

Patterns of Activity in the Categorical Representations of Objects

Thomas A. Carlson, Paul Schrater, and Sheng He

Abstract

■ Object perception has been a subject of extensive fMRI studies in recent years. Yet the nature of the cortical representation of objects in the human brain remains controversial. Analyses of fMRI data have traditionally focused on the activation of individual voxels associated with presentation of various stimuli. The current analysis approaches functional imaging data as collective information about the stimulus. Linking activity in the brain to a stimulus is treated as a pattern-classification problem. Linear discriminant analysis was used to reanalyze a set of data originally published by Ishai et al. (2000), available from the fMRIDC (accession no. 2-2000-

1113D). Results of the new analysis reveal that patterns of activity that distinguish one category of objects from other categories are largely independent of one another, both in terms of the activity and spatial overlap. The information used to detect objects from phase-scrambled control stimuli is not essential in distinguishing one object category from another. Furthermore, performing an object-matching task during the scan significantly improved the ability to predict objects from controls, but had minimal effect on object classification, suggesting that the task-based attentional benefit was non-specific to object categories. ■

INTRODUCTION

The human brain is arguably the most powerful computational architecture known to humanity and represents one of nature's crowning achievements. One of the most striking aspects of the brain is its seemingly limitless capacity for representing information. Deservingly, considerable effort has been put forth to understand its functional and neuroanatomical architectures. Here, we apply pattern-classification methods to fMRI data in an effort to further investigate the nature of how categorical information is represented in the brain.

The debate about how categorical information is represented in the cortex is often summarized into two competing theories: distributed vs. modular. The modular viewpoint proposes that the cortex can be divided into distinct modules dedicated to processing and representing particular types of information (Fodor, 1983). This view has been supported by evolutionary considerations and by findings that some neuropsychological patients exhibited category-specific deficits. In recent years, more evidence supporting this claim has come from functional neuroimaging. Cortical regions were identified that showed preferential activation to specific categories of information, such as the fusiform face area (FFA), the parahippocampal place area (PPA), and more recently, the extrastriate body area (EBA) (Downing, Jiang, Shuman, & Kanwisher, 2001; Epstein

& Kanwisher 1999; Kanwisher, McDermott, & Chun, 1997). However, there remain strong challenges to this domain-specificity view. For example, it was shown that FFA also responded robustly to nonface stimuli in the form of novel synthetic objects called greebles, leaving questions as to whether these modules were indeed category specific (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Conversely, the distributed hypothesis submits that information is diffusely represented in the cortex. This viewpoint, until recently, lacked strong evidence in human neuroimaging and principally relied on evidence drawn from nonhuman primate studies (Tanaka, 1996, 1997; Wang, Tanaka, & Tanifuji, 1996) and computational modeling (Rumelhart & McClelland, 1986; Hinton & Anderson, 1981). fMRI evidence supporting the distributed perspective came when it was reported that previously identified modules in the ventral temporal cortex for faces, houses, and chairs responded robustly to all three categories of objects (Ishai, Ungerleider, Martin, & Haxby, 1999). While the response in these areas to nonpreferred object categories was smaller than that to the preferred category, it nevertheless was significant. Based on these findings, the authors proposed the object form topology hypothesis—that object form is represented in the ventral temporal cortex continuously in a distributed and overlapping arrangement.

More recently, it was argued that the smaller response in ventral temporal cortex to nonpreferred stimuli has a functionally important role as the response conveys

University of Minnesota

information sufficient to determine the stimulus category (Haxby et al., 2001). The argument was based on a novel classification analysis that predicts the object category by computing correlations between regression coefficients for each of the stimulus categories. The analysis showed that object categories could be predicted robustly based on responses from the ventral temporal cortex even after removing areas that responded maximally to that category. For example, face stimulus could be distinguished from others based on activities from ventral areas without FFA. Furthermore, activity in regions such as the FFA could predict non-preferred stimuli such as houses and chairs from other categories of objects.

Addressing the question of how object categories are represented in the cortex requires establishing a link between the object category being viewed by the observer and brain activity. Analysis methods such as the general linear model (GLM) establish this link by determining how well a model can account for the activity in a given voxel (Friston et al., 1995). Haxby et al.'s (2001) analysis establishes this link by creating a template of activity for object categories via regression coefficients, in which the coefficients are simply the mean activity across time in a subset of scans for each voxel,¹ and using them to predict subsequent functional imaging runs. The current analysis sought to establish this link in a novel way through the application of linear discriminant analysis (LDA).

LDA,² a traditional form of statistical classification analysis, was used to understand what features in the activation vector distinguish different object categories. Instead of using stimulus category to identify voxels as significantly active, we seek to find which voxels contribute to the pattern of activity that is indicative of the stimulus category the observer is viewing. The analysis produces a spatial map called a “discriminant” or canonical variant that is a spatial weighting of voxels that can be used to reliably predict the category of object presented to the subject based on a single fMRI time acquisition.

One of the key steps in linking an activation pattern to object categories is to find areas that respond well for all members of an object category. This raises a problem for using mean activation across large spans of time to characterize object category representation. Means are notoriously sensitive to outliers, and very strong responses within a voxel to just a few of the stimuli within a class can be enough to generate a significant mean, even though this voxel would have almost no predictive power for an object category on a time-point by time-point basis. Thus, mean activations cannot always distinguish between voxels that code information about an object category and voxels that code information about particular stimuli within a class (feature encoders). To address this, the current analysis uses individual time acquisitions to test its predictions.

In addition, in contrast to voxel-based methods, categories of objects will be linked to “patterns of activity” across the cortex. This is only natural given that neuronal structures in the brain are highly interconnected and given that the object form topology hypothesis suggests that objects are coded as pattern of activity in a large-scale distributed network. The current analysis treats each functional acquisition as a spatially structured, high-dimensional activation vector that resides in an activity space. Figure 1A shows the transformation between activity space and voxel space. In activity space, the activity in the cortex evoked by a stimulus is treated as a whole and the significance of a voxel is based on its importance to the pattern of activation that distinguishes object categories, as opposed to how well it correlates with a model of the expected response.

An additional concern is that the results from traditional analysis are highly dependent on the control stimulus. As shown in Figure 1B, two hypothetical areas are coding two classes of stimuli. In the traditional sense, the first model is completely modular with the respective regions becoming active only when the preferred stimulus is present; the second model is distributed in that both areas are simultaneously coding each of the two classes of stimulus. Interestingly, the only difference between these two models is the location of the origin, which is defined by the control stimulus! The current analysis principally focuses on activity that defines object categories relative to other categories of objects. This activity, we would argue, is more indicative of the categorical representation of objects as it is defined by their relationship to each other. In the example of Figure 1B, the relationship between the two classes of stimulus remains constant for both models regardless of the location of the origin as shown by the hypothetical d' measure.

Prior to the application of the LDA, two preprocessing steps, in addition to those performed by Ishai, Ungerleider, Martin, Schouten, and Haxby (2000), were taken to maximize the information associated with individual time points and to minimize outliers that would influence the analysis. First, a deconvolution operation was performed on the data to remove the lag due to hemodynamic response and decorrelate the time series data relative to an estimate of the hemodynamic response function (Boynton et al., 1996). The second preprocessing step removed outliers from the data set using a previously described method (Rousseeuw & Van Driessen, 1999).

The application of LDA to fMRI data is carried out in two distinct steps: dimensionality reduction and classification (Figure 2). Dimensionality reduction consisted of removing subcortical regions and areas outside of the brain, and extracting the top 40 principle components in the data. Next, classification was implemented on the reduced data. Individual time acquisitions were labeled

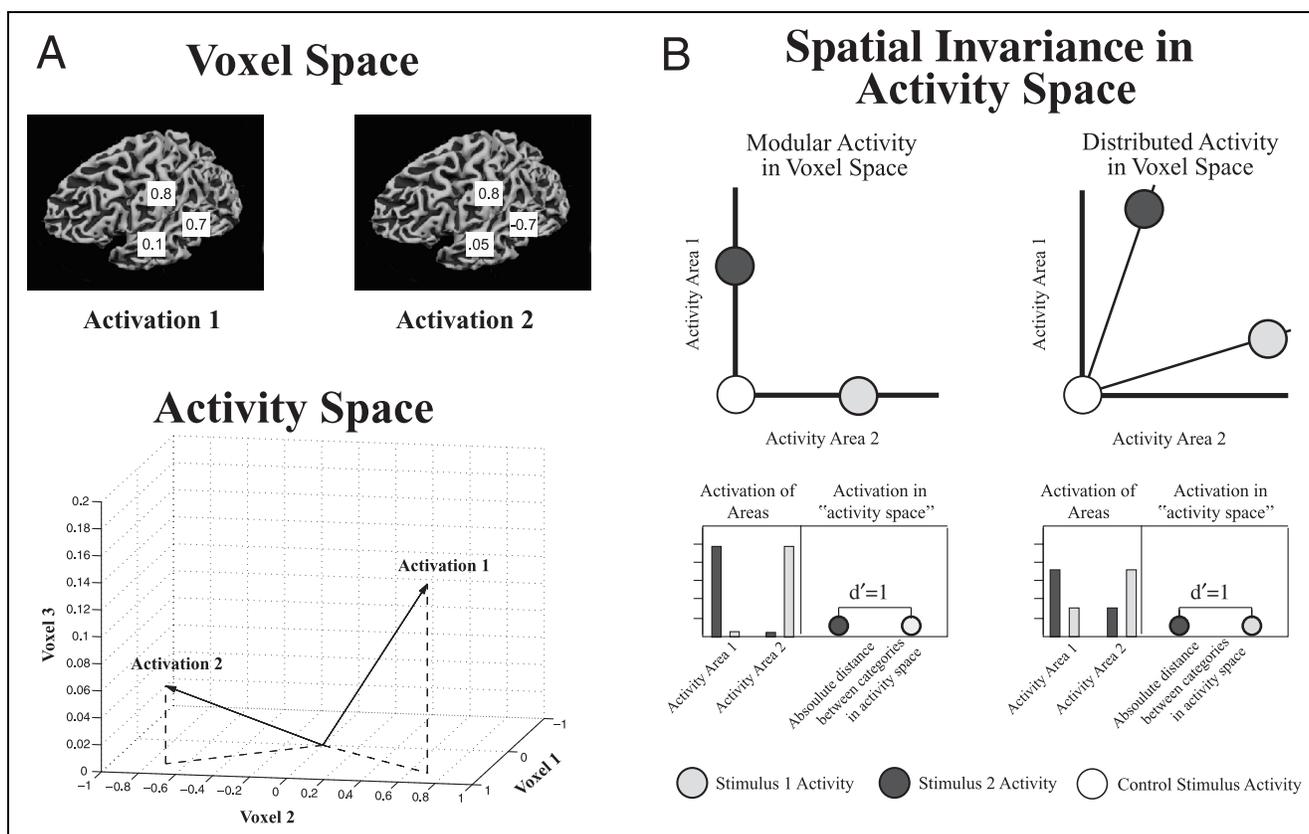


Figure 1. Schematic diagram showing the transformation from voxel space to activity space. (A) Voxel space. Three sample voxels are shown with different levels of activation in the two conditions. For example, Voxel 2 is positively activated in Condition 1 but negatively activated in Condition 2. Activity space. The pattern of activity of the three sample voxels is replotted in the activity space with voxels as axes. Each condition can now be represented by its position in the multidimensional activity space. (B) Spatial invariance in activity space. Two arbitrary categories of objects are plotted in two coordinate systems. In the modular system, each category of objects activates only a single area. In the distributed system, each category of objects simultaneously activates both areas. The only difference between the two models is the placement of the origin, which is defined by the control stimulus.

according to stimulus condition (i.e., chairs, faces, and houses) and a discriminant axis was fit to optimize classification performance. The fits were validated by measuring the predictive performance of the linear classifier using the 632+ bootstrap (Efron et al., 1997), a leave-one-out resampling procedure.

Patterns of activity were identified that support several types of discriminations among the three categories of stimuli (chairs, faces, houses, and phase-scrambled controls). Discriminants for classifying one category of objects from other categories (e.g., faces vs. chairs and houses) will be referred to as category-specific discriminants because they rely on category-specific activity to discriminate between classes of objects. Discriminants that are effective at classifying objects from their phase-scrambled controls (e.g., chairs vs. phase-scrambled chairs) are termed object-control discriminants because they utilize activities that separate objects from their controls. In addition, a set of pairwise discriminants (e.g., faces vs. chairs), which also rely on category-specific activity, were evaluated to facilitate comparisons with the results from another classification analysis (Haxby et al., 2001).

RESULTS

In the following sections, we report results from the LDA of the Ishai et al. data,³ addressing these three questions: 1) How well can the fMRI data be used to classify the stimuli? 2) How much interdependence is there in the different groups of discriminants? 3) How do the discriminants interact spatially?

Performances of Linear Discriminants

Classification performances for the discriminants are shown in Table 1. The discriminants for classifying one object category versus the other two categories had a mean performance of 74.7% correct classification with chance performance at 50% accuracy. Classification for houses was measurably⁴ lower than the other two category-specific discriminants in both the delayed matching condition and the passive viewing condition. Mean performance for the pairwise discriminants was slightly higher at 76.3% accuracy. Discriminant performance can be interpreted as the degree of “uniqueness” between object categories. The lower performance in

distinguishing houses from the other two categories suggests that houses share more common activity with the other two classes of objects and therefore less dissociable. This is also reflected in the performance of pairwise discriminants with more accurate discrimination between chairs and faces than that of the two pairwise comparisons involving houses. Interestingly, the increased cognitive demands of the delayed matching condition did not result in any significant enhancement in the discriminant performance relative to passive viewing for both the category-specific and the pairwise discriminants, the two discriminants that rely on category-specific activity.

Performance for the object-control discriminants was notably higher than the category-specific discriminants. This result is expected given that the object-control discriminants can benefit from shared processes across categories of objects. General responses to objects compared to phase-scrambled controls, such as those

found in the lateral occipital areas (Malach et al., 1995), can contribute to the object-control discriminants. Object-control classification performance differed by less than 2% across categories, well within the 95% bounds on the error of these estimates. There was a large effect of task demand (delayed matching task vs. passive viewing) for all three categories of objects. This is contrasted with the lack of a task demand effect on the category-specific discriminants. Because the stimuli used in the original study were all well above threshold for detection, we do not believe that the subjects had difficulty distinguishing faces, chairs, and houses from their phase-scrambled controls in the passive viewing condition. However, if there were a homunculus sitting in these subjects' head and trying to decide whether the subject just saw an object or a control based on cortical activity patterns, then the delayed matching task certainly made the homunculus's job easier, as it did for the object-control discriminant. Curiously, the

Figure 2. LDA applied to fMRI data. There are two basic steps for the application of discriminant analysis to functional imaging data. (A) Data reduction. A necessary condition for discriminant analysis is that the number of predictors be equal to or less than the number of data points. Data reduction is performed in two steps. First, regions outside the brain and subcortical regions are masked and removed. Second, principal component analysis is used to further summarize the data into component processes. (B) Discriminant analysis. Principal component eigenvalues for individual time acquisitions are labeled by condition (i.e., chairs, faces, and houses) and projected onto an N dimensional space, with N equal to the number of components (2-D example shown). Within this space, a discriminant axis is derived and Bayesian classification is performed on the data projected on this axis.

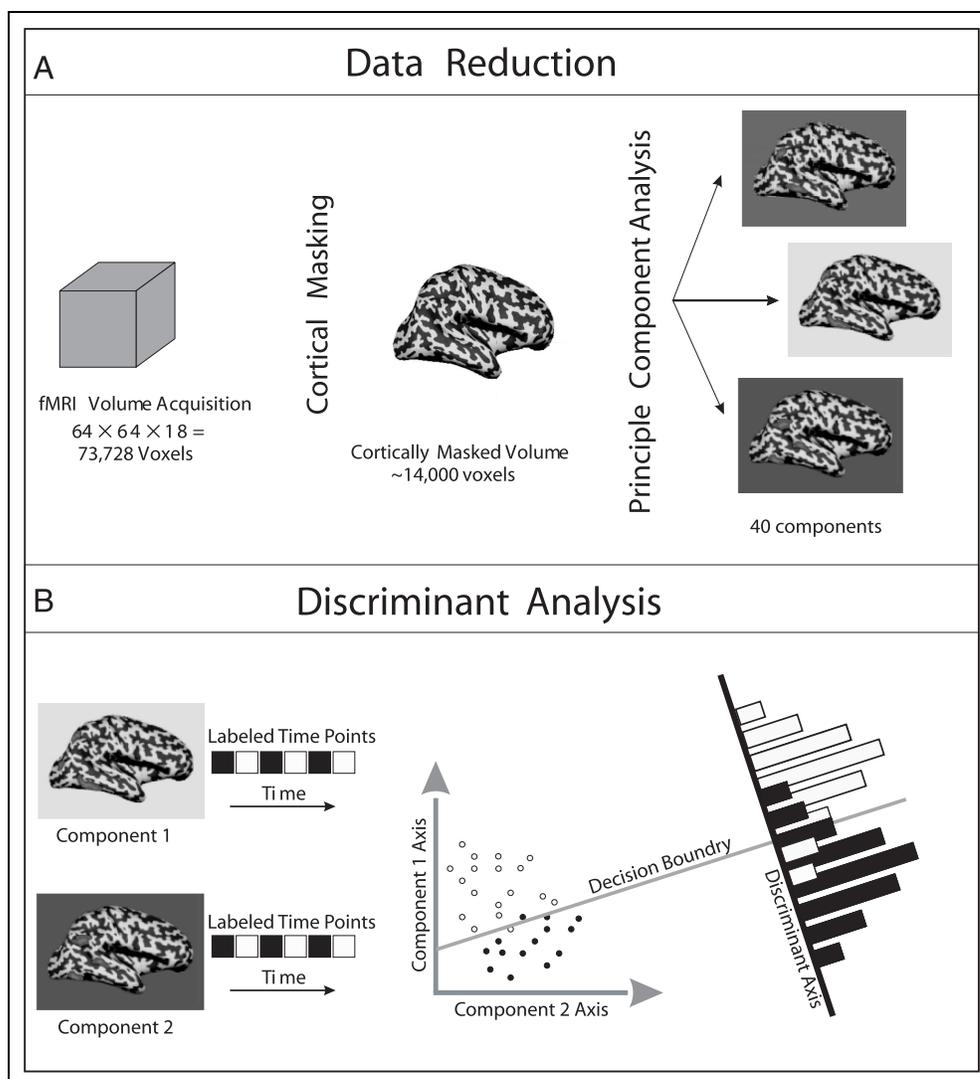


Table 1. Linear Discriminant Performance

<i>Category-Specific Discriminants</i>			
	<i>Chairs vs. Faces and Houses</i>	<i>Faces vs. Chairs and Faces</i>	<i>Houses vs. Chairs and Houses</i>
Delayed matching	75.1% (2.1)	76.6% (2.0)	70.8% (2.0)
Passive viewing	77.4% (2.1)	75.3% (2.0)	72.8% (2.0)
<i>Pairwise</i>			
	<i>Chairs vs. Faces</i>	<i>Chairs vs. Houses</i>	<i>Faces vs. Houses</i>
Delayed matching	80.3% (2.4)	73.2% (2.5)	76.3% (2.5)
Passive viewing	79.2% (2.4)	74.9% (2.4)	73.6% (2.4)
<i>Object-Control Discriminants</i>			
	<i>Chairs</i>	<i>Faces</i>	<i>Houses</i>
Delayed matching	98.7% (1.3)	97.4% (1.4)	99.3 (1.0)
Passive viewing	86.3% (2.2)	84.3% (2.3)	84.6 (2.3)

Performance for discriminants reported as percent correct classification (%). Standard errors of the estimate of classification performance are given in parentheses.

delayed matching task did not make the homunculus's job any easier in deciding whether the subject just saw a chair or a face or a house.

Linear Discriminant Knockouts

LDA identifies the patterns of activity that are important for object classification. It also provides an opportunity to examine how much of the information are common between different object categories. Here we introduce the novel idea of a discriminant knockout (Figure 3). LDA provides an axis in a multidimensional activity space, along which the two conditions can be best separated. What happens when all the information in the cortex along this axis is destroyed, effectively removing this axis? If this process is done recursively, one can remove all the information in the data that can potentially contribute to the discrimination of one category of stimuli from others. The expected result of such a manipulation would be the reduction in performance of the discriminant to the chance levels. More interesting to know is the effect of knocking out one discriminant on the performance of other discriminants, as they are not necessarily orthogonal to one another. For example, we can knock out activity in the data that supports the discrimination of faces against other categories of objects, creating a "virtual lesion," and then evaluate the effect of this manipulation on the performance of other discriminants.

Results for linear discriminant knockouts for both the passive viewing and delayed matching conditions are shown in Figure 3. The dependent measure is relative loss in performance, which is the change in performance

due to the knockout scaled by the baseline performance of the discriminant. Results for a discriminant knockout on the same discriminant represent the amount of information that needs to be removed to reduce the discriminant to chance levels (plots marked with stars). This magnitude provides a baseline that can be used to compare the effects of the other knockouts on the discriminant. The difference between knockouts within a discriminant and knockout of other object categories indicates the amount of unique activity between the two categories. For example, in the passive viewing condition, a chair category-specific knockout for the chair specific discriminant requires a 23% relative loss in accuracy to bring the discriminant to chance performance. In comparison, knocking out the face category-specific discriminant results in only 6% loss in the performance of chair specific discriminant. The ratio of these two numbers (6% vs. 23%) can be viewed as a measure of shared activity.

Each of the three category-specific knockouts reduced classification performance for chair, face, and house specific classifications. These interactions between the category-specific discriminants show that there are indeed aspects of the categorical representation, in terms of activity, that are shared across the three categories of objects. However, this sharing is not complete. The differences between the passive viewing and delayed matching conditions were minimal for category-specific discriminants, consistent with the previous result that increased task demands do not affect the discrimination performance of category-specific discriminants. Interestingly, making one type of objects undiscriminable from the other objects did not affect the discriminability of the

objects from its scrambled controls. More surprisingly, the reverse is true, too. For example, one can lesion the data to the extent that faces are no longer discriminable from scrambled faces, but faces can still be discriminated from chairs or houses. This result is interesting in that it suggests that activity in the cortex that enables the categorical discrimination of objects is independent of the information that separates objects from the phase-scrambled control stimulus.

Although object-control discriminants are highly independent of the object-specific discriminants, there seems to be a significant amount of shared information in detecting different types of objects from their respective controls. In all cases, knocking out one object-control discriminant had a significant effect on all the other object-control discriminants. The strong interaction between object-control discriminants reflects the significant amount of overlapping processes in object detection. This is in contrast with the weaker interactions

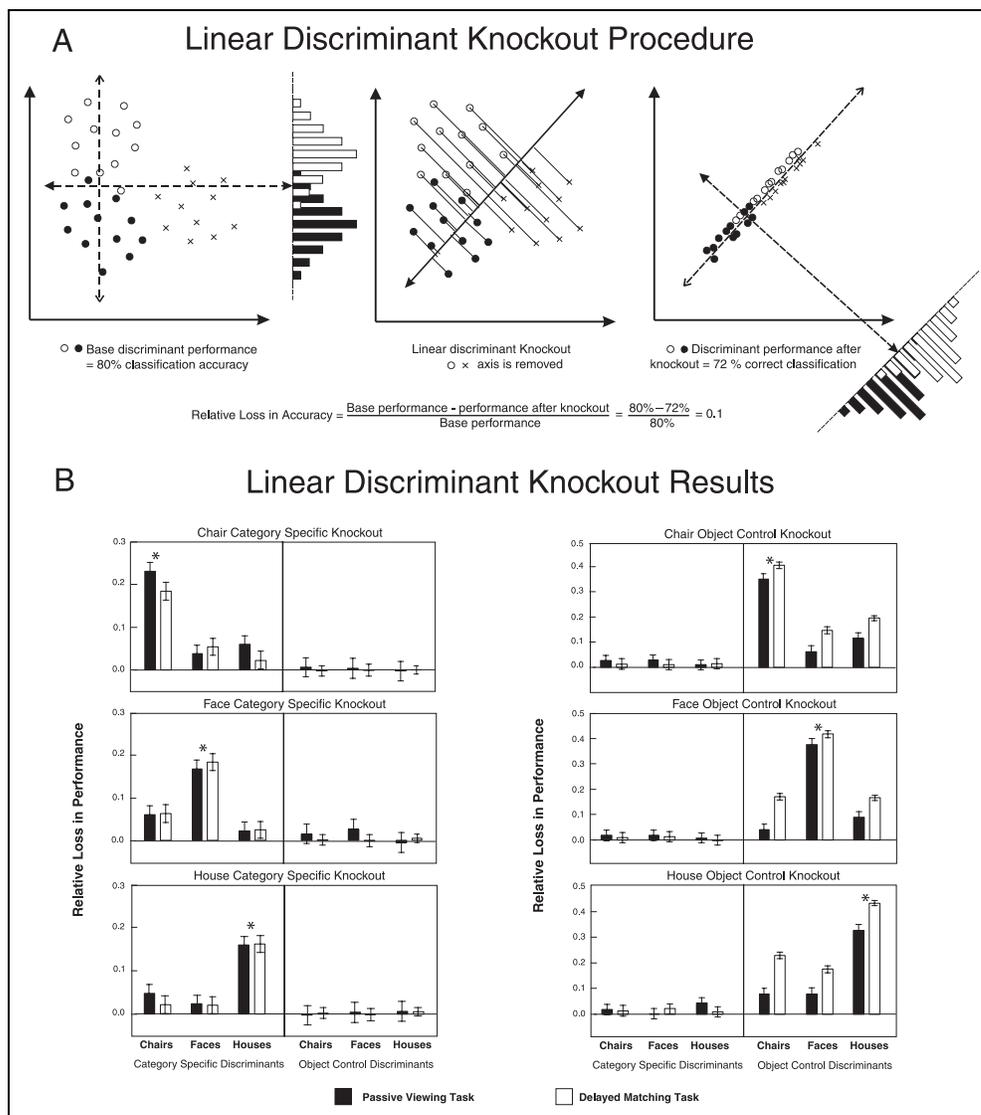
found between the category-specific knockouts. Taken together, this pattern of results seems to suggest that category-specific information is not an essential part of the process that distinguishes objects from nonobjects.

Spatial Aspects of Categorical Representations

Category-specific Representation in the Cortex

The projection of the category-specific discriminants onto the cortex reveal that the majority of strongly weighted voxels were found in the ventral temporal cortex, confirming the strong contribution of the ventral temporal cortex for representing object categorical information. Voxels with significant weights⁵ were distributed across the ventral temporal areas in relatively small clusters corresponding with previously identified category selective areas. Interestingly, while the voxels were distributed there appeared to be very little overlap in the significant voxels across categories.

Figure 3. Discriminant knockouts. (A) Linear discriminant knockout procedure and evaluation. Overlap in activity space is evaluated via linear discriminant knockouts. Three categories of stimuli activity (open circles, filled circles, and X's) are plotted in a two-dimensional space. Initially discriminant performance for classifying open versus closed circles is 80%. A linear discriminant knockout is then performed by removing the axis that separates open circles and X's. The original discriminant (open vs. closed circles) is then retested yielding 72% classification accuracy. The effect of linear discriminant knockout is evaluated in terms of relative loss of accuracy, which is the loss in performance scaled by the base performance of the discriminant. (B) Result of linear discriminant knockouts for object-specific and object-control discriminants for both the delayed matching and passive viewing conditions.

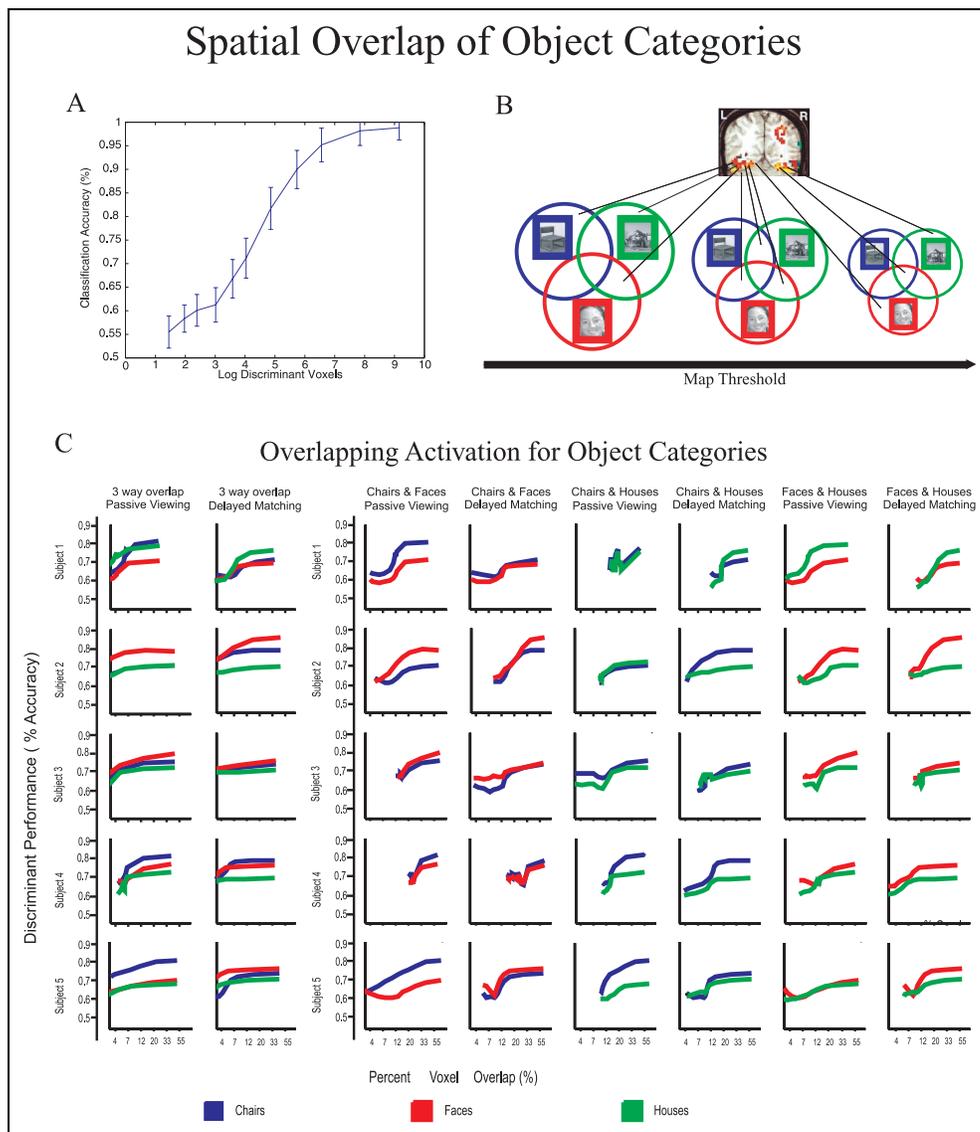


One problem with the aforementioned observation is that the amount of spatial overlap is a function of the threshold applied to the map: the two extreme examples being setting a very low threshold in which the whole brain is significant, thus there would be a large amount of overlap between categories; and setting a very high threshold in which significant voxels would be more likely to cluster in a modular arrangement. To avoid setting an arbitrary threshold, we investigated the relationship between the performance of discriminants and the amount of spatial overlap independent of threshold. The number of overlapping voxels and the discriminant performance are both a function of the threshold; thus, the common variable can bridge a direct relationship between classification performance and the amount of overlapping activation. Discriminant performance and the percentage of voxels overlapping, relative to the total number of significant voxels, were computed for small increments in threshold.⁶ Figure 4

shows the percentage of the number of active voxels participating in multiple discriminants as a function of the discriminant performance. The critical aspect in these plots is the loss of information as the amount of spatial overlap decreases. A decrease in performance coupled with a reduction in the percentage of overlapping voxels indicates the thresholding of voxels that are both important to the prediction and shared across categories. The first two rows in the figure show that the number of shared voxels critical to all three object categories discriminants is relatively small. There was some variation across subjects. Subjects 2 and 3 show only a minimal loss in performance as the amount of overlap drops (indicating minimal information in the shared voxels), while other subjects show sharp decreases in performance when the overlap fell below 4% and 12% (indicating critical information are carried in those shared voxels). The amount of overlap in voxels for pairwise object-specific discriminants was

Figure 4. Spatial overlap of category specific process.

(A) Discriminant performance shown as a function of the number of voxels participating in the prediction via thresholding. (B) Voxels are binned according to their participation in the discriminants. Seven bins, shown on the Venn diagram, were used to encompass all possibilities (chair only, face only, house only, chairs and faces, chairs and houses, faces and houses, and all three categories). (C) The relationship between discriminant performance and spatial overlap evaluated independent of threshold. Discriminant performance and the number of overlapping voxels are both a function of threshold (see A and B). The shared axis allows for the direct comparison of overlapping activation and classification performance. Discriminant performance for object categories, chairs (blue), faces (red), and houses (green) are shown as a function of the percentage of overlapping voxels. The percentage of overlapping voxels is calculated relative to the number of voxels surviving threshold. Thus, while increasing the threshold will reduce the total number of voxels considered, the percentage of overlapping voxels can go up or down depending on the type of voxels remaining after thresholding.



larger. Again, these thresholds did vary across subjects and comparisons, but the percentage of overlapping voxels was generally quite small.

Spatial Aspects of Object-Control Discriminants

Voxels contributing significantly to the object-control discriminants were found in occipital, temporal, and parietal lobes. Given that these discriminants essentially detect differences between objects and noise and are likely to encompass many cortical processes, a complete interpretation of them requires more constraints than available in this data set. Thus, we focus our attention on the differences between the passive viewing and delayed matching tasks that was shown earlier to have an important effect on this discrimination. Under the same threshold, the discriminants for the passive viewing condition had considerably fewer significant voxels than that in the delayed matching condition. Discriminants in the delayed matching condition both recruited new areas of the cortex and expanded regions identified by the passive viewing task. This is not surprising given the increased cognitive demands presumably would require additional processes and enhance processes already occurring in the passive viewing condition. This result, however, highlights the difficulty in interpreting fMRI data. For example, the use of an attention task will often result in more robust responses to the stimuli; however, the observed activation is a combined effect of stimulus and the task.

DISCUSSION

A well-designed functional imaging study provides a wealth of information. The data set submitted by Ishai et al. (1999, 2000) serves as an excellent example of this; as the group has already reported two analyses based on this data set, the current one represents yet another. One of the benefits of performing a reanalysis on a previously published data set is the opportunity to contrast the different analyses and compare the results.

In this reanalysis, we take seriously the idea that the analyses should explicitly look for “patterns of activity” that are meaningful, treating voxels as axes in a multi-dimensional activity space, rather than independently. In doing so, our analysis can more directly investigate hypotheses involving distributed and focal representations. The distributed hypothesis suggests that patterns of activation across the cortex can be used to support different kinds of information. The modular perspective argues that these patterns largely be confined to particular areas. Using LDA in activity space, we can simultaneously and empirically investigate the evidence for both kinds of hypothesis.

The current analysis focused on two types of linear discriminants: category-specific and object-control. First,

with category-specific discriminants, activity specific to object categories was isolated by contrasting one object class with the other two classes of objects. It is important to note, however, that the activity we define as category specific is limited by the number of categories. For example, a pairwise discriminant for two object categories cannot be identified as specific to either object category. In the current study, there are only three categories of objects. Additional research must be performed to further constrain the detection of object-specific activity. Second, with object-control discriminants, activity critical for detecting objects was determined by contrasting an object class and its phase-scrambled control. A key difference between these two types of discriminants is that shared activities across object categories will not contribute to the category-specific discriminants, but they could be important to the object-control discriminants. Our analyses found that these two types of processing evoke patterns of activity in the cortex that are relatively independent of one another, and are differentially influenced by the task demands.

Overlapping Representation between Different Object Categories

Category-specific discriminants can be used to find patterns of activity unique to a category of objects, and as such can be used to address the issue of how object categories are represented in the cortex. The ability to predict the stimulus presented to subjects based on derived category-specific discriminants shows that there are patterns of activity in the cortex unique to each of the three categories of objects that are linearly separable. The weak interactions found after the discriminant knockouts, however, reveal that only a small proportion of the categorical representation of these three objects is shared. Our analysis also found that the category-specific activity is largely confined to the specific regions with very little spatial overlap across categories. On the same issue, Ishai et al. (1999) and Haxby et al. (2001), among others, have demonstrated that object-selective regions do respond to nonpreferred stimuli and proposed the object form topology hypothesis to account for these results. The results of our analysis showing independent category-specific information, on the surface may seem to contradict their hypothesis. How can this be reconciled, especially given that both analyses were performed on the same data? The answer is that the two analyses are looking at two different aspects of the data. In our analysis, activities shared across object categories do not contribute to the object-specific discriminants, as it contains no useful information to differentiate between categories of objects. Ishai et al.’s analysis includes this activity and labels it as a component of the distributed representation. The interpretation of this activity is a question that we feel remains open and could be addressed in future research.

Object Detection Versus Object Category Classification

The object-control discriminants reflect the activities associated with the general processes of detecting an object. The behavior of object detection discriminants was very different from the category-specific discriminants. First, classification performance was much higher for object-control discriminants. The result is expected given that many areas are responsive to objects but not to scrambled controls (Grill-Spector et al., 1998; Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995), and all of the activity in these areas can contribute to the object-control discriminant. On the other hand, a relatively smaller number of voxels have a differential response for the three categories and are capable of contributing to the category-specific discriminants.

Interestingly, task demands and attention had very little effect on the performance of category-specific discriminants and also had negligible effect on the knockouts across the category-specific discriminants. Previous studies have shown that attention could enhance activity in object responsive regions (O'Craven, Downing, & Kanwisher, 1999), and enhancement of response magnitude from the matching task is evident in the current data set (Ishai et al., 2000). The result of our reanalysis suggests that the delayed matching task simply boosted object-related activities, but the benefit was not specific to any particular object category. Given this, it would seem that care must be taken when interpreting the effects of task. While attentional tasks may change activation magnitude, the pattern of activation relevant to the stimulus may not change. Studies that assess activation significance using the baseline provided by a control stimulus may also suffer from the possibility that the effect of task is predominantly due to changes in the processing of the control rather than in the test stimuli.

Differences between Linear Discriminants and Correlational Analysis

The classification performances of the derived linear discriminants were found to be robust, but substantially lower than that found by Haxby et al. (2001) using a correlation classification method. The performance for our pairwise discriminants ranged from 73.2% to 83% in the delayed matching condition, while Haxby et al. (2001) reported near perfect classification for these three classes of objects. This difference is to be expected for several reasons. First, our analysis attempts a much more difficult prediction: Decisions were made for each individual time acquisition rather than for regression coefficients derived from half the total acquisitions. For block designs like Haxby et al.'s, least square regression coefficients are category means across time for each voxel (see the Appendix). In Haxby et al.'s analysis, the prediction performance is limited by the variability of these means, rather than the variability of individual

time acquisitions. If we replaced individual time points with means across half the points (41 time points) in a group, and we assume for the sake of argument that all of these time points are independent, then a d' measure of performance would increase by a factor of $\sqrt{41}$. For all of our comparisons, such an increase would put performance in terms of percent correct at ceiling. Adjusting for the fact that the time points are not independent predicts a smaller gain; however, it would require averaging only four independent time points to drive all our performance values over 95%. Thus, the near-perfect performance reported in Haxby et al. (2001) primarily owes to the statistical advantage of classifying averages rather than individual acquisitions.

Although averaging can potentially explain the differences in performance, there exists an interesting relationship between Haxby et al.'s (2001) analysis and LDA (see the Appendix). We show that Haxby et al.'s analysis classifies the group means across half the data by projecting onto an axis given by the differences between normalized class means computed on the other half of the data. When the class means are all roughly equal (true for the data we analyzed) this procedure simply uses the differences between group means as an axis. LDA uses the differences in means as a discriminant, but it reweights the voxel axes using the data covariance to reduce the contribution of voxel axes with high variability. Because Haxby et al.'s analysis does not factor out voxels with high variability and because they classify means, their analysis confers predictive significance to voxels that are too variable to support classification in our analysis. Thus, we see one of the key advantages of LDA is that it provides a method for computing and incorporating voxel reliability in the predictions.

Benefits of LDA Analysis

Although more powerful classification methods exist, Fisher's LDA has the advantages of simplicity and ease of interpretation. More powerful classification methods involve creating (potentially complicated) regions of activity space associated with a stimulus category. However, regions of activity are difficult to visualize and interpret, requiring a different activity map for each point in the region. In contrast, Fisher's LDA provides a single axis of voxel weights that provide a direct measure of the voxel's contribution to the classification. We believe this gain in interpretability offsets its suboptimality. (Fisher's LDA is only guaranteed optimal when the distribution of brain responses to a category can be modeled as a multi-dimensional Gaussian and when the response covariances for each category are equal, although it can be optimal or almost optimal for other cases.)

General Discussion

In summary, the current article used robust PCA, Fisher's LDA, and a Bayesian classifier to address the question of

categorical object representations. We found that activity in the cortex that defines object categories relative to other categories of objects is independent of activity that defines an object category relative to a control stimulus. The attentional benefit of using a delayed matching task acts principally on nonspecific object processes. We found modest amounts of overlap, both in terms of activity and spatial overlap, between object-specific activity; however, the magnitude of the overlap seems to be insufficient to support the continuous topological arrangement that has been proposed (Ishai et al., 1999).

In addition, we sought to highlight the benefits of this new approach in examining functional imaging data. One key idea in our analysis is that activation patterns can be decomposed into components, some of which are more informative with respect to particular stimulus categories, and LDA can be used to extract these informative components. A second important concept is that of activity space in which voxels, not acting alone and without spatial constraint, can participate in concert for representing and processing of information. Given the highly interconnected nature of the brain, it is our belief that this brain-based analysis method holds promise for future functional imaging research.

METHODS

Data set

Data were acquired from the fMRIDC database (accession no. 2-2000-1113D) (Ishai et al., 2000). A summary of the relevant information from the original study will be presented below to facilitate interpretation of the results (for detailed methods see Ishai et al., 2000).

Subjects

Six right-handed subjects with normal vision participated in the experiment. One subject's data (Subject 9) had compression errors that resulted in a significant loss of data. The remaining five subjects represent the data subjected to the analysis presented in this report.

Stimuli

Stimuli were gray-scale photographs of faces, houses, and chairs presented on a gray background. Control stimuli for the photographs were phase-scrambled images of the respective stimuli that preserved spatial frequency information.

Experimental Procedure

Subjects performed one of two tasks, a delayed matching and passive viewing task. Stimuli presented during passive viewing were presented at a rate of two photographs per second. In the delayed matching task, target

stimuli were presented for 1.5 sec, then after a delay of 0.5 sec two alternative stimuli were presented for 2 sec. Subjects made a keypress response to indicate their choice. Task difficulty was equated across the object categories, which was confirmed both during functional, imaging acquisition and psychophysically. The subjects task, delayed matching and passive viewing alternated between runs. Each run consisted of alternating 21-sec block class of a stimulus category (houses, faces, chairs) followed by a 21-sec block of the respective phase-scrambled control. Two blocks of each of the three classes of objects were presented in each run and were counterbalanced across a total of six runs for each of the experimental conditions.

Data Acquisition

A 1.5-T General Electric Signa scanner with a whole head RFD coil was used for data acquisition. Eighteen contiguous coronal slices were obtained during functional acquisitions (TR = 3 sec, TE = 40 msec, FOV = 20 cm, 64 × 64 matrix, voxel size 3.125 × 3.125 × 5 mm). For each subject, a high-resolution whole head anatomical was acquired in a separate session (124, 1.5-mm thick sagittal slices, TR = 13.9, TE = 5.3, FOV = 24 cm, 256 × 256 matrix).

Data Preprocessing by Ishai et al.

The acquired data set were EPI scan volumes registered with an iterative method (Woods, Mazziotta, & Cherry, 1993), spatially smoothed in plane with a gaussian filter (FWHM was 3.75 mm along the *x*- and *y*-axis), and ratio normalized to the same mean global intensity.

Additional Data Preprocessing

To prepare the data for LDA two additional preprocessing steps were taken. A voxel-by-voxel deconvolution operation was performed on individual run acquisitions to remove the effects of hemodynamic blurring and facilitate labeling of time points. The deconvolution filter was given by a gamma function:

$$b(t) = \frac{(t/\tau)^{n-1}}{\tau\Gamma(n-1)} \exp(-t/\tau)$$

using parameter values ($n = 3$, $\tau = 1.25$) in good agreement with empirical measures (Boynton et al., 1996). It is important to note that this operation is certain to be imperfect given the complexities of the relationship between the MR signal and the neural response; thus, local correlations between adjacent time acquisitions may still exist. Each of the runs was then normalized and the voxel-by-voxel mean was removed. The six experimental runs, for each of the experimental tasks, were then appended into a single volume time

course. Two data-reduction methods were used to reduce the number of variables below the number of recorded time samples. Initially, the dimensionality of the data equals the number of voxels. Thus, each time acquisition has 73,728 ($64 \times 64 \times 18$) dimensions. Voxels in subcortical regions and areas outside the brain were masked and excluded from further analysis, which removed approximately 80% of the total voxels. Next, principle component analysis was performed using singular value decomposition (SVD) to further reduce the dimensionality of the data into component variables less than the 84 time samples (7 acquisitions per block \times 2 blocks per run \times 6 runs = 84 acquisitions per object category). Friston et al. (1993) have suggested retention of all components whose variance is greater than the mean variance. However, we found this procedure retained too many components for this data set, yielding overfit and lack of generalization (see also Ardekani et al., 1998). We used 40 components, approximately half the number of time samples for a class of objects, which was sufficient to capture 80–85% of the total variance in the data. The use of additional components had negligible impact on predictive performance of our classifiers.

Linear Discriminant Analysis

Each time acquisition was given one of six labels according to its corresponding stimulus type (chairs, faces, houses, chairs–noise, faces–noise, houses–noise). Let \mathbf{X} denote the data matrix, with rows given by the masked voxels and columns representing the time samples. As noted above, we performed dimensionality reduction by projecting \mathbf{X} onto the principal components basis \mathbf{B} yielding a reduced data matrix:

$$\mathbf{Y} = \mathbf{B} \mathbf{X}$$

LDA produces a set of orthogonal axes in the data space termed canonical variates or discriminants that best separate class means given the within-class covariance.

The fit is only meaningful if it can be validated against overfit. We validated the fits by measuring the predictive performance of the linear classifier using the 632+ bootstrap (Efron et al., 1997), a state-of-the-art, leave-one-out, resampling procedure. In this procedure, each time point is removed from the data set. The discriminant is fit using a random subset of the remaining data, and the time point of interest is then subsequently classified. By repeating this procedure for numerous random subsets and all the time points, we obtain a measure of the generalizability of the classifier that mitigates against overfitting. It is important to note that individual time acquisitions are tested using data from within the same block, data from other blocks within the same run, and data from separate run acquisitions, which results in 83% of the data coming from separate run acquisitions and the vast majority coming from different blocks

(94%). This is important in that the effect of temporal correlations remaining after the deconvolution operation are minimized because the majority of the information used to fit the discriminant comes from acquisitions in separate runs.

The number of distinct discriminants is one less than the number of classes compared. For two class comparisons, the discriminant L has a particularly simple form:

$$L = -(\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2)$$

where μ_i and Σ_i are the mean and covariance of the i th group. Thus, the best direction to distinguish two classes is the direction along the difference between the means, reweighted by the inverse sum of the within-class covariances. Because mean and covariance estimates are extremely susceptible to contamination by outliers, we computed robust estimates of the mean and covariance using Rousseeuw & Van Driessen's (1999) minimum covariance determinant method. The robust estimates resulted in approximately 10% of the time points in each class being rejected as outliers.

For each of the conditions in the experiment, delayed matching and passive viewing discriminants were derived to identify patterns of activation significant to objects classes. Comparisons included testing each object class against the respective phase-scrambled control, each object class versus the other two object classes (houses vs. faces and chairs, faces vs. houses and chairs, and chairs vs. houses and faces), and pairwise contrasts containing all possible permutations of the three object categories.

Prediction Error

Although LDA provides axes that best distinguish classes, it does not directly provide a measure of the classification performance (Kustra & Strother, 2001). However, if we assume that the data are well described by a multivariate Gaussian distribution,⁷ we can perform classification within a Bayesian framework.

Given we project the data onto the linear discriminant vector \mathbf{L} , the data given the class-conditional distribution of the projected data are given by:

$$P(z_i | i, \mu_i, \Sigma_i) = \frac{1}{C} \exp\left(-\frac{1}{2} z_i^T (\mathbf{L}^T \Sigma_i \mathbf{L})^{-1} z_i\right)$$

where C is a normalization constant and $z_i = \mathbf{L}^T (y_i - \mu_i)$. Classification response is then based on the ratio of posterior distributions

$$r = \frac{P(z_i | i, \mu_i, \Sigma_i) \pi_i}{P(z_j | j, \mu_j, \Sigma_j) \pi_j}$$

where π_i represents the prior probability of the class, and class i is chosen over class j when r is greater than one. Because we used data to derive our discriminant, it is important to have a procedure to test the data that is

not part of the training set, to test predictability. We estimated the prediction error using the 632+ bootstrap procedure (Efron & Tibshirani, 1997), in which testing is performed on a set of data points excluded from the training via resampling. Prediction (classification) accuracy is reported as the ratio of the correct classification of labeled time points to the total number of time points tested.

Discriminant Knockout

In order to test how much of the discrimination for one pair of classes is affected by the removal of the directions required to discriminate another pair, we systematically “knocked out” via projection the best discriminants for a pair of classes until prediction performance was within 5% of chance. Knockouts were performed via a simple matrix, constructed as follows. An orthogonal basis was constructed from the discriminant axis using SVD:

$$\mathbf{L} = \mathbf{U}\mathbf{S}\mathbf{V}^t$$

If \mathbf{L} is an N length column vector, the matrix \mathbf{U} is an N by N orthonormal rotation matrix that has \mathbf{L} as its first column. We construct a new matrix \mathbf{U}_d by replacing the first column with zeros. Then the knockout matrix \mathbf{K}_L is given by:

$$\mathbf{K}_L = \mathbf{U}^t \mathbf{U}_d$$

Multiplying y by the reduced data matrix \mathbf{K}_L resulted in a new data set in which the discriminant axis is removed via projection.

Spatial Overlap

Spatial overlap between discriminants was computed as the ratio of the intersection and the union of nonzero voxels that survive a thresholding procedure. The voxels v_m significant for a discriminant were computed as:

$$v_m = \left\{ v \left| \frac{d(v)}{\max(d)} > \gamma \right. \right\},$$

where $d = \mathbf{B}\mathbf{L}$ is the discriminant \mathbf{L} transformed back into voxel space and γ is the threshold. In words, the discriminants were scaled between -1 and 1 and the voxels greater than the threshold were selected. Using the same threshold for all the discriminants, the proportion overlap is computed as the ratio of the number of surviving voxels in the intersection to the union of the discriminants. Thus, for discriminants j and k , the proportion overlap p_v is given by:

$$p_v = \frac{\#(v_m^j \cap v_m^k)}{\#(v_m^j \cup v_m^k)}$$

In addition, for each γ the prediction accuracy could be computed by thresholding the discriminant (setting to

zero all voxels whose absolute value is less than the threshold) and recomputing the prediction accuracy for each value. The overlap plots were generated by plotting proportion overlap against the recomputed prediction accuracy for corresponding values of the threshold as the threshold varied from 0 to 1.

APPENDIX: RELATIONS BETWEEN THE ANALYSIS IN HAXBY ET AL. (2001) AND LDA

Let \mathbf{X} denote the data matrix as above. For a voxel-wise regression analysis, a model of the form

$$\mathbf{X}_n = \mathbf{G}\beta_n$$

is fit, where \mathbf{X}_n is the n th voxel's time series, \mathbf{G} is a matrix of regressors in which each column is a vector that models the expected temporal response to one of the stimuli, and β is the vector of regression coefficients. The least square estimate of β_n is given by:

$$\beta_n = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{X}_n$$

When the regressors in \mathbf{G} are nonoverlapping blocks,⁸ then the regression coefficients are weighted averages of the time points associated with each class. For example, in an experiment with two stimulus classes, stimulus class \mathbf{a} presented in the first half of $2n$ time points and stimulus \mathbf{b} in the second half, then the regression matrix would look like:

$$\mathbf{G} = \begin{pmatrix} a_1 & 0 \\ a_2 & 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & b_{n-1} \\ 0 & b_n \end{pmatrix} = \begin{pmatrix} \mathbf{a} & 0 \\ 0 & \mathbf{b} \end{pmatrix}$$

where \mathbf{a} and \mathbf{b} are length n vectors of ones. In this case, the matrix pseudoinverse $(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$ reduces to

$$(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T = \begin{pmatrix} \frac{\mathbf{a}^T}{\|\mathbf{a}\|^2} & 0 \\ 0 & \frac{\mathbf{b}^T}{\|\mathbf{b}\|^2} \end{pmatrix} = \begin{pmatrix} w_a & 0 \\ 0 & w_b \end{pmatrix}$$

where $w_a = (1/n, 1/n, \dots, 1/n, 0, \dots, 0)$ and $w_b = (0, \dots, 0, 1/n, 1/n, \dots, 1/n)$ when \mathbf{a} and \mathbf{b} are vectors of ones.

Thus, for nonoverlapping blocks,

$$\beta_n^a = (w_a 0) \cdot \mathbf{X}_n = w_a \cdot \mathbf{X}_n^a = \mu_n^a$$

where the last equality only holds for $w_a = (1/n, 1/n, \dots, 1/n, 0, \dots, 0)$. However, the result is that the regression coefficient for the n th voxel for a stimulus class \mathbf{a} is just the (weighted) average of the time points in which stimulus class \mathbf{a} was presented.

The correlation-based prediction used by Haxby et al. (2001) can be related to the LDA procedure. Specifically, the correlation measure they use can be viewed as the expectation of a discriminant that is given by the difference between normalized class means. The correlation measure, where angle brackets denote an inner product across voxels, the spatial

correlation between the regression coefficients for classes a and b is given by:

$$\rho_{ab} = \frac{\langle \beta_a, \beta_b \rangle}{\sqrt{\langle \beta_a, \beta_a \rangle \langle \beta_b, \beta_b \rangle}} = \frac{\langle \mu_a, \mu_b \rangle}{\| \mu_a \| \cdot \| \mu_b \|}$$

For pairwise comparisons, their classification procedure involves comparing the correlations between classes a and b versus within classes on odd and even runs. Specifically, they compute four correlations: $\rho_{aa}^{oe}, \rho_{ab}^{oe}, \rho_{ab}^{eo}, \rho_{bb}^{oe}$, where the subscripts denote the first and second classes, and the superscripts whether the regression coefficients were computed from odd or even runs (o or e). For example, ρ_{ab}^{oe} is the correlation coefficient between regression coefficients computed for class a from odd runs, and class b from even runs. They consider four comparisons, $\rho_{aa}^{oe} > \rho_{ab}^{oe}, \rho_{aa}^{oe} > \rho_{ab}^{eo}, \rho_{ab}^{eo} < \rho_{bb}^{oe}, \rho_{ab}^{oe} < \rho_{bb}^{oe}$, and the predictive performance for each subject was computed as the proportion of these pairwise comparisons that had the relationship indicated by the inequality signs. Each one of these comparisons is equivalent to comparing whether the difference between the correlations is greater or less than zero:

$$\begin{aligned} \rho_{aa}^{oe} > \rho_{ab}^{oe} &= \rho_{aa}^{oe} - \rho_{ab}^{oe} > 0 \\ &= \left\langle \frac{\mu_a^o}{\| \mu_a^o \|}, \frac{\mu_a^e}{\| \mu_a^e \|} \right\rangle - \left\langle \frac{\mu_a^o}{\| \mu_a^o \|}, \frac{\mu_b^e}{\| \mu_b^e \|} \right\rangle > 0 \\ &= \left\langle \frac{\mu_a^o}{\| \mu_a^o \|}, \frac{\mu_a^e}{\| \mu_a^e \|} - \frac{\mu_b^e}{\| \mu_b^e \|} \right\rangle > 0 \\ &= \left\langle \frac{\mu_a^o}{\| \mu_a^o \|}, w_{ab} \right\rangle > 0 \\ &= \frac{1}{\| \mu_a^o \|} E_t [\langle w_{ab}, \mathbf{X}_a^o \rangle] > 0 \end{aligned}$$

where E_t denotes a temporal average. When the class means are equal in magnitude, the normalization constants factor out and the procedure is equivalent to using the difference between class means (w_{ab}) as a discriminant axis, followed by taking a temporal average. For the data set reanalyzed in this paper, the group mean magnitudes were almost equal. In addition, the classification step is a special case of the Bayesian classifier assuming equal class probabilities and class covariances.

Acknowledgments

Supported by the National Institute of Health and Human Development (HD-07151), the Center for Cognitive Sciences at the University of Minnesota, and McKnight-Land Grant Award from the University of Minnesota, and an award from the James S. McDonnell Foundation.

Reprint requests should be sent to Thomas Carlson, Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA, or via e-mail: tom@gandalf.psych.umn.edu.

Notes

1. Least square regression coefficients were determined via the pseudoinverse of the matrix of regressors. When the regressors are nonoverlapping temporal functions, which is true of most block fMRI designs, then this pseudoinverse solution is a weighted average of the time points where the

weights are produced by the temporal filter used to model the hemodynamics. The weighted average reduces to a simple average when the regressors are box functions. See the Appendix.

2. For those who are not familiar with linear discriminant analysis, a good introduction can be found in Duda, Hart, and Stork (2001).

3. A brief description of the experimental paradigm, data collection, and preprocessing procedures can be found in the Methods section. However, since this is a reanalysis of published data, readers are referred to the original paper for more detailed descriptions (Ishai et al., 2000).

4. Assessed by comparing overlap in 95% confidence intervals.

5. Assessed via thresholding, described below.

6. Spatial maps were scaled with minimum threshold of 0 and a maximum threshold of 1. Discriminant maps, using the aforementioned scale, were computed for 0.03 increments in threshold.

7. Data scatter for each class was well approximated by a multivariate Gaussian distribution. Deviations from a Gaussian distribution were tested using the method described in Rousseeuw & Van Driessen (1999), which assesses the existence of a linear relationship between robust and regular Mahalanobis distances of individual samples from the mean. The average correlation coefficient of this linear relationship was 0.897 with a standard deviation of 0.029.

8. Block overlap makes results in the overlapping time points to contribute to several categories. For a small number of overlapping points, the results for nonoverlapping blocks are still approximately valid.

REFERENCES

- Ardekani, B. A., Strother, S. C., Anderson, J. R., Law, I., Paulson, O. B., Kanno, I., & Rottenberg, D. A. (1998). On the detection of activation patterns using principal components analysis. In R. E. Carson, M. E. Daube-Witherspoon, & P. Herscovitch (Eds.), *Quantitative functional brain imaging with positron emission tomography* (pp. 253–257). San Diego, CA: Academic Press.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of fMRI in human V1. *Journal of Neuroscience*, *16*, 4207–4221.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification and scene analysis*. New York: Wiley.
- Efron, B., & Tibshirani, R. J. (1997). Improvements on cross-validation: The 632+ bootstrap method. *Journal of American Statistical Association*, *92*, 548–560.
- Epstein, R., & Kanwisher, N. (1999). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Friston, K. J., Firth, C., Liddle, P., & Frackowiak, R. (1993). “Functional connectivity: The principal component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism*, *13*, 5–14.
- Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., & Turner, R. (1995). Analysis of fMRI Time Series Revisited. *Neuroimage*, *2*, 45–53.
- Fodor, J. (1983). *Modularity of the mind*. Cambridge: MIT Press.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*, 568–573.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 1409–1422.

- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, *6*, 316–328.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in the ventral temporal cortex. *Science*, *293*, 2425–2430.
- Hinton, G. E., & Anderson, J. A. (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.
- Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (1999). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, *12 Suppl. 2*, 35–51.
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (2000). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences, U.S.A.*, *96*, 9379–9384.
- Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for the perception of faces. *Journal of Neuroscience*, *17*, 4302–4311.
- Kustra, R., & Strother, S. C. (2001). Penalized discriminant analysis of [15O]water PET brain images with prediction error selection of smoothing and regularization hyperparameters. *IEEE Transactions on Medical Imaging*, *20*, 376–387.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in the human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *92*, 8135–8139.
- O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.
- Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*, 212–223.
- Rumelhart, J. L., & McClelland (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1 and 2*. Cambridge: MIT Press.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.
- Tanaka, K. (1997). Mechanisms of visual object recognition: Monkey and human studies. *Current Opinion in Neurobiology*, *7*, 523–429.
- Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, *272*, 1665–1668.
- Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI–PET registration with an automated algorithm. *Journal of Computer-Assisted Tomography*, *17*, 536–546.