-layer neural networks can learn to classify the emotion expressed by a face based on a bitmap photograph of the face [48], even though single pixels in the bitmap do not provide information about emotion; in this case, emotion information is implicit in the pattern of input layer activity and is rendered explicit (via a series of nonlinear transformations) by the classifier. This property of nonlinear classifiers means that, if they perform well, it is unclear whether input voxels (taken on their own) directly code for the dimension of interest, or whether the classifier is extracting information that is implicitly represented in the pattern of activity across voxels.

Decoding the neural representation of visual object categories

The finding (mentioned in Section 1) that different visual object categories are associated with different voxel activity patterns in VT cortex [16,17,19–22] does not, by itself, tell us how these object categories are represented. Several researchers have conducted follow-up analyses to explore the structure of object category representations in VT. In one such analysis, Haxby *et al.* [16,38] found that voxels showing submaximal responses are informative about category membership: Discrimination between pairs of categories (e.g. shoes versus bottles) was still well above chance when the voxels that responded most strongly to those categories (relative to the other categories) were excluded from the analysis; for a related analysis see [22]. This finding suggests that the neural representations of categories such as faces and houses have a broader spatial extent than was previously thought to be the case. Importantly, these results do not imply that all of the voxels in VT are equally involved in representing all categories; for example, voxels in the fusiform face area

Please cite this article as: Kenneth A. Norman et al., Beyond mind-reading: multi-voxel pattern analysis of fMRI data, TRENDS in Cognitive Sciences (2006), doi:10.1016/j.tics.2006.07.005.

Box 3. Using MVPA to study memory search

A recent study by Polyn *et al.* [29] set out to test the contextual reinstatement hypothesis of memory search [49,50]. This hypothesis states that subjects target memories from a particular episode (or type of episode) by activating knowledge about the general properties of that event, which in turn triggers recall of specific details from that event. To test this hypothesis, MVPA methods were used to calculate the degree to which patterns of brain activity recorded during recall matched those seen during the initial encoding phase, on a time-varying basis (see Figure I). During the initial part of the experiment, subjects studied celebrity faces, famous locations, and common objects. A neural network classifier was trained to recognize patterns of brain activity corresponding to studying faces, locations, and objects. Then, subjects were asked to recall (in any order they liked, over a three minute period) the names of all of the faces, locations, and objects that they had studied earlier in the experiment, and the classifier was used to track the re-emergence (during this recall period) of brain patterns from the study phase.

In keeping with the idea that subjects think about general event properties to remember specific details, Polyn *et al.* found that category-specific patterns of brain activity (associated with studying faces, locations, and objects) started to emerge ~5.4 s before recall of specific items from that category [29]. Over the course of the recall period, fluctuations in the strength of 'neural reinstatement' were highly correlated with subjects' recall behavior (Figure I).

This study is not the first to show reinstatement of study-phase brain activity during recall [41–45]. The main difference between the Polyn *et al.* study and these other studies is that, because of the increased sensitivity of the MVPA approach, Polyn *et al.* were able to track the temporal dynamics of reinstatement over the course of the recall period. The finding that reinstatement precedes recall provides some initial evidence in support of the contextual reinstatement hypothesis. In future studies, it will be important to show that the results extend to other types of recall 'contexts' besides semantic categories.

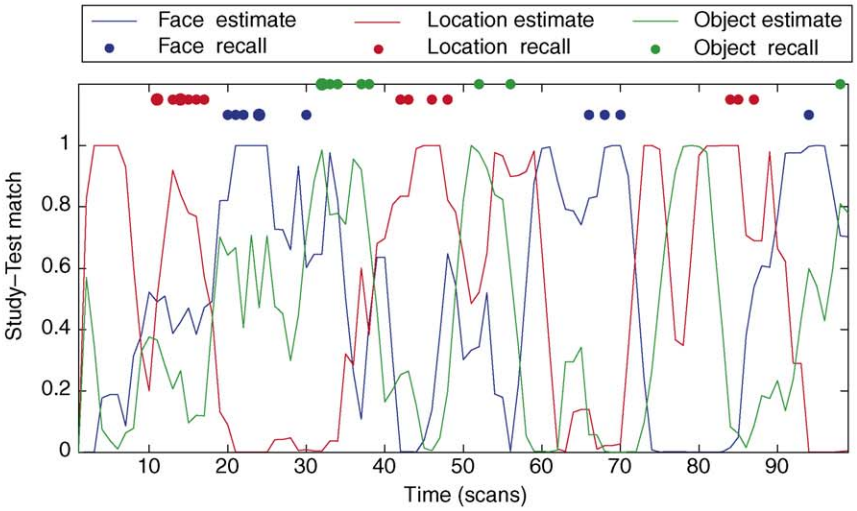

Figure I. Illustration of how brain activity during recall relates to recall behavior, in a single subject. Each point on the *x*-axis corresponds to a 1.8 s interval (during the 3-min recall period). The blue, red, and green lines correspond to the classifier's estimate as to how strongly the subject is reinstating brain patterns characteristic of face-study, location-study, and object-study at that point in time. The blue, red, and green dots indicate time points where subjects recalled faces, locations, and objects; the dots were shifted forward by three time-points, to account for the lag in the peak hemodynamic response. The graph illustrates the strong correspondence between the classifier's estimate of category-specific brain activity, and the subject's actual recall behavior. (Reprinted with permission from [29].)

do not appear to discriminate between shoes and bottles [17]. Overall, the results suggest that information is represented in a partially distributed fashion in VT: The neural substrates of different categories overlap, and the degree of overlap is proportional to the similarity of the categories. In support of this claim, O'Toole *et al.* [22] applied a classifier to VT cortex and showed that the classifier's ability to discriminate between different object types decreased as the visual similarity of those objects increased (see also [21]).

Decoding the neural representation of line orientation. A recent study by Kamitani and Tong [23] used MVPA to study the neural representation of line orientation in visual cortex. Electrophysiological and optical imaging studies have established that orientation-selectivity in primary visual cortex (V1) exists at the level of cortical columns, which cycle through all orientations approximately every millimeter (see, e.g. [39]). Given that multiple orientation-selective columns fit within a single 3-mm cubic fMRI voxel, the Kamitani and Tong study constitutes a test of whether MVPA methods can be used to characterize neural codes that exist at the subvoxel level.

To assess the sensitivity of different regions of visual cortex to orientation information, linear support vector machines were trained to recognize patterns of brain activity associated with viewing gratings (striped patterns) with different orientations. There were eight classifiers in total, corresponding to angles from 0 degrees to 157.5 degrees in 22.5-degree increments. After training, the classifiers were applied to new patterns and the output of the classifiers was used to estimate the orientation of the viewed gratings (by selecting the orientation corresponding to the classifier with the largest output value). The accuracy of these linear classifiers, applied to a particular region, can be used as an aggregate index of the sensitivity of individual voxels in that region (Box 2).

The study found that, although orientation is coded at a subvoxel level in early visual cortex, there were small irregularities in how strongly different orientations were represented by each voxel. By combining information across multiple voxels, the classifier was able to exploit these small irregularities to generate an accurate readout

Please cite this article as: Kenneth A. Norman et al., Beyond mind-reading: multi-voxel pattern analysis of fMRI data, TRENDS in Cognitive Sciences (2006), doi:10.1016/j.tics.2006.07.005.

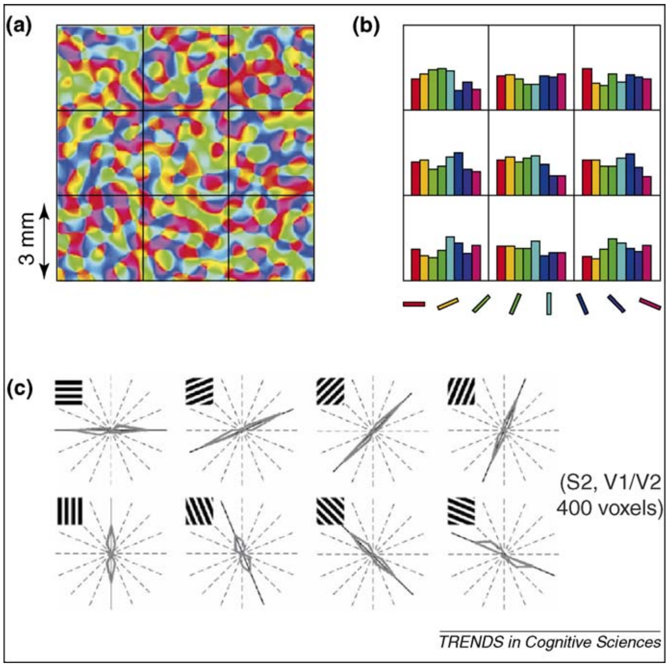

Figure 2. Orientation decoding from fMRI activity in visual cortex. Parts **(a)** and **(b)** (adapted with permission from [51]) illustrate how voxels acquire weak sensitivity for line orientation. Part (a) shows a simulated orientation tuning map for a patch of visual cortex (different colors indicate different orientations), with a voxel-sized (3 mm) grid superimposed on the map. Part (b) shows the distribution of orientation selectivity values for each of the nine 'voxels' shown in Part (a). Although all of the orientations are represented inside each voxel, the distribution of selectivity values is slightly different for each voxel. The classifier is able to exploit these small per-voxel irregularities in selectivity to decode orientation from multi-voxel patterns (see also [24], Supplementary Figure 1). Part **(c)** (adapted with permission from [23]) illustrates the performance of the classifier in [23] for a single subject. The polar plots show the classifier's orientation predictions for eight different (actual) line orientations; predictions were based on 400 voxels in V1 and V2. For these voxels, most of the classifier's predictions exactly matched the correct orientation, and the classifier's (rare) mistakes were all tightly clustered around the correct orientation.

of line orientation based on activity in areas V1 and V2. They also found that the classifier was more likely to confuse similar (versus dissimilar) orientations (Figure 2). When the classifier was applied to other visual subregions, lower-level regions showed more orientation-selectivity than higher-level regions: V1/V2 showed the best sensitivity, V3 was slightly worse, V4 was slightly worse still, and there was no orientation selectivity in area MT+. Overall, these results are highly consistent with the results of prior studies of how orientation is represented in different parts of visual cortex [40].

A study conducted by Haynes and Rees [24] provides converging evidence for the representation of line orientation information in early visual cortex. They used a visual masking technique to prevent subjects from consciously perceiving the orientation of presented gratings; behaviorally, subjects' ability to discriminate between line orientations was at chance. Despite this total lack of behavioral discrimination ability, a linear discriminant classifier (applied to V1 voxels) was nonetheless able to decode the orientation of the masked lines with greater than chance accuracy.

Conclusions

MVPA has evolved extensively in the 5 years since the publication of the Haxby *et al.* [16] object categories study, and we expect that MVPA methods will continue to evolve rapidly in the coming years. A promising development in this regard is the debut, earlier this year, of an annual 'brain activity interpretation' competition (see http://www.ebc.pitt.edu/competition.html). This competition should facilitate the development of better algorithms for feature selection and classification by allowing researchers to benchmark different algorithms on a common dataset. Other factors should also boost MVPA performance: Improvements in the spatial resolution of fMRI will make it possible to resolve even finer-grained cognitive distinctions [46], and improvements in computer speed will make it possible to search through an even larger number of voxel sets (to find the most informative set; see Box 1). For all of these reasons, we believe that MVPA has a bright future (see Box 4) as a tool for characterizing how information is represented and processed in the brain.

Box 4. Questions for future research

- The line-orientation studies described in Section 3 [23,24] showed that MVPA can be used to confirm the properties of a well-understood neural code. Can we use MVPA to decipher the properties of neural codes that are less well-understood, e.g. the neural code for face identity?
- What are the limits on the kinds of representations that can be resolved by applying MVPA methods to standard-resolution fMRI data? What are the limits on the kinds of representations that can be resolved using high-resolution fMRI data?
- Existing methods for aligning brain data across subjects (using structural data) are suboptimal for MVPA analyses, because these methods blur out high-spatial-frequency patterns that (in individual subjects) carry information about cognitive states. Can we devise improved methods for translating between subjects' functional brain states, such that a classifier trained on high-spatial-frequency information in subject A will generalize well to subject B?
- One of the weaknesses of extant MVPA methods is that the classifier is not provided with information about spatial relationships between the to-be-classified voxels (i.e. which voxels are nearby in 3-D space). As such, classifiers have no natural way to leverage the *topography* of cortical representations – the fact that spatially proximal voxels tend to represent similar things. How can we make better use of spatial information in MVPA analyses?
- What is the most effective way of searching the brain for voxel sets that optimally satisfy a multivariate criterion (e.g. classifiability)?
- Can we use MVPA methods to improve our ability to recognize cognitive states (from fMRI) in real time? If so, can we use these methods to provide 'cognitive biofeedback' to help subjects learn to control their thoughts [52,53] or external devices?
- Can we use MVPA to study how representations change as a function of learning?
- Most of the MVPA studies reviewed here treat cognitive states as discrete, unitary entities (e.g. is the person viewing a shoe or a bottle). This conflicts with the view, prevalent among psychologists, that cognitive states can be viewed as points in a high-dimensional 'cognitive space', where the distance between two points in this cognitive space corresponds to the psychological similarity of the two cognitive states (see [54] for discussion of how to build a cognitive 'face space' based on face similarity ratings). What are the most effective methods for mapping (in a continuous fashion) between brain states and points in cognitive space?

References

1 Friston, K.J. and Buchel, C. (2003) Functional connectivity. In *Human Brain Function* (2nd edn) (Frackowiak, R.S.J. *et al.*, eds), Academic Press

2 Friston, K.J. *et al.* (2003) Dynamic causal modelling. *Neuroimage* 19, 1273–1302
3 McIntosh, A.R. *et al.* (1996) Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage* 3, 143–157
4 McIntosh, A.R. and Lobaugh, N.J. (2004) Partial least squares analysis of neuroimaging data: applications and advances. *Neuroimage* 23, S250–S263
5 Calhoun, V.D. *et al.* (2001) Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum. Brain Mapp.* 13, 43–53
6 Peters, B.O. *et al.* (1998) Mining multi-channel EEG for its information content: an ANN-based method for a brain-computer interface. *Neural Netw.* 11, 1429–1433
7 Parra, L. *et al.* (2002) Linear spatial integration for single-trial detection in encephalography. *Neuroimage* 17, 223–230
8 Muller-Putz, G.R. *et al.* (2005) EEG-based neuroprosthesis control: a step towards clinical practice. *Neurosci. Lett.* 382, 169–174
9 Vallabhaneni, A. and He, B. (2004) Motor imagery task classification for brain computer interface applications using spatiotemporal principle component analysis. *Neurol. Res.* 26, 282–287
10 Wang, T. *et al.* (2004) Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. *Clin. Neurophysiol.* 115, 2744–2753
11 Philiastides, M.G. and Sajda, P. (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb. Cortex* 16, 509–518
12 Carmena, J.M. *et al.* (2003) Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* 1, E42
13 Hung, C.P. *et al.* (2005) Fast readout of object identity of macaque inferior temporal cortex. *Science* 310, 863–866
14 Tsao, D.Y. *et al.* (2006) A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674
15 Kriegeskorte, N. *et al.* (2006) Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3863–3868
16 Haxby, J.V. *et al.* (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2429
17 Spiridon, M. and Kanwisher, N. (2002) How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35, 1157–1165
18 Tsao, D.Y. *et al.* (2003) Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* 6, 989–995
19 Carlson, T.A. *et al.* (2003) Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15, 704–717
20 Cox, D.D. and 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, 261–270
21 Hanson, S.J. *et al.* (2004) Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a 'face' area? *Neuroimage* 23, 156–166
22 O'Toole, A.J. *et al.* (2005) Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* 17, 580–590
23 Kamitani, Y. and Tong, F. (2005) Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685
24 Haynes, J.D. and Rees, G. (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8, 686–691
25 Kamitani, Y. and Tong, F. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* (in press)
26 Mitchell, T.M. *et al.* (2004) Learning to decode cognitive states from brain images. *Mach. Learn.* 5, 145–175
27 Davatzikos, C. *et al.* (2005) Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage* 28, 663–668
28 Haynes, J.D. and Rees, G. (2005) Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.* 15, 1301–1307
29 Polyn, S.M. *et al.* (2005) Category-specific cortical activity precedes recall during memory search. *Science* 310, 1963–1966
30 O'Craven, K.M. and Kanwisher, N. (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* 12, 1013–1023
31 LaConte, S. *et al.* (2003) The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. *Neuroimage* 18, 10–27
32 LaConte, S. *et al.* (2005) Support vector machines for temporal classification of block design fMRI data. *Neuroimage* 26, 317–329
33 Strother, S. *et al.* (2004) Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. a preliminary group analysis. *Neuroimage* 23 (Suppl 1), S196–S207
34 Mourao-Miranda, J. *et al.* (2005) Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *Neuroimage* 28, 980–995
35 Edelman, S. *et al.* (1998) Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26, 309–321
36 Grill-Spector, K. and Malach, R. (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst.)* 107, 293–321
37 Duda, R.O. *et al.* (2001) *Pattern Classification,* (2nd edn), Wiley
38 Haxby, J.V. (2004) Analysis of topographically organized patterns of response in fMRI data: distributed representations of objects in ventral temporal cortex. In *Attention and Performance XX* (Kanwisher, N. and Duncan, J., eds), Oxford University Press
39 Bartfeld, E. and Grinvald, A. (1992) Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proc. Natl. Acad. Sci. U. S. A.* 89, 11905–11909
40 Vanduffel, W. *et al.* (2002) The organization of orientation selectivity throughout macaque visual cortex. *Cereb. Cortex* 12, 647–662
41 Wheeler, M.E. *et al.* (2000) Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11125–11129
42 Nyberg, L. *et al.* (2000) Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11120–11124
43 Wheeler, M.E. and Buckner, R.L. (2003) Functional dissociation among components of remembering: control, perceived oldness, and content. *J. Neurosci.* 23, 3869–3880
44 Kahn, I. *et al.* (2004) Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. *J. Neurosci.* 24, 4172–4180
45 Smith, A.P.R. *et al.* (2004) fMRI correlates of the episodic retrieval of emotional contexts. *Neuroimage* 22, 868–878
46 Sayres, R. *et al.* (2005) Identifying distributed object representations in human extrastriate cortex. In *Advances in Neural Information Processing Systems Vol. 18 (Weiss, Y. et al.*, eds), pp. 1169–1176, MIT Press
47 Padmala, S. and Pessoa, L. (2005) The dream of a single image for a single event: decoding near-threshold perception of fear from distributed single-trial brain activation. *Soc. Neurosci. Abstr. 2005 Abstract Viewer/Itinerary Planner* No. 193.2
48 Cottrell, G.W. *et al.* (2001) Is all face processing holistic? The view from UCSD. In *Computational, Geometric, and Process Perspectives on Facial Cognition* (Wenger, M.J. and Townsend, J.T., eds), pp. 347–396, Erlbaum
49 Tulving, E. and Thompson, D. (1973) Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80, 352–373
50 Bartlett, F.C. (1932) *Remembering: A Study in Experimental and Social Psychology.* Cambridge University Press
51 Boynton, G.M. (2005) Imaging orientation selectivity: decoding conscious perception in V1. *Nat. Neurosci.* 8, 541–542
52 deCharms, R.C. *et al.* (2004) Learned regulation of spatially localized brain activation using real-time fMRI. *Neuroimage* 21, 436–443
53 deCharms, R.C. *et al.* (2005) Control over brain activation and pain learned by using real-time functional MRI. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18626–18631
54 Wenger, M.J. and Townsend, J.T., (eds) (2001) *Computational, Geometric, and Process Perspectives on Facial Cognition*, Erlbaum

Please cite this article as: Kenneth A. Norman et al., Beyond mind-reading: multi-voxel pattern analysis of fMRI data, TRENDS in Cognitive Sciences (2006), doi:10.1016/j.tics.2006.07.005.