How Distributed Is Visual Category Information in Human Occipito-Temporal Cortex? An fMRI Study

Mona Spiridon[1] and Nancy Kanwisher[1]
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology, NE20-464
Cambridge, Massachusetts 02139

Summary

We used fMRI to study the distribution of object category information in the ventral visual pathway. Extending the findings of Haxby et al. (2001), we find that categories of stimuli can be distinguished by the pattern of activation they elicit across this entire pathway, even when the stimuli within a category differ in viewpoint, exemplar, or image format. However, regions within the ventral visual pathway are neither interchangeable nor equipotential. Although the FFA and PPA permit excellent discrimination between preferred versus nonpreferred stimuli (e.g., faces-bottles and houses-bottles, respectively), we find that neither region alone permits accurate discrimination between pairs of nonpreferred stimuli (e.g., bottles-shoes). These findings indicate that the ventral visual pathway is not homogeneous, but contains some regions (including FFA and PPA) that are primarily involved in the analysis of a single class of stimulus.

Introduction

Central to any theory of visual object recognition is the question of how objects are internally represented. Here we use fMRI to address one aspect of this question: are some classes of objects represented in localized regions of extrastriate cortex specialized for representing primarily that class of object (Downing et al., 2001; Epstein and Kanwisher, 1998; Kanwisher et al., 1997; McCarthy et al., 1997; Puce et al., 1996), or are all objects represented in distributed neural codes that span much of the occipitotemporal visual pathway (Haxby et al., 2001; Ishai et al., 1999)?

fMRI research emerging from a number of labs has described focal regions of occipitotemporal cortex that respond selectively to certain categories of objects: the fusiform face area (FFA) for faces (Halgren et al., 1999; Kanwisher et al., 1997; McCarthy et al., 1997; Puce et al., 1996), the parahippocampal place area (PPA) for places (Aguirre et al., 1998; Epstein et al., 1999; Epstein and Kanwisher, 1998; Haxby et al., 1994), and the extrastriate body area (EBA) for human bodies (Downing et al., 2001). The definition of selectivity used here, adapted from the neurophysiology literature (Tovee et al., 1993), is that a region of cortex must respond at least twice as strongly to "preferred" as to "nonpreferred" stimuli. Implicit in this definition is the fact that each of the category-selective regions of cortex produces some positive response to nonpreferred stimuli, compared to a low level baseline such as a fixation point in an otherwise blank screen.

The existence of partial responses to nonpreferred stimuli is open to two different interpretations. According to one, the partial response to a shoe in the FFA (for example) encodes information about the shoe and forms part of the neural representation of the shoe. On this view, the information about shoes would be encoded across much of extrastriate visual cortex, including the FFA and other regions of cortex that respond selectively to things other than shoes.
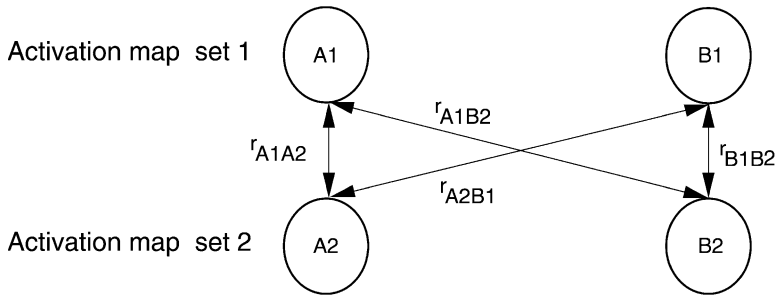
The alternative interpretation is that the response to a shoe in the FFA reflects the automatic but epiphenomenal engagement of this system by any visual stimulus, with the resulting weak response not forming part of the representation of the shoe. That is, even a specialized face processor may have no way to "know" that a given input is not a face, other than to try processing it as such. Thus, the mere existence of partial responses to nonpreferred stimuli in a category-selective region of cortex does not guarantee that these responses encode any information about the category of those stimuli, or that any such information forms a critical part of the representation of those stimuli.

In attempting to determine which of these interpretations is correct, a great deal is at stake. If nonpreferred responses play an important functional role in vision, then the implications will extend far beyond the FFA and PPA, to encompass many if not most prior studies in visual neuroscience. It has been a standard assumption in both neuroimaging and neurophysiology that the best way to determine the function of a neuron or cortical region is by discovering the stimulus that elicits the highest response from it. If it turns out that nonpreferred responses carry as much information as preferred responses, that will call into question both the standard techniques and many of the prior results in the field of visual neuroscience.

How can we determine whether responses to nonpreferred stimuli form part of the representation of those stimuli (as in the first account), or whether these responses are epiphenomenal and do not form part of the representation (as in the second account)? At a bare minimum, the representational account requires that nonpreferred responses contain information that could in principle help discriminate between nonpreferred stimuli. This hypothesis is the focus of the present investigation, and of the prior study by Haxby et al. (2001) that motivated it. Another critical requirement of the representational account (considered further in the Discussion) is that the discriminative information must actually get used when the subject makes category discriminations.

Haxby et al. (2001) asked whether nonpreferred responses across human occipitotemporal cortex carry information about the category of object seen. Their strategy was as follows. Each subject was scanned with fMRI while he/she viewed eight different categories of stimuli. The resulting data from each subject were split in half, generating an activation map for each of the

[1]Correspondence: spiridon@mit.edu (M.S.), ngk@mit.edu (N.K.)

Figure 1. Calculation of the Percentage of Correct Discrimination for a Given Pair of Stimulus Categories A and B

Each circle represents an activation map for a given category based on half the data collected for that category and subject. Thus A1 is the activation map for category A based on half the data and A2 is the activation map based on the other half. *r* represents the correlation coefficient between two activation maps. The percent correct discrimination between categories A and B is 100% if all of the following inequalities are correct: $r_{A1A2} > r_{A1B2}$, $r_{A1A2} > r_{A2B1}$, $r_{B1B2} > r_{A2B1}$, $r_{B1B2} > r_{A1B2}$. If only two or three of these relations are correct, the percent correct discrimination drops to 75% or 50%, respectively. The percent correct discrimination for all seven categories is the average of the percent correct discrimination over 84 pairwise comparisons.

eight stimulus categories from each of the halves of the data (i.e., 16 activation maps). Next, one set of activation maps (from one half of the data) was used to predict which category produced each of the eight activation maps generated from the other half of the data. This analysis was conducted by measuring the similarity (i.e., correlation) between each activation map in one set to each of the eight activation maps in the other set (and vice versa). The measure of the discrimination accuracy was based on all pairwise comparisons between identical and nonidentical stimulus categories (for more details, see Figure 1).

Several important findings emerged from this study. First, activation maps could be quite accurately categorized based on the activation maps from the other half of the data, demonstrating that the patterns of activation for each category were highly replicable. Second, and more importantly, information about which stimulus category produced a given activation map could be determined even when the region of cortex that responded maximally to that category was omitted from the analysis. This led Haxby et al. (2001) to suggest that cortical regions maximally selective to a given category may not form a necessary part of the representation of that category. Finally, even when only the region that responded maximally to a given category was included in the analysis, performance in determining which of the nonpreferred categories had been presented was still well above chance. These results indicate that nonpreferred responses across occipitotemporal cortex are systematic and replicable, and contain information about the category of object seen.

However, the Haxby et al. (2001) study leaves several critical questions unanswered, which we address here. First, is the information present in the pattern of cortical activation abstract enough to permit generalization to new images of members of the same category? In the original Haxby et al. (2001) study, the same set of grayscale stimulus images was used to derive both activation maps for each category. Therefore, the high performance on category discrimination might simply indicate that fMRI is a reliable technique producing similar responses when viewing identical image sets. To establish that each category produces a specific activation pattern in occipito-temporal cortex that does not simply reflect similarities in low level features of those stimuli, we tested whether the results of Haxby et al. (2001)

would hold when different stimuli from the same category are compared. Haxby et al. (2001) argued against a role for low level features in their findings by showing similar results when one set of stimuli consisted of photographs, and the other consisted of line drawings. However, this analysis used only three categories (faces, houses, chairs), and these are among the most distinctive in terms of their cortical response patterns, so it is not clear how this conclusion will generalize to other categories. Here we collected from each subject in a single scanning session independent data sets in which the subjects viewed stimuli from each category that differed in category exemplar, viewpoint, and image format (photographs versus line drawings). Our design enabled us to test the abstractness of the category information, by measuring categorization performance based on fMRI data sets collected from stimuli that differed in viewpoint, exemplar, or stimulus format (Figure 2).

A second question addressed in the present study was whether any categories of objects are "special" in the degree to which information about these categories is concentrated within focal regions of cortex. The main analyses of Haxby et al. (2001) treat the different kinds of category discriminations as equivalent. However, considerable evidence suggests that information about faces and places (including houses) may be represented in a more localized fashion across cortex than other stimuli. Here we addressed these questions by running separate analyses on discriminations involving faces and houses, and also by running analyses on cortical regions selectively responsive to faces and houses.

Our results replicate some aspects of the findings of Haxby et al. (2001), but nonetheless show that faces and places are "special" in three important respects. First, discriminations involving faces and houses are more accurate than discriminations involving other categories. Second, focal regions of cortex that respond selectively to faces (the FFA) and places (the PPA) are category specific not only in their mean response but also in the sense that the profile of response across the voxels within each region contributes little or no information to discriminations between nonpreferred stimuli. Third, we found no evidence for a focal or distributed region that selectively discriminates one of the other object categories from the alternatives.
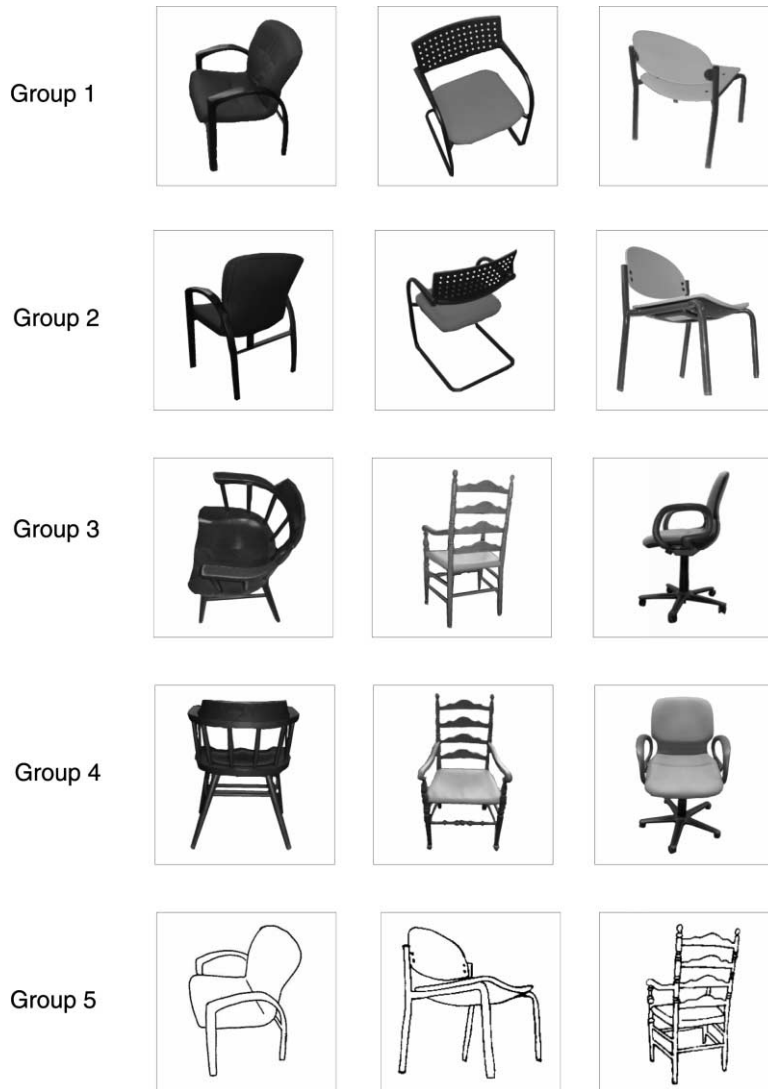
Figure 2. Examples of Stimuli for the Category "Chairs"

There are eight stimuli each in groups 1 to 4 and 32 stimuli in group 5. The other categories are handled in the same fashion.

Results

Analysis over All Visually Active Voxels

The average percent correct discrimination performance of all pairwise comparisons between the seven categories is obtained using an analysis technique similar to that of Haxby et al. (2001) (see Figure 1). Figure 3 (A, blue bar) shows that overall performance is quite high (mean across subjects: 96% of the pairwise comparisons correct), replicating the findings of Haxby et al. (2001). Note, however, that pairwise percent correct is an inflated measure. With seven categories, if the correct category is ranked second out of the seven alternatives, not first, this would produce 86% pairwise correct.

In order to evaluate how performance varies across different types of category comparisons, we separately scored performance on (1) "face versus object" discriminations, i.e., the four pairwise discriminations in which faces are contrasted with each of the four small inanimate object categories (bottles, shoes, scissors, and chairs), (2) the analogous four "house versus object"

discriminations, and (3) "object versus object" discriminations, i.e., the six discriminations involving each of the possible pairs from inanimate objects. Figure 3 shows performance on each of these discrimination types (including the overall measure across all 84 discriminations, see Experimental Procedures section) as a function of stimulus generalization, that is, whether the two sets of activation maps were based on stimulus images that were identical (as in the Haxby et al. [2001] study) or that differed in viewpoint, exemplar, or image format (Figure 2).

A two-way ANOVA across subjects found a significant main effect of discrimination type, $F(3,15) = 16.2$, $p < .001$, indicating higher performance for discriminations between faces versus objects and houses versus objects than for discriminations between pairs of small inanimate objects (see Figures 3B–3D). However, neither the main effect of stimulus generalization nor the interaction of stimulus generalization by discrimination type reached significance, $F(3,15) = 1.3$, $p > 0.3$ and $F(9,45) = 1.7$, $p > 0.1$, respectively. Thus, performance does not decline substantially when the two sets of activation
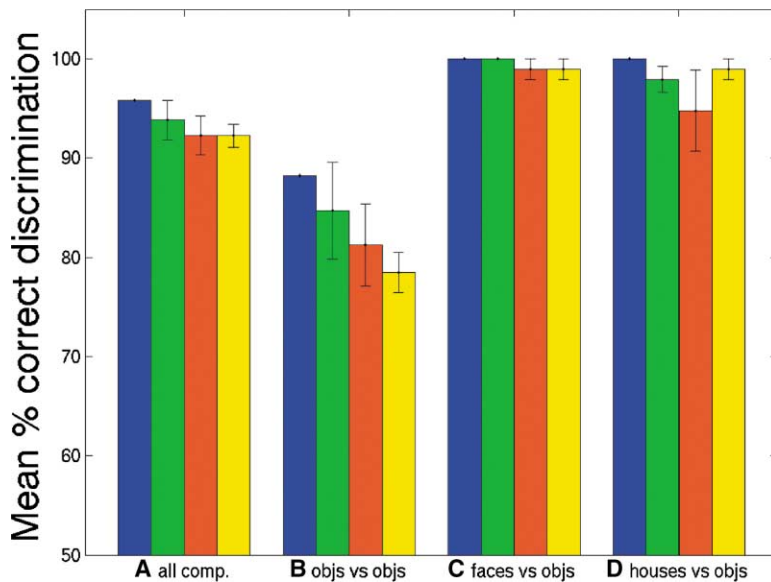
Figure 3. Mean Percentage Pairwise Correct Discrimination and Standard Error across Six Subjects for Different Discrimination Types Based on All Visually Active Voxels

The two sets of activation maps used in the analysis were produced from stimulus sets that were either identical grayscale photographs (blue), photographs with different viewpoints (green), photographs with different exemplars (red), or different formats (i.e., photographs and line drawings) (yellow). Chance is 50%. The percent correct discrimination is calculated from the average across: (A) all pairwise comparisons between the seven object categories, (B) all pairwise comparisons between two small man-made objects, (C) all pairwise comparisons between faces and small man-made objects, and (D) all pairwise comparisons between houses and small man-made objects.

maps are based on stimulus images that differ in viewpoint, exemplars, or image format, compared to when the two sets of activation maps are based on subjects viewing identical stimulus sets (as in the Haxby et al. [2001] study). However, a post hoc paired sample t test did indicate significantly lower performance for discriminations between the four object categories when one set of activation maps were based on photographs and the other on line drawings, compared to when the two stimulus sets were identical (t = 4.7, p < .01), suggesting that performance may decline for some kinds of discriminations and image transformations.

This overall pattern of results did not change substantially when all face-selective voxels, or all house-selective voxels, were omitted from the analysis, indicating that sufficient information exists outside the FFA to enable discrimination of faces versus nonfaces, and sufficient information exists outside the PPA to discriminate houses from nonhouses.

Are the FFA and PPA Selective in the Information They Contain?

The previous analyses across all visually responsive voxels (or all voxels except the face-selective or house-selective voxels) suggest that information about object categories is quite distributed across the ventral visual pathway, with information relevant to faces and houses present in regions that do not respond selectively to these categories. If information about object categories were completely distributed, then it would not matter what region of cortex one sampled when making a given category discrimination. To test this hypothesis, we next measured pairwise categorization accuracy for the same three kinds of discriminations investigated in the previous section (faces versus objects, houses versus objects, and objects versus objects) as a function of the region of cortex sampled (the FFA, PPA, retinotopic cortex, or other nonretinotopic cortical regions) and stimulus generalization type (either identical stimuli, or photographs to line drawings).

One problem that arises in this analysis is that the number of voxels in each of these regions varies many-fold, so a measure of the information in each area confounds the information per voxel with the number of voxels in that area. To deal with this problem, we first measured performance as a function of the number of voxels included in the analysis for each of these areas separately (see Figure 4). The voxels were randomly selected within each area. As expected, performance increased with number of voxels. For the analysis over all visually active voxels, performance was well over half the asymptote level by the time 30 voxels were included in the analysis. In addition, at least 30 voxels could be found for each region in each subject. We therefore used 30 voxels for each of the four cortical regions in this analysis. Because voxels were necessarily clustered together in the FFA and PPA, we chose 30 voxels from similarly clustered regions for the retinotopic and other nonretinotopic regions. (see following section for analysis based on unclustered voxel sets). For each subject, one cluster was centered on the FFA, another on the PPA, and another within the retinotopic cortex; the remaining ten clusters had random locations in the visually responsive occipito-temporal cortex not overlapping with any of these three regions and not overlapping with each other.

The resulting means across subjects of the pairwise percent correct performance in each condition are shown in Table 1. A $3 \times 4 \times 2$ ANOVA across subjects on the pairwise percent correct performance for the resulting 24 conditions revealed a significant main effect of discrimination type, $F_{(2,10)} = 15.1$, $p < .005$ and a significant interaction of discrimination type by cortical region, $F_{(6,30)} = 12.5$, $p < .001$. As in the previous analysis across all voxels, the main effect of stimulus generalization did not reach significance, $F_{(1,5)} = 4.0$, $p > 0.05$. The triple interaction of stimulus generalization $\times$ discrimination type $\times$ cortical region revealed only a nonsignificant trend, $F_{(6,30)} = 1.9$, $p > 0.1$. No other main effects of interactions reached significance, i.e., all F's are smaller than one.
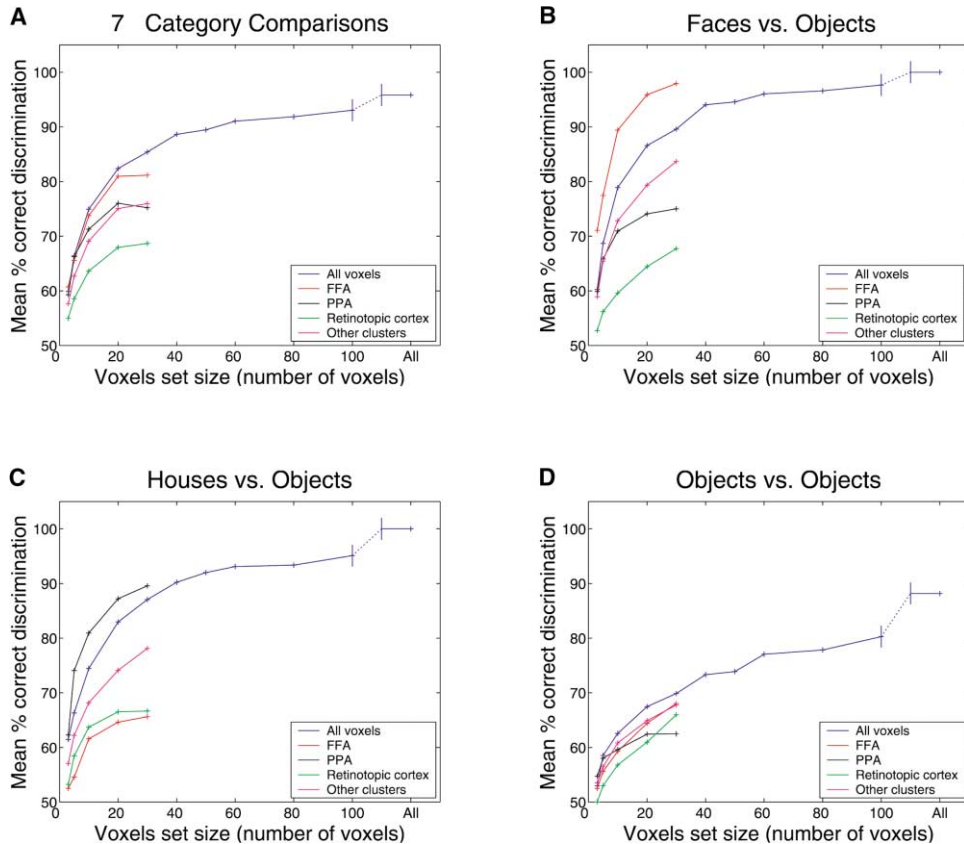
Figure 4. Mean Percentage of Correct Discrimination across Six Subjects as a Function of the Number of Voxels in Different Cortical Regions

For each subject and each subset size, the percentage of correct discrimination is averaged over many different subsets of randomly selected voxels. Chance level is 50%. Overall performance on (A) all 84 discriminations for the seven categories, (B) face versus object discrimination averaged over four pairwise comparisons (faces versus chairs, faces versus bottles, faces versus scissors, faces versus shoes), (C) same as (B) but for houses instead of faces, (D) object discrimination averaged over six pairwise comparisons (chairs versus bottles, chairs versus scissors, chairs versus shoes, bottles versus scissors, bottle versus shoes, scissors versus shoes).

The critical interaction of discrimination type by cortical region indicates that cortical regions differ from each other in the category discriminations they are most informative about. This finding leads directly to the central question of whether the FFA and PPA carry information about nonpreferred stimuli. To address this question, we measured performance on the three different discrimination types separately within the entire FFA, the entire PPA, and the 30 voxel clusters in retinotopic cortex. These data, which are shown in Figure 5, reveal that the FFA performs better than chance on face versus object discriminations ($p < .001$), and also significantly better on face versus object discriminations than house versus object ($p < .05$) or object versus object ($p < .01$)

discriminations. Similarly, the PPA performs significantly better than chance on house versus object discriminations ($p < .005$), and also significantly better on house versus object discriminations than object versus object discriminations ($p < .05$). The comparison of PPA performance on house versus object to face versus object discrimination performance did not reach significance ($p = .08$), which is not surprising given that the PPA responds less strongly to faces than to other objects and thus exhibits some ability to distinguish face from nonface objects (Epstein et al., 1999; Epstein and Kanwisher, 1998).

These comparisons reveal that the FFA and PPA are significantly more informative about discriminations in-

Table 1. Mean Pairwise Percent Correct Discrimination and Standard Deviation across Six Subjects for Different Clusters

Format	Identical (Photos-Photos)			Different (Photo-Drawings)		
Discrimination Type	F vs O	H vs O	O vs O	F vs O	H vs O	O vs O
FFA cluster	97.9 ± 3.2	65.6 ± 16.7	68.1 ± 8.6	97.9 ± 3.2	59.4 ± 17.6	66.7 ± 15.4
PPA cluster	75.0 ± 19.8	89.6 ± 7.6	62.5 ± 15.6	62.5 ± 17.7	86.5 ± 9.2	60.4 ± 17.0
"V1" cluster	67.7 ± 6.1	66.7 ± 11.7	66.0 ± 18.0	63.5 ± 13.9	71.9 ± 8.6	54.9 ± 18.1
Other clusters	83.6 ± 5.4	78.1 ± 8.2	67.8 ± 3.8	76.0 ± 8.1	73.9 ± 8.2	61.9 ± 6.6

"F vs O" stands for faces versus objects, "H vs O" houses versus objects, and "O vs O" objects versus objects.
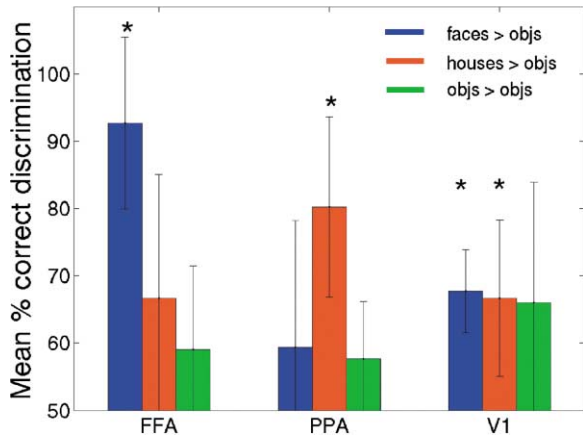
Figure 5. Mean Percentage of Correct Discrimination and Standard Deviation in Three Different Regions

The performance is calculated for different types of pairwise discriminations in the entire FFA, the entire PPA, and a 30 voxel cluster in the calcarine sulcus (probably V1), based on activations from identical images. A t test comparing performance to chance level (50%) produces the following p values: FFA: faces versus objects ($p < 0.001$), houses versus objects ($p = 0.08$), objects versus objects ($p = 0.14$); PPA: faces versus objects ($p = 0.28$), houses versus objects ($p = 0.003$), objects versus objects ($p = 0.08$); V1: faces versus objects ($p < 0.001$), houses versus objects ($p = 0.02$), objects versus objects ($p = 0.08$). The stars (*) indicate p values below 0.05.

volving preferred than nonpreferred stimuli. But do the FFA and PPA carry any categorical information at all about nonface nonhouse objects? As shown in Figure 5, object versus object discriminations were not significantly higher than chance in either the FFA or the PPA. A few of these comparisons were close to significance (see the legend to Figure 5 for details), so it is possible that a small amount of information about nonpreferred categories may be present in the FFA and PPA. However, as can be seen in Figure 5, performance on discriminations involving nonpreferred categories in the FFA and PPA was no higher than performance on the same discriminations based on retinotopic cortex. This result suggests that the FFA and PPA do not contribute any new discriminative information for nonpreferred categories beyond the presumably low level feature correlates of these categories already extracted in retinotopic cortex.

Selectivity for Other Object Categories

Thus, the FFA and PPA appear to carry categorical information primarily or only about their preferred stimuli. Do similarly selective regions exist for any of the categories of small man-made objects? Only one category, scissors, produced a clustered region of somewhat selective activation that was in a roughly similar location across subjects. But analyses based on this region alone produced only 66% correct discrimination of scissors versus faces, houses, and other man-made objects. This performance is lower than the scissor discrimination performance found in the FFA (72.5%) and similar to the one found in the PPA (62.5%), so it does not appear to support particularly effective scissor discrimination.

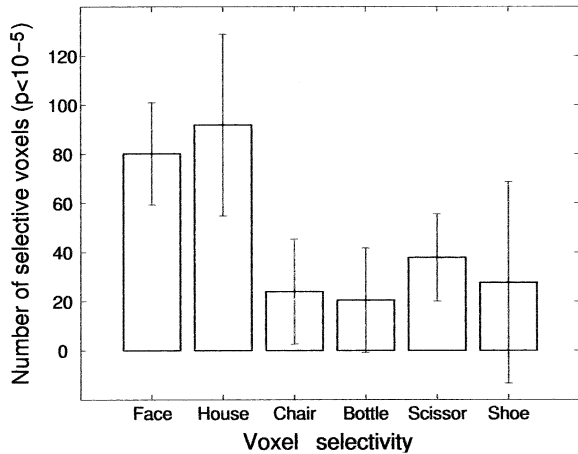Perhaps selectivity for object categories other than

Figure 6. Mean across Subjects of the Number of Voxels Selective for Faces, Houses, Chairs, Bottles, and Shoes

The selective voxels are the voxels for which the activation is significantly higher for one category than for the other ones ($p < 10^{-5}$, uncorrected for multiple comparisons).

faces and houses exists but is not segregated within focal regions of cortex. We tested whether distributed selectivities exist by measuring discrimination performance on distributed voxel sets that are selective to one category over the other ones. Figure 6 shows the average number of voxels selective for each category with $p < 10^{-5}$. This number is much smaller for small man-made objects than for faces or houses. Since discrimination performance depends on the number of voxels, we measured discrimination performance based on six different sets of 30 voxels, each set comprised of the most selective voxels for one category (faces, houses, chairs, shoes, scissors, and bottles). Table 2 shows that face-selective voxels perform better on face discrimination than on discrimination involving other stimulus classes, and house-selective voxels perform somewhat better on house (and face) discriminations than other discriminations. However, for bottle-, scissor-, shoe-, and chair-selective voxels, performance on discriminations involving their preferred category is no better than it is for discriminations involving nonpreferred categories. Furthermore, for bottles, shoes, and scissors, the voxels most selective for these categories performed no better than chance on any discriminations (except scissor-selective voxels on faces). Thus, we find no evidence for sets of voxels that selectively discriminate other stimulus categories, even when we allow such voxels to be distributed across the cortex

Discussion

Our data indicate that some information about object categories, including faces and houses, is present in the pattern of neural responses across broad regions of cortex in the ventral visual pathway. This result replicates the findings reported by Haxby et al. (2001). Further, and somewhat surprisingly, this discriminative information appears not to be based solely on low level features that covary with object categories because per-

Table 2. Mean Percent of Correct Discrimination and Standard Deviation across Six Subjects on Pairwise Discriminations Involving Each Category

Percentage of Correct Discrimination

	Faces	Houses	Chairs	Bottles	Scissors	Shoes
30 Most Selective Voxels for:						
Faces	**93.3 ± 11.0**	**82.3 ± 12.1**	**78.1 ± 10.3**	67.7 ± 17.0	**71.9 ± 9.5**	**66.7 ± 15.6**
Houses	**86.5 ± 17.4**	**81.7 ± 16.6**	**76.0 ± 14.5**	**77.1 ± 16.1**	**75.0 ± 18.9**	**72.9 ± 11.6**
Chairs	**75.0 ± 18.9**	**71.9 ± 17.6**	**71.7 ± 11.3**	**78.1 ± 19.6**	**75.0 ± 14.8**	66.6 ± 17.1
Bottles	60.4 ± 23.6	51.0 ± 16.9	57.3 ± 22.2	54.2 ± 23.5	48.9 ± 25.1	50.0 ± 18.9
Scissors	**71.9 ± 14.7**	65.6 ± 24.3	64.6 ± 23.9	62.5 ± 15.8	64.2 ± 16.9	64.6 ± 15.1
Shoes	60.4 ± 24.9	56.3 ± 18.9	58.3 ± 27.0	57.3 ± 21.8	53.1 ± 20.4	54.2 ± 18.6
FFA (16)	**82.5 ± 9.4**	56.2 ± 15.3	60.4 ± 10.2	64.6 ± 16.6	62.5 ± 13.7	50.0 ± 14.7
PPA (14)	62.5 ± 26.5	**67.5 ± 10.3**	62.5 ± 12.5	**61.5 ± 10.8**	58.3 ± 17.5	51.0 ± 15.5

The selective voxels for each category were the 30 voxels most strongly selective for that category independent of cortical location or clustering. The correct discrimination of the category for which the voxels are selective is measured over all pairwise discriminations in which this preferred category is contrasted with the other categories. For the discrimination of nonpreferred categories, no discriminations involving the preferred category were included. Performance values that are significantly higher than chance are in bold and underlined. The high performance on nonpreferred categories in the face- and house-selective voxels are due to the face- and house-selective voxels outside the FFA and PPA; this can be seen from the fact that when these voxels are removed from the analysis, discrimination performance on most nonpreferred categories is not significantly better than chance whereas discrimination performance on preferred categories remains high (bottom two rows of table). The number of selective voxels from the original set of 30 voxels that are located in the FFA and PPA are given in parentheses (averaged across six subjects).

formance does not decline substantially when the activation maps for each category are based on images that differ in category exemplar, viewpoint, and image format. It remains to be determined whether such apparently distributed category information would still be present across broad regions of the ventral visual pathway if the stimulus sets producing the activation maps differed in other dimensions such as size and retinal position.

However, the ventral visual pathway is not homogeneous and the voxels and regions within it are not interchangeable or equipotential in the information they carry about different category discriminations. This fact is revealed by a significant interaction in category discrimination performance between the kind of discrimination that is made and the region of cortex that is used as the basis for that discrimination. Further, although the FFA is highly informative about discriminations between faces and other categories, it provides little or no information about discriminations between different nonface object categories. Similarly, the PPA allows good discrimination of houses from other categories, but provides little or no information about discriminations among nonhouse objects. Finally, neither the FFA nor the PPA outperforms retinotopic cortex in discriminations between nonpreferred stimuli, suggesting that any small contribution made by the FFA or PPA on these discriminations may be based not on abstract category information, but on low level visual features that were confounded with particular categories in our stimulus sets.

Thus, although categorical information is apparently spread over a broad expanse of the ventral visual pathway, our analysis finds little or no evidence that the FFA and PPA carry discriminative information about nonpreferred stimuli. This finding is inconsistent with the suggestion of Haxby et al. (2001) that "regions such as the parahippocampal place area or the fusiform face area are not dedicated to representing only spatial arrangements of human faces but, rather, are part of a more

extended representation for all objects" (pg. 2427). The most likely reason why we reached different conclusions from those of Haxby et al. (2001) is that they did not separately measure performance on discriminations involving preferred stimuli from those involving only nonpreferred stimuli, leaving them unable to accurately assess the performance of the FFA and PPA on discriminations between two nonpreferred stimuli.

Of course, fMRI has limited spatial resolution, and our inability to detect categorical information about nonpreferred stimuli in the FFA and PPA does not prove that this information is not present in these areas. Future studies using higher spatial resolution may be able to detect this information. (Indeed, one would expect even a filter designed only for discriminating exemplars from nonexemplars of category X to have some ability to discriminate exemplars versus nonexemplars of category Y, insofar as X and Y share some visual features.) But the mere existence of discriminative information is no guarantee that this information forms part of the representation of the object and is used in object categorization. How then would we determine whether any such information in the FFA (for example) plays an important role in discriminations involving stimuli other than faces?

A recent neuropsychological study addresses both problems. Wada and Yamamoto (2001) report on the case of a man with an unusually small lesion restricted to the region of the right FFA who was severely disrupted on a variety of face recognition tasks, yet fully preserved on all of the tests of object recognition they ran. If we assume that this man's lesion included the right FFA, then these data suggest that even if some small amount of category-discriminative information exists in the FFA for nonfaces (undetected in our study), this information appears not to play a necessary role in the recognition of those nonface objects. The fact that this patient, like most prosopagnosics, knows that a face is a face but does not know whose face it is suggests that the FFA

is not necessary for discriminating faces from nonfaces, but is necessary for discriminating between different individual faces. This hypothesis is also consistent with Haxby et al.'s (2001) finding, replicated here, that discriminating faces from nonfaces can be done on the basis of activation maps even when the FFA is excluded from the analysis. Converging evidence for this hypothesis would be obtained if future experiments with higher spatial resolution find that the neural response within the FFA alone is sufficient to distinguish between individual faces.

More generally, the present results are encouraging for the validity of preferred responses as an index of the function of an area, voxel, or neural population. While it remains possible that nonpreferred responses carry some important part of the representation in some neural representations, the present data suggest that within the ventral visual pathway the focal regions that respond most strongly to faces and places are also the regions that are most informative about discriminations involving those categories. Such analyses cannot indicate whether these regions are in fact the main or only regions used in making such discriminations. Nonetheless, there would seem to be substantial advantages in sampling a restricted region of cortex where the relevant discrimination information is carried with high signal-to-noise ratio (SNR), compared to sampling a broad expanse of cortex spanning several centimeters that contains numerous voxels with low SNR.

Experimental Procedures

Subjects
Eight healthy subjects (four males and four females) participated in this study. Two subjects (one male and one female) were excluded prior to the analysis due to excessive head motion.

Stimuli
Stimuli were collected from seven different categories used by Haxby et al. (2001): faces, cats, houses, chairs, scissors, shoes, and bottles. Scrambled images were also presented to the subjects, but were not used in the analysis because they differ in too many respects from the other images, making their discrimination trivial. For each category, we collected 16 different exemplars with two different viewpoints each. Stimuli were either digitized grayscale photographs with a white background or line drawings of the same photographs. Stimulus size was approximately 9° of visual angle. A small dot in the center of the image was present during the whole scan and subjects were instructed to maintain fixation. Stimuli were presented in the center of the screen on a gray background.

Procedure
The image sets for each category were divided into five groups (Figure 2). Group 1 contained the first eight exemplars from one viewpoint. Group 2 contained the same exemplars as group 1 seen from a different viewpoint. Groups 3 and 4 were analogous to groups 1 and 2 but contained the other eight exemplars. Group 5 contained line drawing versions of all 32 images. Each subject was run on 24 different scans, 16 scans of grayscale photographs and 8 scans of line drawings. Each scan consisted of 16 s epochs, one for each category, preceded and followed by 8 s fixation periods. The pictures were presented for 200 ms with an 800 ms blank interval. Each picture was presented twice per epoch. Subjects were instructed to press a button whenever they saw two identical images consecutively (1-back task). This happened twice per epoch. The task was designed to keep the subjects alert and attentive. Images from each category were presented in different epochs within each scan in a way that balanced the serial position of the category across scans.

An additional scan was performed on two subjects, during which a radial wedge consisting of flickering black and white checks was presented for 16 s in each of four directions (left, right, top, bottom). This allowed us to locate V1 functionally (Grill-Spector et al., 1998).

Each group of grayscale photographs was presented in four scans. This procedure allowed us to divide the set of 16 scans in two halves (eight scans each) that contained either different viewpoints (groups 1 and 3 in the first half and groups 2 and 4 in the second half), different exemplars (groups 1 and 2 in the first half and groups 3 and 4 in the second half), or identical images (each half contained images from all groups). Each group of line drawing stimuli was presented in two out of the eight scans.

fMRI Data Acquisition
Scanning was done on a 3T Allegra scanner at the MGH-NMR Center in Charlestown. A head coil and a Gradient Echo pulse sequence with TR, 2 s; TE, 30 ms; flip angle, 90° was used. Twenty 4 mm thick slices were oriented parallel to the temporal lobe (near axial slices). The slices covered the entire temporal lobe plus most of the occipital lobe. For one subject, the slices were oriented perpendicular to the calcarine sulcus.

fMRI Data Analysis
Motion correction using the Fsfast analysis toolbox was performed prior to the data analysis (Burock and Dale, 2000). For each subject, the set of visually active voxels was defined as the voxels in the temporal and occipital cortex that were significantly more active during stimulus epochs than fixation for at least one category (t test, $p < 10^{-5}$) in the first half of data (averaged across eight scans). The set of visually active voxels is different for each way of grouping. For each category, the activation maps represent the response amplitude of the category epoch relative to fixation averaged across eight scans. In order to remove outliers in the signal, the response amplitude was determined by fitting a sixth-order polynomial. Seven activation maps per category were obtained (two for the identical image case, two for the different viewpoint case, two for the different exemplar case, and one for the line drawings). The percent correct discrimination was calculated from pairwise comparisons of correlation coefficients according to the method of Haxby et al. (2001). A given pairwise comparison between categories A and B is deemed a correct identification of category A if the correlation coefficient between the two maps of category A is higher than that between those of categories A and B. The percent correct discrimination is obtained from the average of all pairwise comparisons (between the first and the second data set and vice versa) involved in the discrimination. For example, in the comparison involving all categories, the percent correct discrimination is based on $7 \times 6 \times 2 = 84$ pairwise comparisons.

Performance as a Function of the Number of Voxels
For the analysis in Figure 4, correct discrimination performance was calculated as a function of voxel set size. For each voxel set size, the correct discrimination is based on the average performance over many different subsets of randomly selected voxels. For each cortical region, the voxels were randomly selected from the set of voxels that constitutes this region. The number of subsets was proportional $1/\sqrt{N}$, where N is the number of voxels, so that the error is approximately independent of N.

ROI Analysis
The location of the fusiform face area (FFA) was determined using the voxels in half of the data of the identical images case that were significantly more active for faces than chairs, bottles, scissors, and shoes (t test, $p < 10^{-5}$, Hanning radius for smoothing equal 2). Similarly, voxels that were significantly more active for houses than other small man-made objects defined the parahippocampal place area (PPA). The substantially lower PPA response to houses compared to scenes makes this an only approximately correct PPA localizer (Epstein and Kanwisher, 1998). Retinotopic cortex was determined anatomically by adjacency to the calcarine sulcus in four subjects. In two other subjects, an additional scan (see Procedure) allowed us to functionally define the borders of retinotopic areas by mapping the representations of the vertical and horizontal visual

field meridians (Engel et al., 1994; Sereno et al., 1995). For the cluster analysis, the FFA, PPA, and retinotopic cortex clusters were defined as the 30 visually active voxels centered around the above-mentioned functional regions. We use the term "retinotopic cortex" rather than "V1" because the cluster of 30 voxels might include additional retinotopic areas (V2,V3).

References

Aguirre, G.K., Zarahn, E., and D'esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron *21*, 373–383.

Burock, M.A., and Dale, A.M. (2000). Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. Hum. Brain Mapp. *11*, 249–260.

Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science *293*, 2470–2473.

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., and Shadlen, M.N. (1994). fMRI of human visual cortex. Nature *369*, 525.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598–601.

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? Neuron *23*, 115–125.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., and Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. Neuron *21*, 191–202.

Halgren, E., Dale, A.M., Sereno, M.I., Tootell, R.B., Marinkovic, K., and Rosen, B.R. (1999). Location of human face-selective cortex with respect to retinotopic areas. Hum. Brain Mapp. *7*, 29–37.

Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P., and Grady, C.L. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J. Neurosci. *14*, 6336–6353.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., and Haxby, J.V. (1999). Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. USA *96*, 9379–9384.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. *17*, 4302–4311.

McCarthy, G., Puce, A., Gore, J.C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. *9*, 604–609.

Puce, A., Allison, T., Asgari, M., Gore, J.C., and McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. *16*, 5205–5215.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.

Tovee, M.J., Rolls, E.T., Treves, A., and Bellis, R.P. (1993). Informa-tion encoding and the responses of single neurons in the primate temporal visual cortex. J. Neurophysiol. *70*, 640–654.

Wada, Y., and Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. J. Neurol. Neurosurg. Psychiatry *71*, 254–257.