Class Information Predicts Activation by Object Fragments in Human Object Areas

Yulia Lerner[1,2], Boris Epshtein[1], Shimon Ullman[1*], and Rafael Malach[1*]

Abstract

■ Object-related areas in the ventral visual system in humans are known from imaging studies to be preferentially activated by object images compared with noise or texture patterns. It is unknown, however, which features of the object images are extracted and represented in these areas. Here we tested the extent to which the representation of visual classes used object fragments selected by maximizing the information delivered about the class. We tested functional magnetic resonance imaging blood oxygenation level-dependent activation of highly informative object features in low- and high-level visual areas, compared with noninformative object features matched for low-level image properties. Activation in V1 was similar, but in the lateral occipital area and in the posterior fusiform gyrus, activation by ''informative'' fragments was significantly higher for three object classes. Behavioral studies also revealed high correlation between performance and fragments information. The results show that an objective class-information measure can predict classification performance and activation in human object-related areas. ■

INTRODUCTION

The ability of the human visual system to recognize and classify objects is one of the most impressive capacities and widely investigated topics in cognitive science. Such tasks are performed by the human brain with remarkable efficiency that exceeds that of any known artificial machinery. For humans, complex scenes are organized perceptually as a collection of separate objects. However, via what mechanisms are these individual objects classified and identified? Past approaches to this question include both holistic (Riesenhuber & Poggio, 1999; Marr & Nishihara, 1978) and part-based mechanisms (Martelli, Majaj, & Pelli, 2005; Palmeri & Gauthier, 2004; Biederman, 1987). A holistic model for recognition, as advocated by Gestalt psychologists (Wertheimer, 1938), suggests that object identification takes on a global form, that is, objects are perceived as whole shapes rather than as a summation of their parts. Parts- and feature-based models for recognition suggest that identification begins with the detection of many independent features (or subparts) of the object, which is then followed by an integration process (Robson & Graham, 1981). These approaches are not mutually exclusive, and the approach discussed further below combines aspects of both. It has also been suggested that visual object recognition involves a hierarchical process (Riesenhuber & Poggio, 1999; Marr & Nishihara, 1978). According to this view, recognition involves a multistage detection and integration process, by which information about subparts of increasing complexity is collected, eventually leading to the identification of the full object.

Contemporary research relating to these problems provides a natural common ground for the convergence of two fields of study: computational modeling and functional brain imaging. These two disciplines are highly complementary, suggesting that the advancement of either can benefit from close interaction with the other. This interdependent relationship between computational modeling and human vision provides a basis for the current study, and is further elaborated by the parallels drawn between our computational and experimental results.

Computational studies over the past decade have shown the usefulness of selected object fragments as useful visual features for classification and recognition (Fergus, Perona, & Zisserman, 2003; Agarwal & Roth, 2002; Ullman & Sali, 2000). In these classification models, distinctive regions selected from object images during a learning stage are extracted, and an object class is represented as a configuration of such regions, or object fragments. According to a recent computational study of object recognition, the accuracy with which particular fragments of an object allow its correct identification depends on the amount of visual information contained within those fragments (Ullman, Vidal-Naquet, & Sali, 2002). The most informative features for visual classification were shown to be fragments of intermediate complexity (in terms of size of resolution), which show

[1]Weizmann Institute of Science, Rehovot, Israel, [2]Tel Aviv Sourasky Medical Center, Tel Aviv, Israel
*These authors contributed equally to the work.

similarity to visual features described in the posterior inferotemporal cortex of the monkey (Fujita, Tanaka, Ito, & Cheng, 1992). These features are maximally informative for classification when they are constructed hierarchically rather than used as integral features (Epshtein & Ullman, 2005).

Taken together, these computational models reveal a natural hierarchy of category-related fragments that corresponds to a plausible functional architecture of object recognition. However, the question remains: how is this functional hierarchy for recognition organized in the brain, if at all?

It is well established that significant regions within both the ventral and dorsal visual pathways respond selectively to objects. Specifically, a focal point of object-related activation was found within the lateral occipital complex (LOC) (Grill-Spector, Kourtzi, & Kanwisher, 2001; Kanwisher, Chun, McDermott, & Ledden, 1996; Malach et al., 1995). For instance, event-related potential and functional magnetic resonance imaging (fMRI) studies in humans have shown that the LOC responds more strongly to images of specific objects as opposed to visual noise, texture patterns, or scrambled object-parts (Grill-Spector et al., 1998; Malach et al., 1995). Identification of object-selective regions is a prerequisite for investigating more specific questions about the functional organization of visual system. As the previously discussed computational models suggest, it is possible that recognition and classification are processed in the brain by a hierarchy of features, selected to be informative for the recognition of different object classes. According to this view, the local low-level functionality of early visual areas is gradually transformed into increasingly complex visual representations along the posterior–anterior processing axis. Indeed, evidence from primate studies strongly supporting this idea has shown functional transition in brain activation from V1 cells, which responded to local orientations (Hubel & Wiesel, 1968), to V4 cells, which responded to more complicated visual cues such as specific colors and patterns (Gallant, Braun, & Van Essen, 1993; Schein & Desimone, 1990). Moreover, fMRI studies in humans suggest a corresponding hierarchy, advancing from low-level processing in the V1 to the more holistic object representation in the LOC (Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001; Grill-Spector et al., 1998). Specifically, scrambling experiments, in which images of objects were broken up into an increasingly larger number of fragments, revealed a corresponding increase in sensitivity to image scrambling along the hierarchical axis. In another set of experiments, a hierarchy of shape-selective regions was found within the LOC itself: whereas posterior regions were activated by object fragments, anterior regions responded selectively to either whole or half-object images (Grill-Spector et al., 2001).

Based on these and related studies, it appears that computational modeling and imaging studies converge to a similar view of classification by the visual cortex. Computational models that rely on visually informative fragments predict a class-specific hierarchy for the recognition and classification of objects. Similarly, imaging points to a corresponding functional hierarchy in the organization of the human brain. However, in order to draw a direct link between computational and physiological models of recognition, and to better understand the nature of the features used by the visual system, it is necessary to determine how computationally identified fragments affect activation of object-related brain regions. In the present study, we explored this issue by searching to what extent optimal fragments defined by an objective computational criterion produce preferential activation in object-related areas of the human brain. Specifically, we examined information-specific sensitivity by comparing cortical activation in response to optimal versus matching random fragments. As detailed below, the amount of information delivered by an image feature was based on a simple measure of its frequency within and outside the class of interest. Computational modeling has shown that features selected by this information criterion are useful for reliable recognition, but it is unclear whether they elicit preferential activation in high-order visual areas. Our results show a significant correspondence between information content and activity in higher order object-related areas. That is, object-selective regions showed preferential activation for computationally acquired informative fragments as compared to less informative ones.

METHODS

Subjects

Thirteen subjects with normal or corrected-to-normal vision ranging in age from 26 to 37 years participated in one or more of the following experiments: 13 in the conventional border mapping and in the fMRI Experiments 1 and 2; 6 in the control; and 12 in the behavioral tests. Prior to initiating the experiments, subjects provided written informed consent agreeing to participate in the study. Tel-Aviv Sourasky Medical Center approved the experimental protocol.

MRI Setup

Subjects were examined in a 1.5-T Signa Horizon LX 8.25 GE scanner. A custom quadrature surface coil was used for the meridian mapping experiment (Nova Medical, Wakefield, MA), and a standard head coil was used for the fragment experiments. Blood oxygenation level-dependent (BOLD) contrast was obtained with gradient-echo echo-planar imaging (EPI) sequence (TR = 3000 msec, TE = 55 msec, flip angle = 90°, field of view 24 × 24 cm^2, matrix size 80 × 80). The scanned volume included 25 to 26 nearly axial slices of 4 mm thickness and 1 mm gap covering the entire cortex.

A whole-brain spoiled gradient (SPGR) sequence was acquired for each subject to allow exact cortical segmentation, reconstruction, and volume-based statistical analysis of signal changes during the experiments. T1-weighted high-resolution (1.1×1.1 mm^2) anatomic images (124 images, 1.2 thickness) of the same orientation as the EPI slices were obtained to assist the incorporation of the functional data into the 3-D Talairach space (Talairach & Tournoux, 1988).

Extracting Informative Features

Informative classification features were extracted automatically for three object classes (faces, cars, horses) using an information maximization algorithm (Ullman et al., 2002). Briefly, the method examines a large number of candidate object fragments extracted from a set of training images, and selects from them an optimal subset of informative features.

Candidate fragments are first extracted from the training images at multiple locations, sizes, and resolutions. These fragments are searched for in all the database images. Normalized cross-correlation was used as similarity measure, although other methods, such as gradient-based measures, are also possible (Lowe, 2004). A fragment is considered to be present in an image if the similarity measure exceeds a predefined detection threshold. An optimal threshold is determined automatically for each fragment at a level that maximizes the delivered mutual information (MI), as explained below. The MI delivered by each fragment f_i, about the class C is computed using the formula (Cover & Thomas, 1991):

$$\mathrm{MI}(f_i(\theta_i); C) \equiv \sum_{\substack{f_i(\theta_i)=\{0,1\} \\ C=\{0,1\}}} p(f_i(\theta_i), C) \log\left(\frac{p(f_i(\theta_i), C)}{p(f_i(\theta_i))\, p(C)}\right)$$

Here θ_i is the detection threshold associated with fragment f_i. A binary variable C represents the class, namely, $C = 1$ if the image belongs to the class, and 0 otherwise. $f_i(\theta_i)$ is a binary variable: 0 means that fragment number i was not detected in the image (maximal correlation was smaller than θ_i), and 1 otherwise.

The delivered information depends on the detection threshold θ: if the threshold is too low, the fragment will be detected with high frequency in both class and nonclass images, and with high threshold the fragment will be missed in both class and nonclass images. The value θ_i of threshold yielding maximal information for the fragment f_i is computed for each fragment. The most informative fragments were then selected successively. After finding the fragment with the highest MI score, the search identified the next fragment that delivered the maximal amount of additional information. Additional fragments are added in a similar manner using a max–min procedure (Fleuret, 2004; Ullman et al., 2002). Informative fragments selected by the algorithm for three classes of objects are shown in Figure 1A. In the selection process, a large object fragment may have low information content because the probability of finding this particular fragment in a novel image becomes low. Such a fragment may contain, however, subfragments with higher information content than the full fragment. This is undesirable for our testing, and therefore, in selecting low-information features, the fragments we used were maximal in the sense that they did not contain smaller subregions with higher MI than the full fragment.

The nonclass images were collected from the various Web databases. To validate that the results do not depend on the choice of nonclass, we divided an original nonclass into two subclasses. Then MIs of the selected fragments were measured using each of the subclasses separately. The difference in the values of MI was small [(mean(abs($\mathrm{MI}_1 - \mathrm{MI}_2$)/max($\mathrm{MI}_1$,$\mathrm{MI}_2$) = 4.7%], indicating that computation was not sensitive to the actual choice of nonclass images.

Stimuli

Visual stimuli were generated on a PC and presented via an LCD projector (Epson MP 7200) onto a translucent screen. Subjects viewed the stimuli through a tilted ($\sim$45°) mirror positioned above each subject's forehead.

In a single trial, subjects were presented with an image composed of a configuration of seven different fragments, which were assigned randomly to the aperture locations, and belong to one of three object categories: faces, cars, and horses (Figure 1A). The fragments were initially selected as rectangular patches and then the circular apertures were placed. The comparison showed no significant change in MI between the original rectangular fragments and the same fragments with circular mask—the absolute difference between MIs was 0.06 ± 0.05. All fragments had the same size because we preferred to avoid a mixture of the size effect with effect of information. Fragments were categorized as either ''informative'' or ''random,'' with ''informative'' fragments computed to be maximally informative as described above. A set of ''random'' fragments (controlled for size) was collected for the purpose of comparison. In Experiment 1, ''random'' fragments were chosen from within a single image only. In Experiment 2, ''random'' fragments were chosen from across a range of images within a given category. Importantly, ''random'' fragments were chosen exactly of the same size and contrast properties as ''informative'' ones (see Analysis of Low-level Image Properties section). The external boundaries of the fragments were smoothed using a Gaussian Blur filter with radius 1.2 pixels. Fragment similarity in terms of image properties was later also verified by their identical V1 activation (see Results). Note that the way of selecting ''random'' fragments did not exclude the possibility that some of the ''random'' fragments

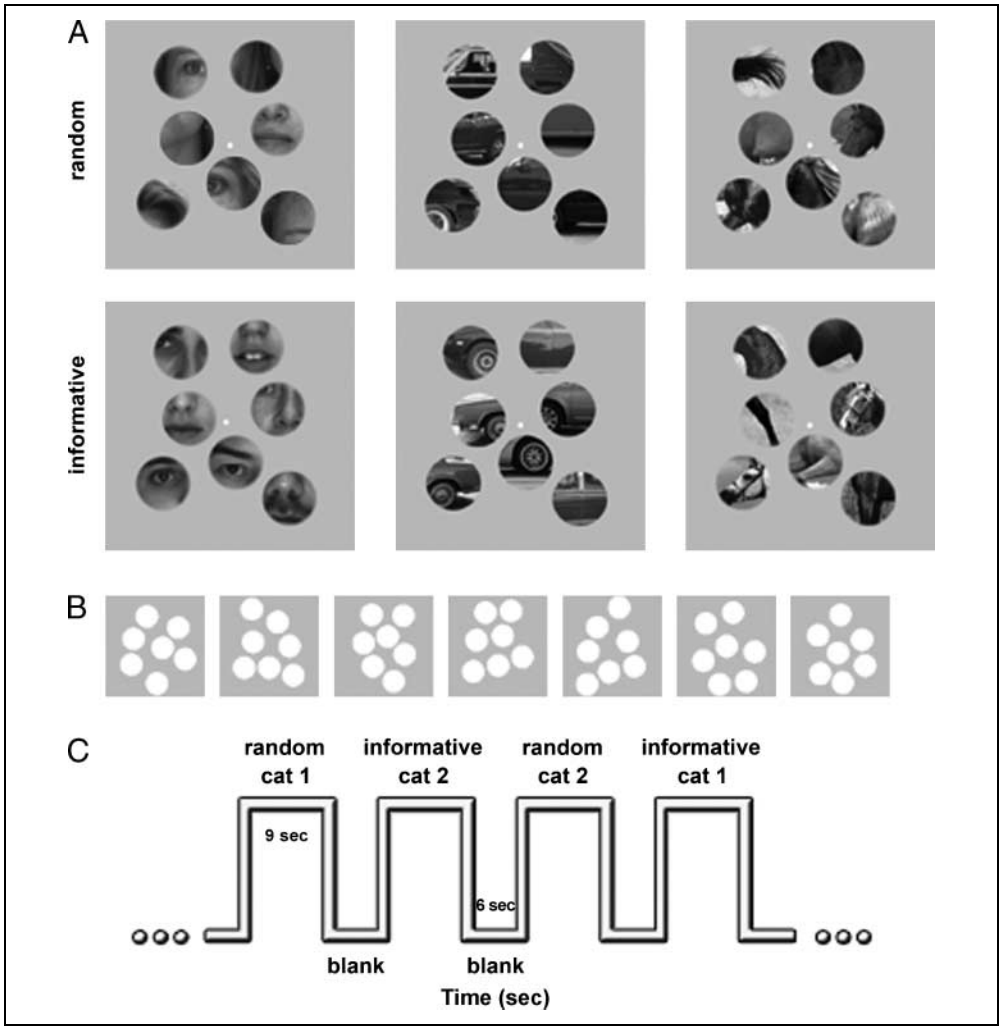


Figure 1. Stimuli and experimental design. (A) Examples of two types of the visual stimuli used in Experiments 1 and 2: top—''random'' fragments, bottom—''informative'' fragments. Stimuli were presented as a configuration of seven different fragments in three object categories: faces, cars, and horses. A white fixation point was presented in the center of all images. In Experiment 1, ''random'' fragments were selected from a single image, and in Experiment 2, from a range of pictures within a category. (B) Examples of configurations in which the fragments were arranged. The configuration differed between blocks. (C) A part from the time axis of Experiments 1 and 2. An interleaved short epoch design was used. The experiments lasted 522 sec and consisted of 33 visual blocks of 9 sec each. Each condition was repeated eight times and consisted of eight different and one repeated images.

happen to be ''informative'' but, on average, this is not likely. The two sets were compared in terms of the information delivered, and they were significantly different (see Results).

EXPERIMENTS

Fragments Experiment 1

Experiment 1 was conducted on images of faces and horses, and included four experimental conditions: face-''informative,'' face-''random,'' horse-''informative,'' and horse-''random.'' A single session lasted 522 sec and was composed of 33 blocks, each 9 sec long (Figure 1C). Each block consisted of 1-sec presentations of eight unique stimuli taken from the same condition: seven of the stimuli were presented once and the eighth was presented twice. Configurations differed between blocks (Figure 1B). Also, for every block consisting of a particular configuration of ''informative'' fragments, an equivalent block was constructed consisting of the same configuration but with ''random'' fragments selected from a single image (Figure 1A and B).

Each experimental session began with the presentation of a 21-sec-long blank screen followed by a 9-sec-long patterned stimuli block (excluded from all analyses), and ended with the presentation of a 12-sec-long blank screen. Subjects (n = 13) were asked to fixate on a small central dot and to perform a 1-back memory task intended to control subjects' alertness.

Fragments Experiment 2

To check if the way in which ''random'' fragments were selected affects the outcomes, and to extend this work to additional fragment category, this experiment was conducted. It was almost a replica of Experiment 1, but this time ''random'' fragments were selected from a range of pictures within a category. Cars and horses object categories were tested.

Fragments Experiment 3

Experiment 3 pursued two goals: (i) to check to what extent the level of activity is related to informative content per se, and (ii) to evaluate the involvement of con-

figurational ''holistic'' effects in the observed level of activity. This experiment was conducted on face fragments and included three informativeness levels: high, mid, and low. A single session lasted 582 sec and was composed of 37 blocks, each 9 sec long. Eight unique stimuli and one repeated were presented in each block. Fragments were arranged either in a normal face configuration or in a ''jumbled'' configuration. Normal and jumbled configurations were presented for all tested levels of information (Figure 7A). A central small fixation point was added to the images.

Each experimental session began with the presentation of a 21-sec-long blank screen followed by a 9-sec-long patterned stimuli block (excluded from all analyses), and ended with the presentation of a 12-sec-long blank screen. Subjects ($n = 8$) were asked to perform a 1-back memory task and to record their responses via button presses indicating same/different decision. Eye movements were recorded during the experiment (see below).

Control Experiment 1

Low-level image features may still affect the preferential activation of the ''informative'' fragments. To search this possibility, the control experiment was conducted. The experiment had a similar block design to the Fragments Experiment 3. Again, only the face category was tested. There were six types of visual stimuli in the experiment: ''Informative'' and ''random'' fragments were presented in a normal view, in an inverted view or were broken into nine parts and then randomly scrambled (see Figure 8). Each stimulus consisted of a single fragment. High-frequency edges in the scrambled stimuli were smoothed using a 2-D Gaussian function ($\sigma = 0.05$). A fixation point was presented in all stimuli. Subjects ($n = 8$) were asked to perform a 1-back memory task and push the buttons on the response box in order to indicate the same/different stimulus.

Control Experiment 2

The Control Experiment 2 was conducted for further investigation of configurational effects. It had the same block design as the previously described Fragments Experiment 1. Only the face category was tested. A total of four experimental conditions were tested. The first condition consisted of three fragments of a single face arranged in their normal facial configuration. The second condition consisted of three fragments arranged in the same configuration as in the first condition, but taken from three different face images. In this condition, the fragments were both ''informative'' and were arranged in a normal face configuration (e.g., eyes on top and mouth on the bottom), yet the final image was not of a specific single face, but features were selected from different faces. The last two conditions were equivalent to those used in the first two experiments. Namely, compositions of seven ''random'' fragments were chosen either from within a single face image or from across a range of face images. A fixation point was presented in all stimuli (see Figure 9, top panel for example stimuli).

Retinotopy Mapping Experiment

To define the borders of early retinotopic visual areas, the representation of vertical and horizontal visual field meridians was mapped for each subject (Sereno et al., 1995). Specific details about the procedure were described by Levy, Hasson, Harel, and Malach (2004). Briefly, triangular wedges of object images were presented either vertically (upper or lower) or horizontally (left or right) at 4 Hz in 18-sec blocks alternated with 6-sec blanks. Subjects were required to fixate on a small dot located at the center of the screen. Additionally, a ''house–face localizer'' experiment (Levy, Hasson, Avidan, Hendler, & Malach, 2001) was used to delineate the anatomical location of the posterior fusiform gyrus (pFs).

Eye Movements Control

The eye movement data were collected using MR-compatible eye tracker (ASL Eye Tracker, Model R-LRO6) in the magnet for seven subjects that participated in the Fragments Experiment 3 during the MR scans. Data points obtained during the blank epochs, and when the eye tracker was failed to indicate accurate eye position (for example, due to blinking), were excluded. The data for high and low ''informative'' fragments were analyzed separately, and horizontal and vertical projections of the clusters were shown. Additionally, separately for each observer, and each condition, the density of the fixation points was estimated using Gaussian kernels with a bandwidth equal to pupil size.

Behavioral Measurements

Human Performance in the Original Experiment

Behavioral measurements were taken outside the magnet. Subjects ($n = 12$) were presented with the same stimuli composed of seven fragments that were shown in the original fMRI experiments. Sixty-three additional images not belonging to three original object classes and including body parts, planes, and animals, with randomly selected features, were added to increase the difficulty of the task. Subjects were asked to identify the objects by saying face/car/horse/other. Thirty-two (16 ''informative'' and 16 ''random'') different images for each category—faces, cars, and horses—were used. In each trial, a single image was presented for 200 msec, followed by a 1250-msec-long blank screen, allowing subjects sufficient time to

respond. Percentage of incorrect response was calculated (Figure 6A). Prior to participating in the behavioral experiment, all subjects underwent a short training session (1.5 min) on a different set of stimuli.

Performance in a Single-fragment Experiment

To allow a direct comparison with computational measures, a single-fragment experiment was conducted. Subjects (n = 11) were presented with stimuli consisting of a single fragment each, in the analogy with computer simulations which were based on single fragments (see below). Eighty fragments (40 ''informative'' and 40 ''random'') with the highest MI were selected in each of three categories—faces, cars, and horses. Subjects were instructed about four categories while the fourth category (other) did not correspond to any stimulus. Each single image was presented for 200 msec, followed by a 1250-msec-long blank screen. Percentage of errors (misses and false alarms) was calculated (Figure 6B). Prior to participating in this experiment, all subjects underwent a short training session (60 sec).

Computational Simulations

Classification errors (sum of misses and false alarms) were also computed using simulations for each fragment in the experiment. Each fragment was searched in all the database images, and the number of correct detections and classification errors was recorded. Each detection in the incorrect class (e.g., a horse fragment detected in a car image) was considered an error. This procedure was performed for all three categories (Figure 6C).

DATA ANALYSIS

fMRI data were analyzed with the ''BrainVoyager'' software package (Brain Innovation, Maastricht, The Netherlands) and complementary in-house software. The first three images from each functional scan were discarded. The functional images were superimposed onto 2-D anatomic images and incorporated into the 3-D datasets through trilinear interpolation. The complete dataset was then transformed into Talairach and Tournoux (1988) space. Preprocessing of functional scans included 3-D motion correction, slice scan time correction, linear trend removal, and filtering out of low frequencies (up to 5 cycles/experiment). The cortical surface of each subject was reconstructed from the 3-D SPGR scan. Reconstruction included segmentation of the white matter using a grow-region function, smooth covering of a sphere around the segmented region, and expansion of the reconstructed white matter into the gray matter. The surface of each hemisphere was then unfolded, cut along the calcarine sulcus, and flattened.

Statistical Analysis

Single-subject Analysis

A General Linear Model (Friston et al., 1994) statistical analysis was used. A hemodynamic lag of 3 or 6 sec long was fitted to the model for each subject by maximizing the extent of the overall visual activations. A box-car predictor was constructed for each experimental condition except fixation, and the model was independently fitted to the signal of each voxel. A least-squares algorithm was used to calculate a coefficient for each predictor.

Multisubject Analysis

In addition to the single-subject analysis, data were analyzed across all subjects. Thus, the time series from images of all subjects were converted into Talairach space and normalized using the z-value. For each subject, the relative contribution of the predictors for each contrast was estimated separately. These values were then used to obtain significance at the multisubject level. Calculations of significance values from the activation maps were based on the individual voxel significance and on the minimum cluster size of 10 voxels (Forman et al., 1995). Multisubject maps were obtained using a random effect procedure (Friston, Holmes, Price, Buchel, & Worsley, 1999) and projected on a single, flattened, Talairach normalized brain. Statistical level ranged from $p < .05$ (darker colors) up to at least $p < .001$ (brighter colors) and was indicated by color scales.

Analysis of Low-level Image Properties

Using the two-dimensional Fourier transforms (FT) analysis for the stimuli (whole images), we compared the spectral power of the ''informative'' and ''random'' fragments in the low- and high-frequency range. The low-frequency range consisted of all frequencies where horizontal and vertical frequency components were smaller than half of the maximum frequency (12.5 cycles/deg), similarly for the high range. The overall power (total energy was computed for each image separately as sum of all squared amplitudes in FT, and then mean across image values was calculated) in the entire frequency domain, as well as power in the lower range (0–6.25 cycles/deg) and higher range (6.25–12.5 cycles/deg). Additionally, a similar analysis was done on an individual fragment basis.

Internal Localizer Approach

Details of the approach were previously published (Lerner, Hendler, & Malach, 2002). Briefly, one subset of the repetitions of a condition was used to localize the region of interest (ROI; biased statistically as a part of the test), whereas a signal in the complementary subset of the repetitions (unbiased statistically) was used to evaluate the

activation level. Specifically, for the each localizer test (e.g., ''high informative'' vs. ''low informative''), two statistical tests were conducted. In the first test, half of the epochs in a condition were used to define the ROI, and their activity was therefore statistically biased, whereas the other half of the epochs were used to obtain an independent measure of a statistically unbiased BOLD activation, and these halves were reversed for the second statistical tests. The average activation obtained from the unbiased epochs is presented in the figures (e.g., Figures 7 and 8).

RESULTS

Preferential Activation to ''Informative'' Fragments in the Human Visual Cortex

To what extent are object-selective brain regions sensitive to object fragments obtained either by maximizing information for classification, or selected randomly, from the same objects and with similar image properties? To address this question, we examined the activation level to visual stimuli composed of two types of fragments (''informative'' vs. ''random'') from three different object categories—faces, cars, and horses. Figure 1 illustrates the basic design of the experiments. An example of ''informative'' fragments for each object category and a comparable example of ''random'' fragments are presented in Figure 1A. The fragments were selected separately for each condition and presented in 9-sec blocks (Figure 1C) in the different configurations. Examples of such configurations are shown in Figure 1B. The two approaches in the way in which ''random'' fragments were selected (see Methods for details) did not yield significantly different results, and thus, data from the two experiments were combined in all further analyses.

To characterize the overall activation patterns observed during ''informative'' versus ''random'' fragment blocks, cortical maps were constructed representing activation as averaged across all subjects (Figure 2, top; $n = 13$, random effect, $p < .05$). The data reveal significantly higher activity in response to ''informative'' as opposed to ''random'' fragments in several areas of the visual cortex. Specifically, activation was found in both the ventral and dorsal streams, extending bilaterally to premotor areas. Ventrally, activation was found in two subdivisions of the LOC: in the pFs situated lateral and anterior to areas V4/V8, and in the lateral–occipital region (LO) located in the vicinity of the inferior occipital sulcus (IOS) or gyrus (IOG). Dorsally, preferential activity to ''informative'' fragments (vs. ''random'' ones) was found in the intraparietal sulcus (IPS). Additionally, an anterior activation focus was observed in the posterior part of the superior frontal sulcus (near the precentral sulcus, and slightly more rostral in the right hemisphere compared to the left). Condition-related activity was highly consistent across all subjects (see Figure 3 for comparable single-subject activation levels).

Additionally, we analyzed the fMRI response to ''informative'' versus ''random'' stimuli within ventral subregions of object-selective voxels for each of the object categories separately. The patterns of activation revealed by such ''single-category'' tests are shown in Figure 2 (bottom). Note that the single-category maps were consistent with behavioral performance in the categorization task—as the number of missclassifications increases from category to category (see Figure 6), the preferential activation to ''informative'' fragments was reduced.

ROIs were sampled for each subject, and a corresponding quantitative analysis of the activation levels was done for the object-related areas. Results are depicted in Figure 4. As can be seen in both the LO and pFs subdivisions, the activation was significantly higher for ''informative'' fragments compared with randomly selected fragments ($p < .005$). Importantly, preferential activation to more ''informative'' fragments was found across all object categories. Although a higher selectivity for faces was globally observed in the face-related pFs ($p < .0005$), the bias for ''informative'' fragments in this area was evident for the nonpreferred horse and car stimuli as well. Also, activity in the IPS showed a significantly stronger preferential effect for ''informative'' fragments in the face and horse images ($p < .01$), although the overall signal in this area was generally low. These findings suggest that the same area can contribute differentially to different object categories. In contrast, no substantial differences were found in the higher order collateral sulcus as well as in the early retinotopic visual areas. An intriguing question is how the selectivity to ''informative'' fragments develops along the hierarchy of visual areas. This is shown in Figure 5. Interestingly, a tendency (albeit not significant) for preferential activation for ''informative'' fragments can be detected as early as area V4 ventrally and V3A dorsally (Figure 5).

The ''random'' fragments used for stimulation were selected from the image so that their edges and contrast properties were similar to the corresponding ''informative'' fragments (see Analysis of Low-level Image Properties section). A Fourier power-spectral analysis did not reveal any significant difference between the informative and ''random'' fragments (for details, see Table 1). Moreover, other possible differences not having directly to do with class-specific fragments, such as a number of image ''interest points,'' were analyzed. The number of ''interest points'' calculated using a Harris corner detector (Harris & Stephens, 1988) was similar in the informative and ''random'' fragments (mean number per fragment: ''informative''—24.5; ''random''—25.7). The similarity in low-level properties was also evident in the similar activation level we found in V1 activity for the ''informative'' and ''random'' fragments. The information supplied by the randomly selected fragments was less than the information supplied by the informative features (average MI for ''random'' fragments: 0.2 ± 0.007; average MI for ''informative''

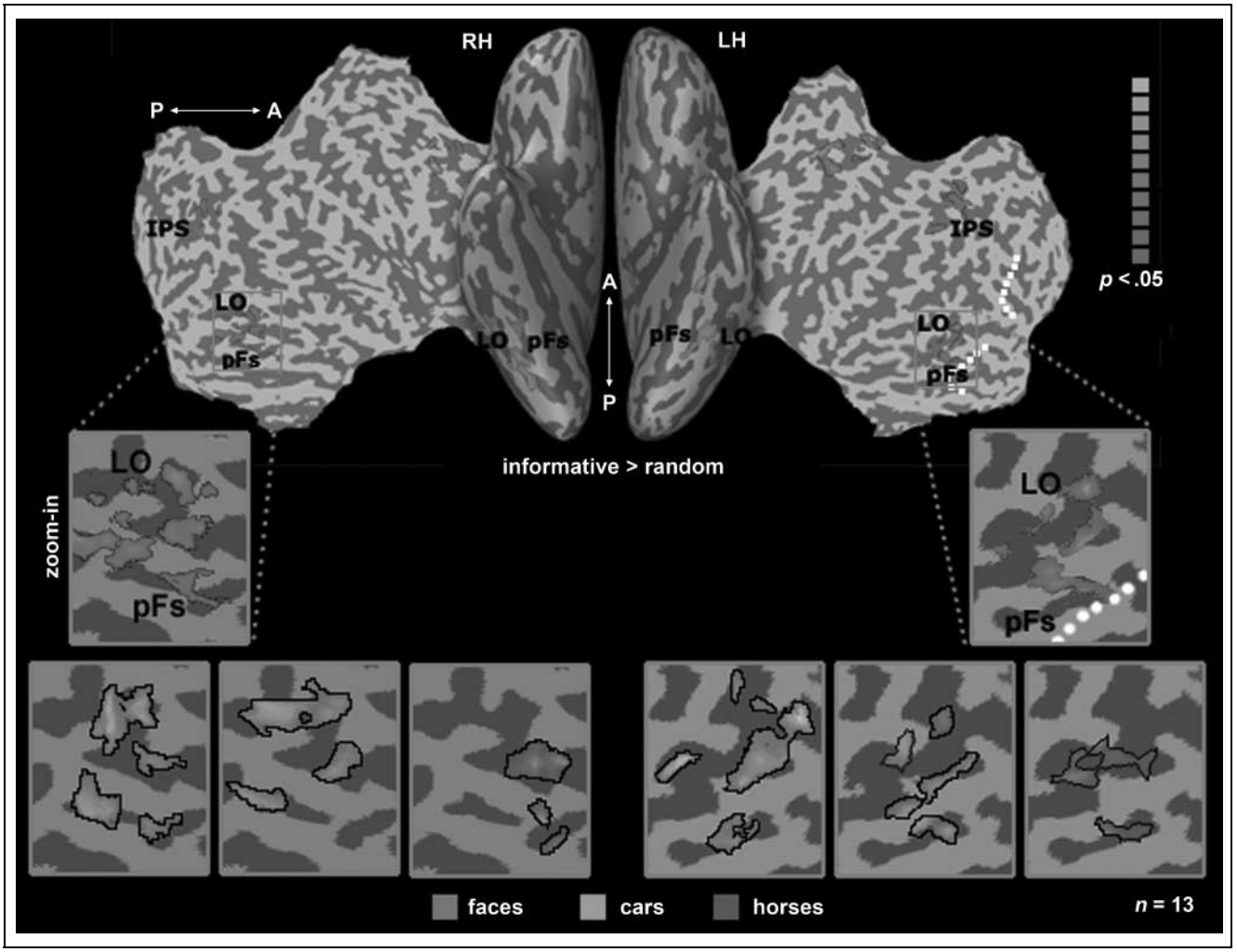
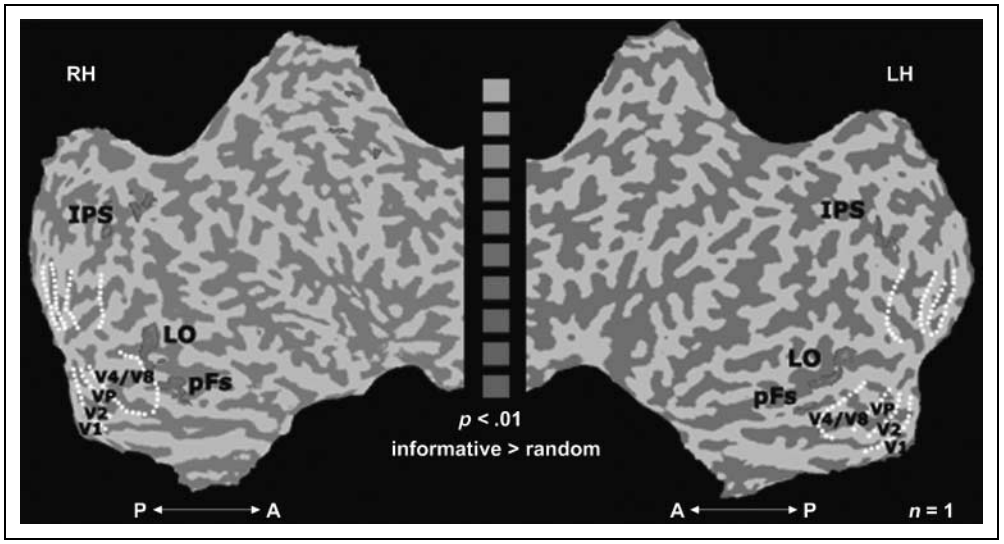


Figure 2. Multisubject fMRI activation (n = 13). (Top) Inflated and flattened cortical maps showing all the regions which demonstrated preferential activation to highly ''informative'' object fragments (test ''informative'' vs. ''random,'' random effect, $p < .05$). The activated voxels were localized in two major foci: LOC ventrally and IPS dorsally. (Bottom) Preferential activation to ''informative'' versus ''random'' stimuli is shown within each category. Single-category activation is shown only for the ventral object-related areas (zoom-in from the region marked by the orange box on the flattened brain). The color scale indicates significance levels. Dotted lines denote anterior borders of retinotopic areas. pFs = posterior fusiform gyrus; LO = lateral occipital; IPS = intraparietal sulcus; LH = left hemisphere; RH = right hemisphere; A = anterior; P = posterior.

Figure 3. Single-subject activation pattern. Note again remarkable preferential activation to ''informative'' fragments compared to ''random'' presented on flattened brain. The meridian borders are indicated by white dotted lines and denote the retinotopic visual areas. Abbreviations as in Figure 2.



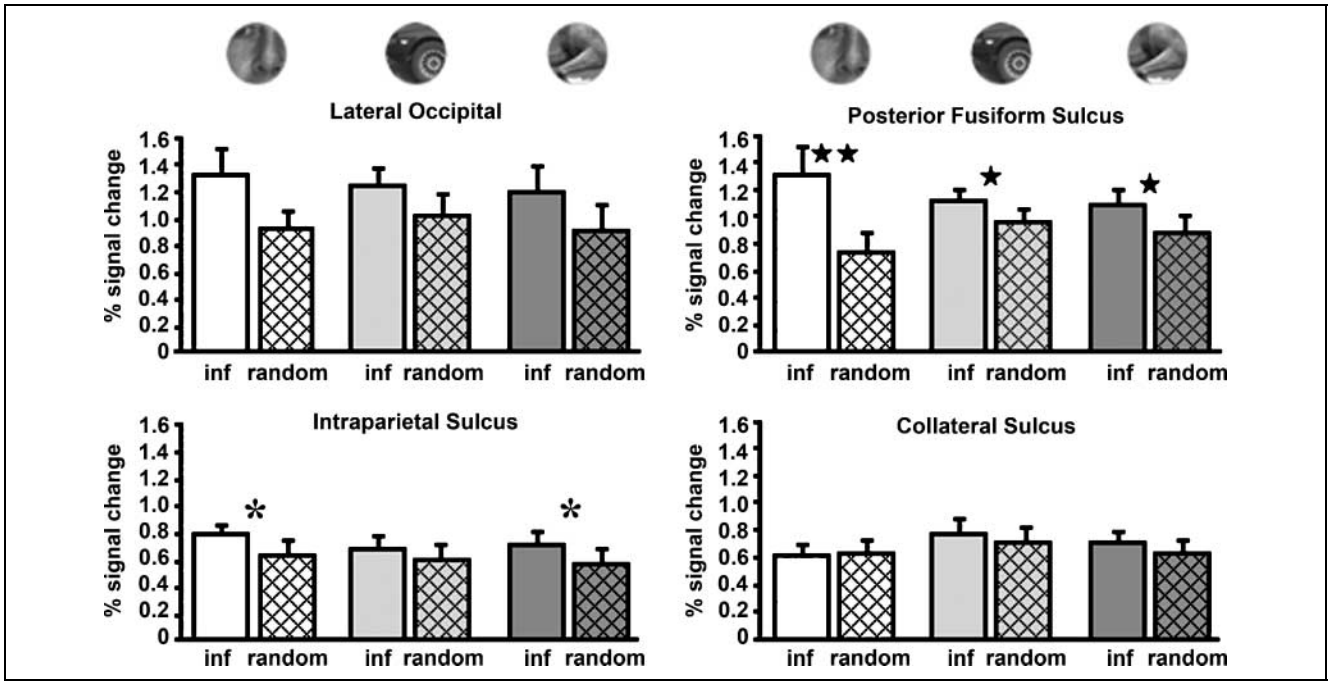


Figure 4. Sensitivity to ''informative'' fragments as reflected in activation profiles. Average activation levels obtained in the high-order visual areas. Object categories (*x*-axis) are indicated in the apertures above the graphs. The *y*-axis denotes an fMRI percent signal change relative to blanks. Asterisks denote a significant difference between image types (''informative''/''random'') within a category calculated by a *t* test: $*p < .05$, $\star p < .005$, $\star\star p < .0005$. No differences were found in the object-related collateral sulcus. Error bars indicate *SEM*. inf = ''informative''.

fragments 0.7 ± 0.05, $p < 2.6e{-}15$). Consistent with this measure, our findings indicate that the randomly selected fragments could be used for recognition, but performance is significantly reduced compared with the ''informative'' fragments.

Relationship between Brain Activity, Behavioral Performance, and Computational Information Measurements

The first experiments established that computationally determined ''informative'' fragments produced increased sensitivity in high-order visual areas. We next examined two additional relationships within our data: (1) physiological versus behavioral, and (2) behavioral versus computational.

To what extent do the fMRI data correlate with the behavioral measurements collected in our study? To answer this question, the percent of incorrect detection was calculated based on experiments performed outside of the magnet by 12 subjects that participated in the fMRI studies (Figure 6A; see Methods for details). Subjects observed the same multiple-fragments stimuli that were shown in the Fragments Experiments 1 and 2. For all three categories, classification performance was significantly higher for ''informative'' compared with ''random'' fragments ($p < .05$). Interestingly, the tendency for somewhat higher activation during the presentation of face stimuli (as observed in fMRI results) correlated with behavioral performance. Namely, the number of the missclassifications for both ''informative'' and ''random'' nonface fragments was higher than those found for the faces fragments.

In the final analysis, we quantitatively compared human behavioral performance in the single-fragment experiment (see Methods) to the performance of our computational learning algorithm. To do this, an error (sum of the misses and false alarms) analysis of both behavioral and computational data was derived for 40 ''informative'' and 40 ''random'' fragments with the highest MI. The reason for this selection was that the ''informative'' fragments we used started with high MI, but as their number increased, we started to get fragments with MI that was not much higher than the ''random'' ones. For human performance, false alarms were calculated as the number of events in which fragments belonging to one class were identified as belonging to another class (e.g., false alarms for faces is the number of times that car and horse fragments were identified as face fragments). In the case of the computational data analysis, false alarms were similarly calculated by treating fragments from the class of interest as the ''target class'' and all other fragments as ''nonclass.'' Results of such error analysis reveal an intriguing correlation between human behavioral performance and the performance of our learning algorithm (Figure 6B and C). Particularly notable is the close similarity of results for ''informative'' fragments. Interestingly, humans' evolution-related bias for the face category appeared only in a smaller number of errors for this category but not in an overall trend for the ''informative''

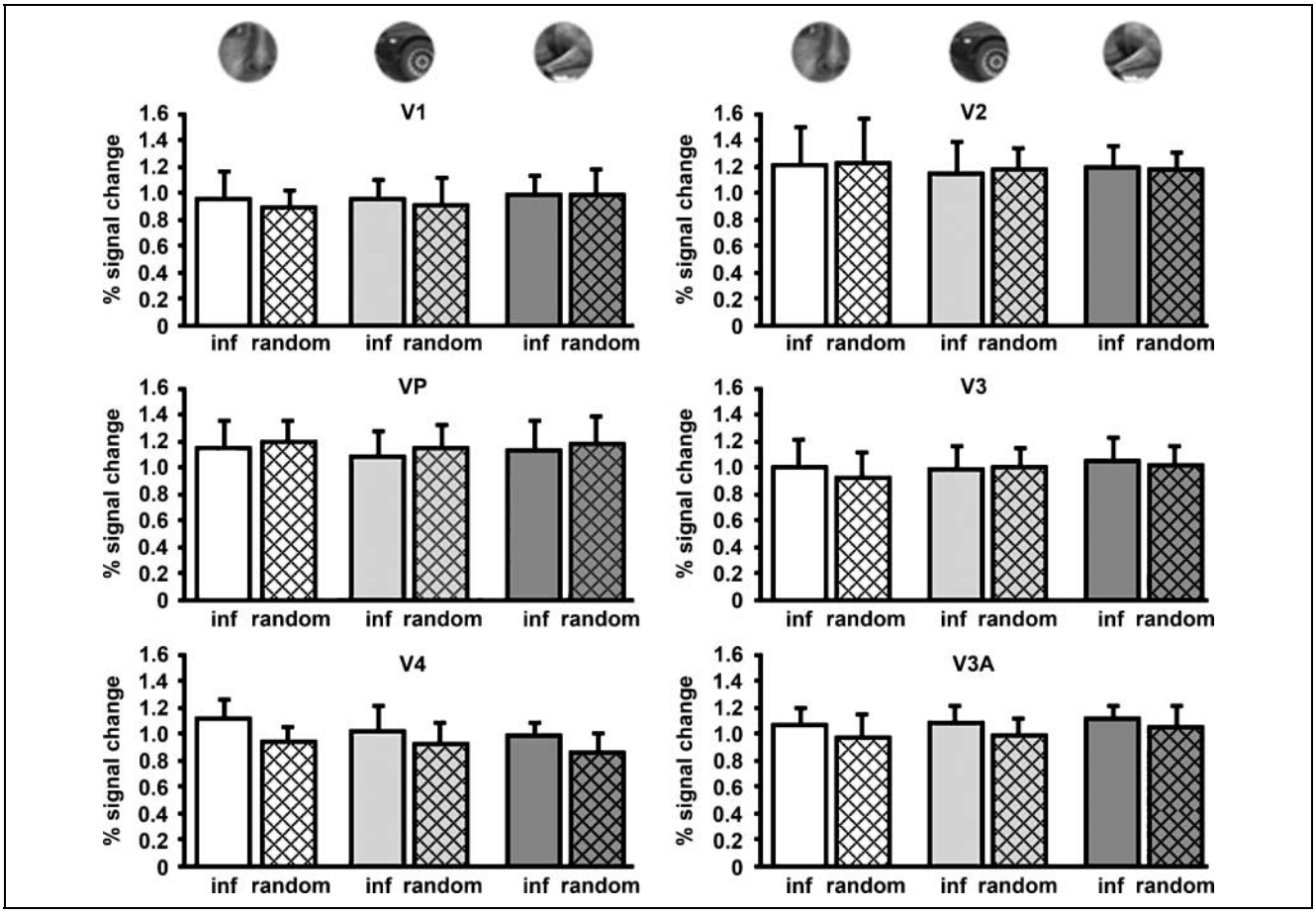


Figure 5. Sensitivity to informative fragments in retinotopic visual areas. Average activation levels obtained in the early visual areas. Object categories (*x*-axis) are indicated in the apertures above the graphs. The *y*-axis denotes an fMRI percent signal change relative to blanks. No favor for ''informative'' fragments was found. Note a trend for preferential activation to ''informative'' fragments in V4 and V3A. Error bars indicate *SEM*. inf = ''informative''.

Table 1. Spectral Power Analysis (Arbitrary Units)

	Informative	*Random*
Image-based Analysis		
Faces	2.87e4 ± 0.07	2.92e4 ± 0.12
Cars	3.36e4 ± 0.04	3.26e4 ± 0.09
Horses	2.83e4 ± 0.04	2.85e4 ± 0.08
Fragment-based Analysis		
Faces	1.8e4 ± 0.6	1.4e4 ± 0.6
Cars	1.3e4 ± 0.5	1.5e4 ± 0.8
Horses	1.6e4 ± 0.7	1.5e4 ± 1.1

A Fourier power-spectral analysis was done on the whole images and on the fragments. First, the overall power in the image/fragment was computed for each image/fragment separately as sum of squared amplitudes in Fast Fourier Transform in the entire frequency domain. Then, a mean across images or fragments was computed. Note no significant differences between the ''informative'' and ''random'' fragments.

fragments favor (our subjects were not the car or horse experts).

Potential Low-level Effects

Was the observed level of differential activity determined by the fragment's informative content, or by some additional image property that may have been inadvertently correlated with the difference between ''informative'' and ''random'' fragments? Thus, it could be that other differences between ''informative'' and ''random'' stimuli, which were not directly related to the informativeness per se, may have accounted for the observed activity. One approach to examine this possibility is through a parametric manipulation of the information level. To this end, we conducted an experiment using three different informativeness levels—high, mid, and low for the face fragments (Fragments Experiment 3). Results were obtained in eight subjects and are depicted in Figure 7.

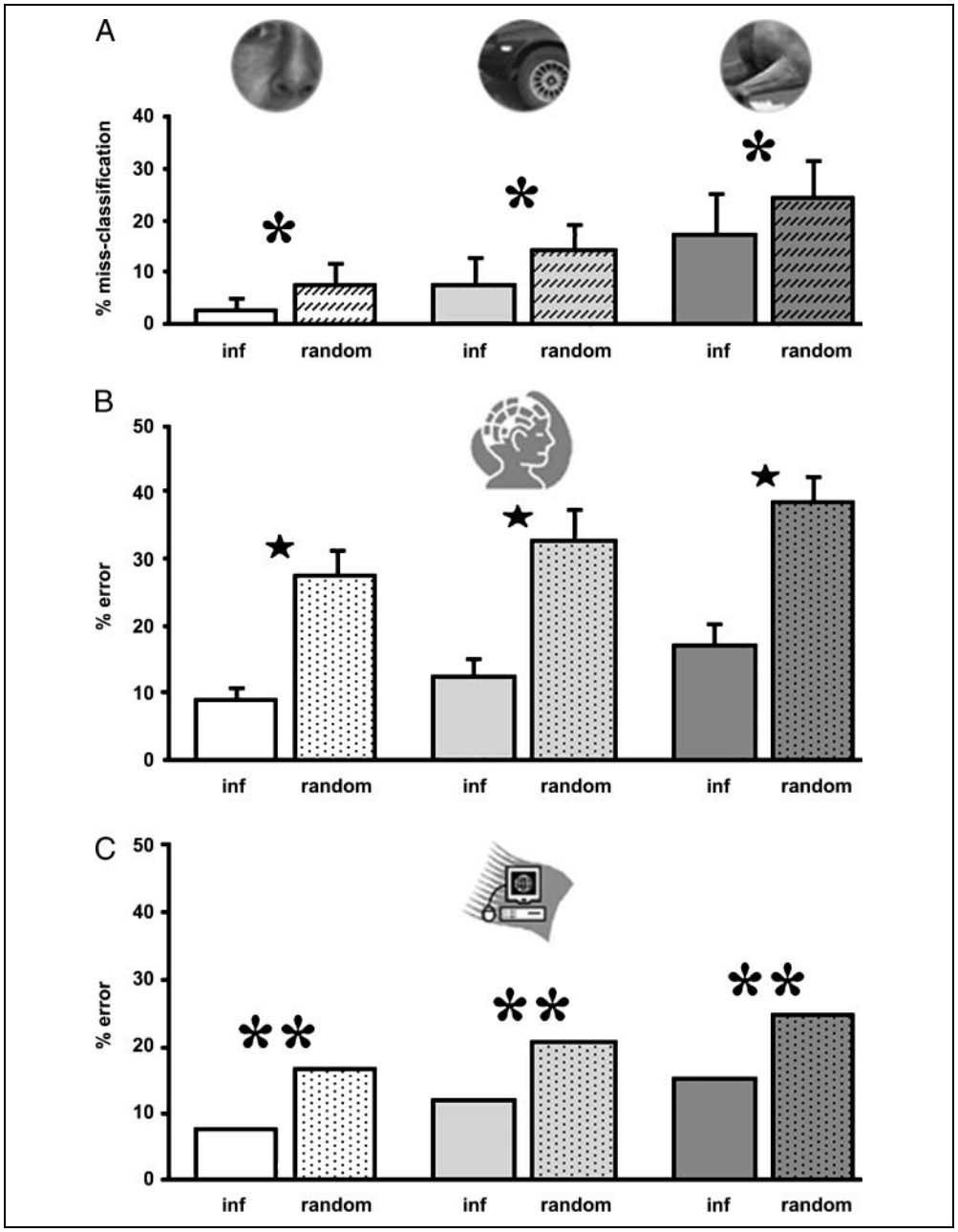


Figure 6. Recognition performance during different conditions. (A) Incorrect classification response. Twelve subjects which participated in the imaging studies were presented with the same multiple-fragments stimuli that were shown in the Fragments Experiments 1 and 2. Number of the missclassifications increases for ''random'' fragments in each category. (B) Human performance ($n = 11$) in the single-fragment experiment for ''informative'' and ''random'' fragments. (C) Computer performance showing the errors for ''informative'' and ''random'' fragments. Note the intriguing similarity between human and computer performance for ''informative'' fragments. The categories are shown in the apertures above the graphs. Error bars indicate *SEM*. *$p < .05$, **$p < .005$, ⋆$p < .0005$. inf = ''informative''.

As can be seen, a monotonic decrease in activity in the LOC could be determined as a function of fragment informativeness. A significant systematic reduction in the activity level was found with decrease of informativeness (Figure 7B). Note that this reduction was stronger in the pFs compared to LO (activation patches are shown in Figure 7C).

To further examine the possibility that low-level image features such as edges or power spectrum did not contribute to the preferential activation of the ''informative'' fragments, we conducted a control experiment (Control Experiment 1). Here we used an image scrambling approach (Lerner et al., 2001; Grill-Spector et al., 1998). Each fragment was broken into nine randomly scrambled parts. Activation profiles, achieved using the internal-localizer method (see Methods) and averaged across eight subjects are shown in Figure 8. As before, in the unscrambled condition, the ''informative'' fragments produced a significantly higher activation ($p < .05$) compared to the ''random'' ones. Scrambling of the ''informative'' fragments produced a significant ($p < .005$) reduction in activation, and, importantly, the preferential activation of the ''informative'' fragments compared to the ''random'' fragments disappeared upon scrambling. Note that inverting the fragments did not abolish the preferential activation to ''informative'' fragments, which is in line with the rather weak inversion effects found in high-order visual areas (Yovel & Kanwisher, 2005). In agreement with our previous research (Lerner et al., 2001), early retinotopic visual areas (V1, for example) demonstrated

preferential activation to the scrambled condition. However, no differences between ''informative'' and ''random'' conditions were observed in this area.

Configurational Effects

An important issue with regard to the differential activation in high-order object areas is the involvement of configurational ''holistic'' effects that may go beyond the local information content of individual fragments. Holistic effects, which depend on the presence of several parts in the correct configuration, may arise in the fragment-based scheme in two ways. First, ''informative'' fragments are found at different scales and resolutions including large, low-resolution ones (Harris & Stephens, 1988), and the presence of such fragments is sensitive to the overall object configuration. Second, individual fragments can be combined in a hierarchical scheme in a manner that preserves their qualitative spatial relation, resulting in stronger activation for correct configurations (Harris & Stephens, 1988). To examine the effect of configurations, we presented equally ''informative'' fragments arranged either in a face-like configuration or in a ''jumbled'' one (see Figure 7, Fragments Experiment 3). Using an internal localizer approach (see Methods), for each subject, we sampled ROIs that demonstrated a preferential activation to high ''informative'' fragments. No significant interaction effect between information level and the type of fragment was found in any of the ROI. Our results showed a significant increase

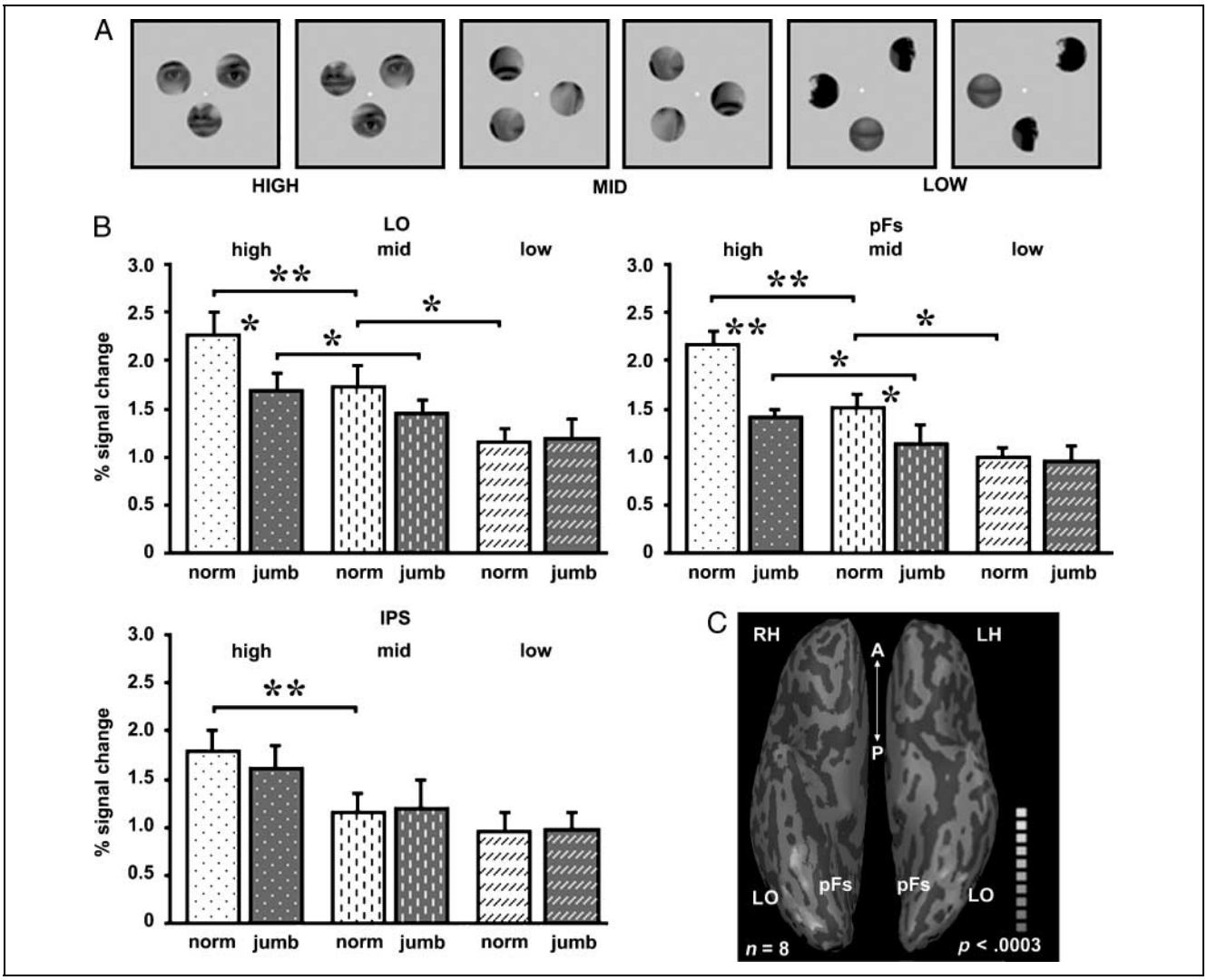


Figure 7. Fragments Experiment 3. (A) Examples of stimuli used in the experiment. ''Informative'' and ''random'' face fragments were presented in the normal (norm) and jumbled (jumb) configuration in three ''informative'' levels: high, mid, and low. (B) Activation profiles revealed in the high informative versus low informative contrast in the high-order visual areas. Asterisks denote a significant difference between conditions calculated by a t test: $*p < .05$, $**p < .005$. Error bars indicate *SEM*. (C) Multisubject activation ($n = 8$) shown on the inflated brain. The activated voxels in two major foci in the ventral stream LO and pFs demonstrate the preferential activation to highly ''informative'' object fragments (test ''high'' vs. ''low''). The color scale indicates significance levels. pFs = posterior fusiform gyrus; LO = lateral occipital; LH = left hemisphere; RH = right hemisphere; A = anterior; P = posterior.

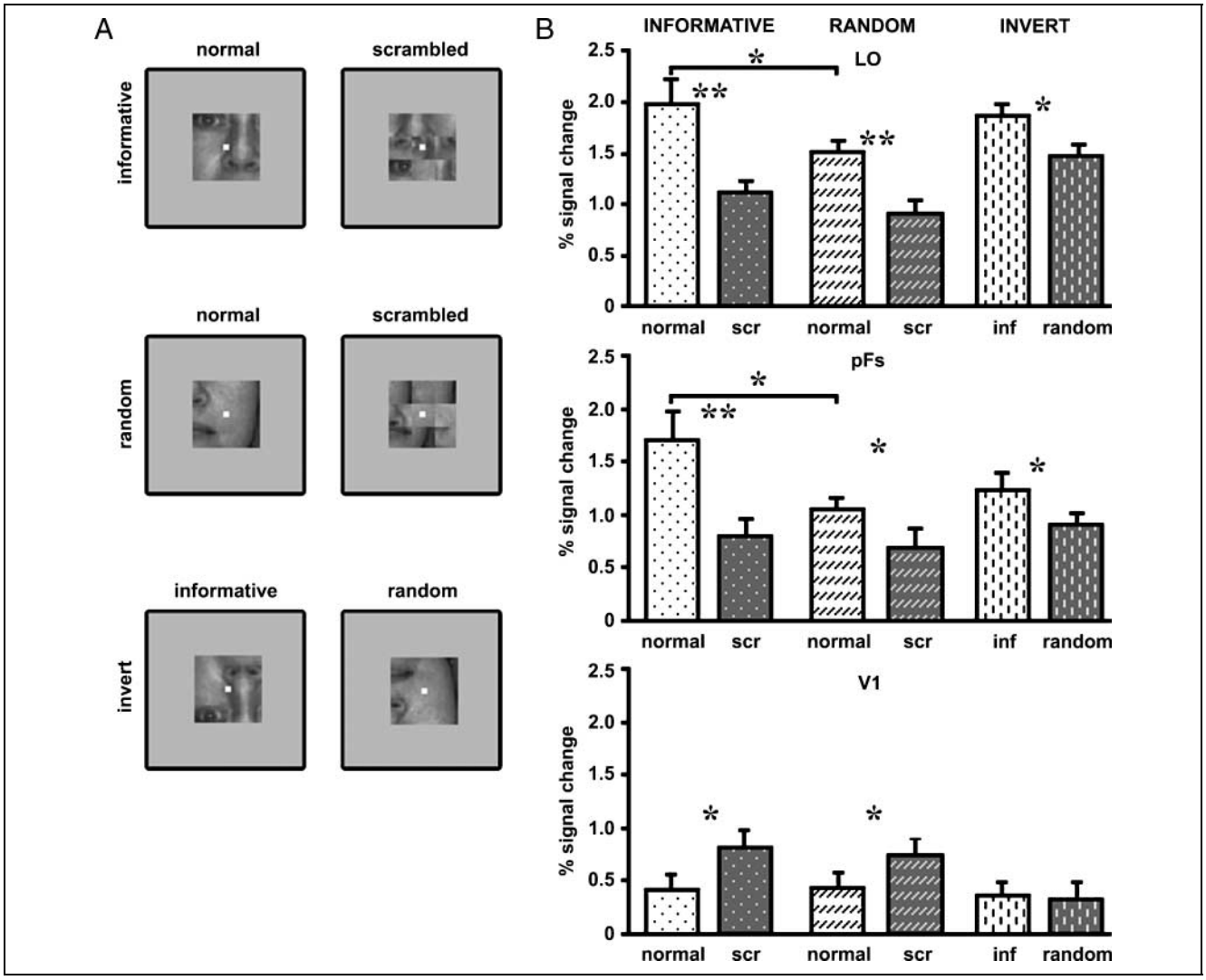


Figure 8. Control Experiment 1: stimuli and average activation level (n = 8). (A) Examples of visual stimuli used in the experiments. Two types of face fragments—''informative'' and ''random''—were presented in the normal view or scrambled into nine parts. Additionally, ''informative'' and ''random'' fragments were shown in the inverted mode. (B) Average activation levels obtained in the high-order visual areas LO and pFs and retinotopic area V1 during comparing the normal and scrambled conditions. Only statistically unbiased epochs are shown (see Internal Localizer Approach section). Note the absence of preferential sensitivity to ''informative'' conditions compared to ''random'' in V1. The y-axis denotes an fMRI percent signal change relative to blanks. Asterisks denote a significant difference between conditions calculated by a t test: *$p < .05$, **$p < .005$. Error bars indicate *SEM*. inf = ''informative''.

in activation for the normal, compared to the jumbled, configurations for the high ''informative'' fragments in the LO ($p < .05$) and the pFs ($p < .005$), as well as for the mid-''informative'' level in the pFs ($p < .05$) (Figure 7B).

We used another control experiment (Control Experiment 2) to determine if recognition performance depends on whether the fragments came from the same source or multiple face images. The comparison was made using two different fragment configurations. In the first configuration, we evaluated performance on typical face templates consisting of either three ''random'' fragments selected from a single face or three ''informative'' fragments selected from different faces and placed in a normal face configuration. The combined configuration appeared slightly distorted as the eye regions were taken from different faces. The second configuration compared performance on seven randomly arranged ''informative'' fragments selected from either a single face or different faces. Activity levels measured in six subjects and their corresponding stimuli examples are illustrated in Figure 9 (bottom and top panels, respectively). A comparison of Fragments Experiments 1 and 2 suggests that performance in these experiments did not depend on whether the object fragments were selected from a single image or multiple images. Our results revealed a greater activation to the few (3) aligned face windows than to many (7), mostly misaligned windows. However, no significant difference was found between images of seven fragments belonging to a single face and those belonging to a range of different faces (pFs; % signal change: 0.82 ±

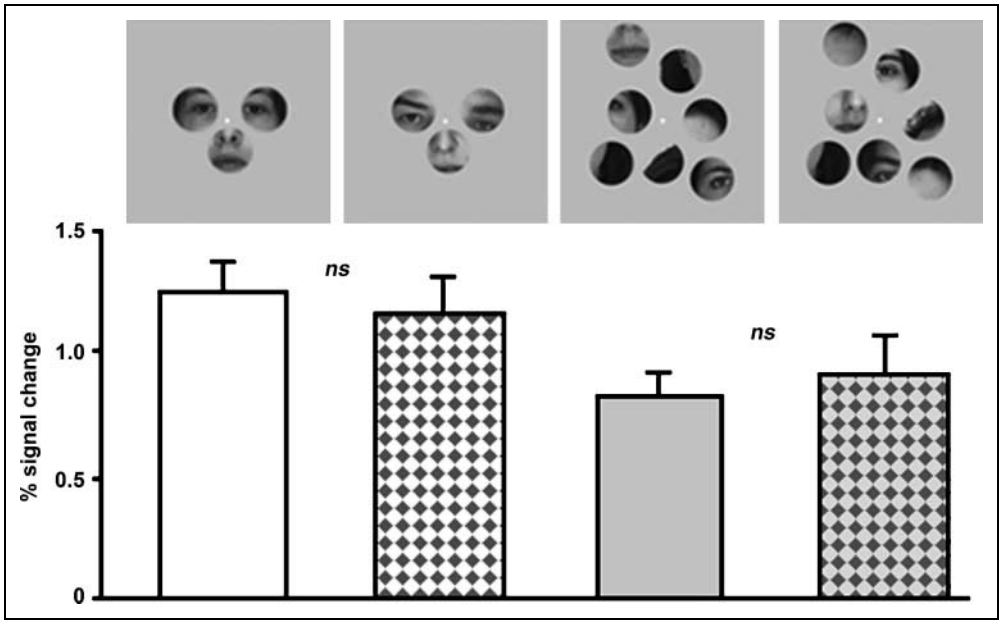


Figure 9. Control Experiment 2: stimuli and average activation level (n = 6). Activation profiles measured in pFs to the different conditions shown on the bottom panel. The y-axis indicates an fMRI percent signal change relative to blanks. No significant difference was found between the condition composed of single-face fragments and the condition consisted of different faces fragments.

0.11 vs. 0.91 ± 0.16; p = .1), although activation was slightly higher in the condition with different faces fragments. Also, no effect related to three fragments selected from either a single face or different faces was found (pFs; % signal change: 1.25 ± 0.12 vs. 1.16 ± 0.15; p = .2). Thus, visually perceived face distortions were not reflected by the activation level in this area. Additional effects of fragments configuration were further tested below.

Potential Nonvisual Effects

Is it possible that the parietal activation reflects differences in the attention of the subjects across conditions? One outcome of heightened attention to the ''informative'' fragments may be a better performance on the 1-back memory tasks the subjects were required to perform. Note that this task is not an object categorization task and was aimed solely to assess attentional effects. To examine this possibility, we collected the subjects' responses while they indicated the stimulus identity during the scan. Analysis of these data revealed that average errors for the ''informative'' and ''random'' conditions were very close (1.0 ± 0.8% and 0.9 ± 0.8%, correspondingly) as well as average reaction time in these conditions (481 ± 58 and 499 ± 53 msec). Moreover, no correlation was discerned between the normalized BOLD signal and average errors number (r^2 = .001). Thus, it is reasonable to assume that attentional differences to the different types of fragments were not dominant.

It could be argued that the ''informative'' fragments, which carry the most interesting information about the objects, may have been associated with more intensive eye movements. This could potentially produce a stronger cortical activation compared to ''random'' selected fragments. To check this possibility, we recorded, in the MRI scanner, subjects' eye movements while they observed fragments with high and low information content and compared their associated eye movements. Figure 10 shows two-dimensional clusters of fixation points in red and blue (corresponding to high and low informativeness, respectively). The results of these measurements show similar patterns of eye movements during epochs with different information levels. Formally, kernel density estimation, using Gaussian kernel with a bandwidth equal to pupil size, showed no difference for fixation densities between conditions for the same subject—at each location the difference was smaller than precision of double variables ($<10^{-15}$). Therefore, the difference in cortical activation is unlikely to be solely explained by eye movements.

DISCUSSION

The present study tested the relations between predictions raised by computational considerations and human vision, regarding features used for object classification. In particular, the goal was to test whether features (composed of object fragments) selected, through an objective computational measure, to maximize the information they deliver for classification also play a preferred role in human vision.

The main outcome of this study is the intriguing demonstration that a completely objective criterion for object fragment selection results in a higher activation of human object areas. Specifically, by comparing the activation to highly ''informative'' fragments with activation to similar but less ''informative'' ones (see Figure 1), we have discovered that ''informative'' fragments produced a preferential activation in human object-related areas.

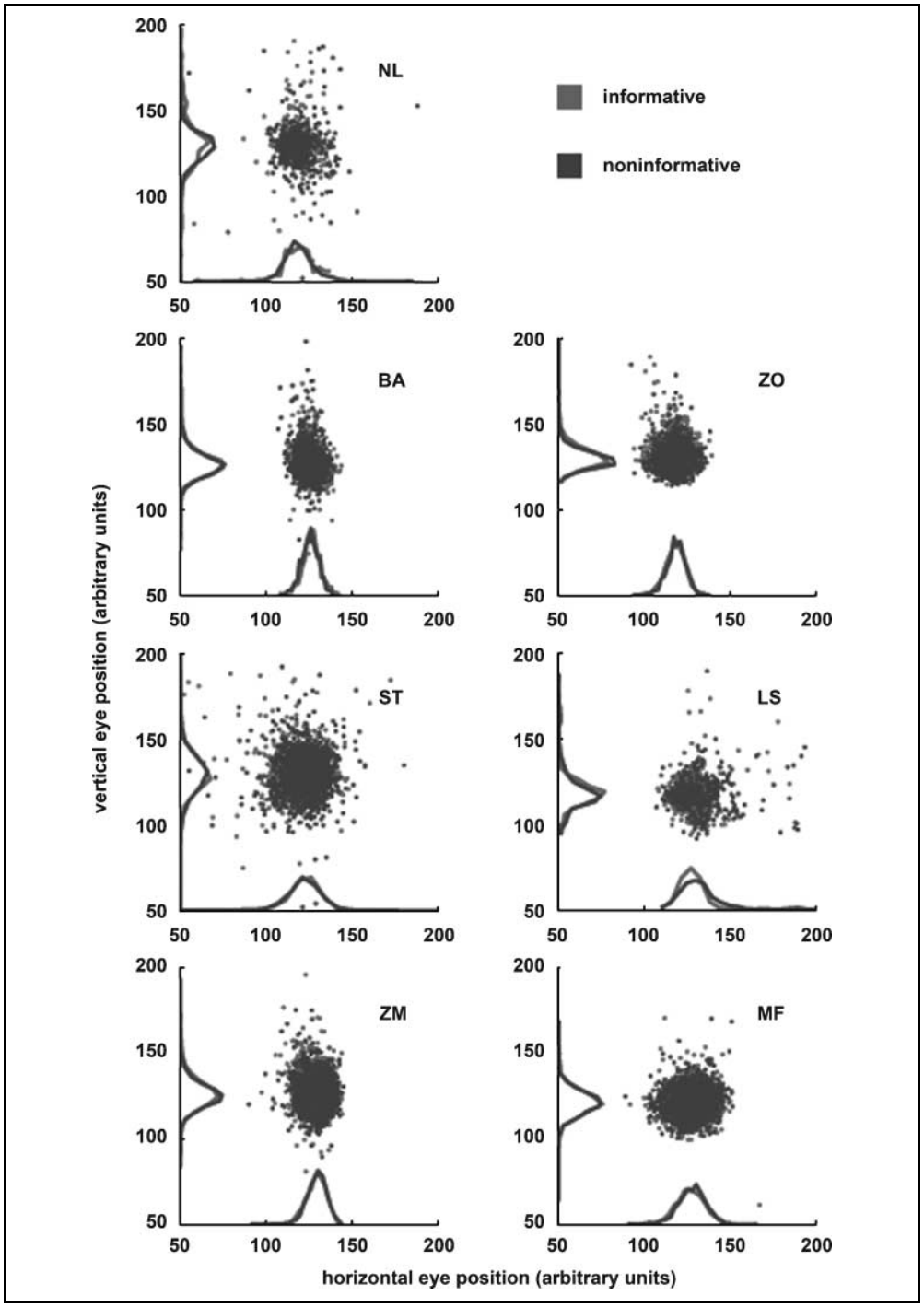


Figure 10. Eye-movements data analysis. Two-dimensional distribution of fixation points is presented for seven different subjects. The data for high and low informative conditions are shown in red and blue, respectively. The horizontal and vertical projections are shown along axes.

Our control experiments argue against the possibility that the preferential activation was due to nonvisual effects such as attention effects or eye movements. Thus, it should be emphasized that the task (1-back) was of a similar attentional demand for both the ''informative'' and less ''informative'' fragments. Although we cannot fully rule out the possibility that the ''informative'' fragments led to a reflexive enhancement of attention, the fact that the preferential activation was most evident in specific regions (the pFs; see Figure 4) argues for a selective, recognition-related effect, rather than a more global attentional enhancement.

The highly ''informative'' fragments in our study were usually also more recognizable, and a possible question to consider is therefore the relation between these factors and their contribution to the increased fMRI activation. The empirical findings show that recognizability and informativeness are, in fact, highly correlated (although it remains possible that recognizability may be affected by additional parameters). This correlation is expected from

the working hypothesis of the study, regarding the causal relationship between informativeness and recognizability. The findings that psychophysical recognizability and fMRI activation in high-order visual areas are closely related to our simple and objective information measure suggest that the algorithm is successful in capturing essential parameters of the recognition processes. To our knowledge, this study is the first to demonstrate such effects following fragment selection, which is based on a general, strictly objective, criterion that does not depend on prior assumptions concerning the shape or low-level features of the stimulus categories.

These results are consistent with general expectations because an image region which is objectively informative can also be expected to be useful in human vision. It is not clear, however, that the simple information measure used in the computations would be sufficient for predicting human performance. Information can be measured mathematically between any different variables, and the theory needs to specify which variables to extract and use. The computation we used was based on a simple measure of using the fragment directly as a feature, based on its frequency within and outside the class of interest (''patch information''). A tight correlation between recognition and this patch information is not expected a priori, and was not proposed and tested prior to the current study. In different theories of recognition, such as geon-based (Biederman, 1987), which uses 3-D shape primitives, or eigen-faces (Turk & Pentland, 1991), which uses global object templates, such information measure will not be predictive of performance. The patch-information measure used here allows us for the first time to directly compare, for example, a part of a horse's torso with a piece of a leg, and predict their usefulness for recognition; the empirical results are consistent with these predictions.

In the present study, highly significant preferential activation for ''informative'' fragments was found in both LO and pFs subregions of the LOC, the typical object-related areas. These results were consistent for different object categories, such as faces, horses, and cars. Interestingly, sensitivity in the fusiform gyrus also for the nonface objects points to preference for faces (Kanwisher, McDermott, & Chun, 1997) as well as to aspects of distributed representation (Haxby et al., 2001).

The behavioral tests also showed an advantage of the ''informative'' fragments in recognition performance. Interestingly, a comparison of human recognition measurements to the performance of the computational learning algorithm revealed a strong correlation for all ''informative'' fragments, and especially for the face fragments. These results demonstrate that processes which have been shown to be highly effective in computational simulations can be used to predict effective stimulus parameters applicable to human vision.

Although our results demonstrate the importance of individual informative class-specific features to human recognition, it is important to note also the role of configuration effects that were observed. The results of both control experiments and, particularly, Fragments Experiment 3 have shown that configurational effects, that is, the arrangement of the fragments relative to each other, contributed to the activation level. Thus, ''jumbling'' the location of the ''informative'' fragments relative to each other, without changing the contents of each fragment, resulted in a significant reduction in activation (see Figure 7). This suggests that, in addition to the local information content of the fragments, a more ''holistic'' component played a role. However, it should be emphasized that such holistic configurational effects may also be the result of using larger, and perhaps, low-resolution ''informative'' fragments which encompass large parts of the object image, with reduced information about image detail (Ullman et al., 2002). Additional possibilities are that the ''informative'' fragments may be better inducers of completion (Lerner, Harel, & Malach, 2004), or high-order grouping (Hasson, Hendler, Ben Bashat, & Malach, 2001) effects. Given the rich and complex nature of human object recognition, it is likely that high-order object areas use both the individual features as well as a more holistic aspect that depends on their relative position. Such a representation is, in fact, compatible with fragment-based models (Epshtein & Ullman, 2005) and with the partial completion effects observed in the past (Lerner et al., 2004), as well as with the fMRI reports of holistic processing (Imber, Shapley, & Rubin, 2005; Altmann, Bulthoff, & Kourtzi, 2003; Kourtzi, Erb, Grodd, & Bulthoff, 2003; Stanley & Rubin, 2003).

Previous studies showed that high-level object-related areas are activated by whole objects as well as by their large parts (Lerner et al., 2001; Grill-Spector et al., 1998). The gradual decline in activation to fragmented objects may be related to receptive field size as well as to stimulus selectivity. Namely, the object fragments falling in the receptive fields within this region need to contain a large portion of the object to produce the activation comparable to activation elicited by whole objects.

Different object classes differ in the resolution and size distributions of the optimal object fragments (Ullman et al., 2002). For face images, optimal fragments for recognition are usually of intermediate size at high resolution, whereas car images have typically a higher proportion of larger optimal fragments, sometimes at intermediate resolution. Interestingly, it has been recently proposed that a global organizing principle of human object recognition areas may be associated with focal versus global vision. More specifically, cortical areas which have been associated with face processing, such as the fusiform face area (FFA), appear to deal with detailed, high-resolution processing, and consequently, receive a more extensive input from foveal representation, whereas areas associated with buildings and scenes are associated with processes of large-scale integration and, as such, are more related to global, peripheral visual field representa-

tions (Hasson, Harel, Levy, & Malach, 2003; Malach, Levy, & Hasson, 2002; Levy et al., 2001). It is interesting to note that from an objective set of principles, it naturally emerges that, although both classes include a range of different features, faces are associated with a higher proportion of fragments requiring higher resolution vision, compared with cars which are associated with a larger proportion of more global fragments.

In addition to the ventral stream activation in the occipito-temporal cortex, our results also show consistent activations in parietal and prefrontal cortices. It is interesting to note in this respect that selective object activation has been consistently reported in dorsal stream parietal areas occupying similar anatomical locations to the ones we found (Shmuelof & Zohary, 2005). A likely interpretation of the parietal activation might be the involvement of these regions with action planning (Culham & Valyear, 2006). It could be that the enhanced activation to ''informative'' fragments in the parietal cortex is related to mechanisms of target detection and object manipulations associated with such action planning roles. Prefrontal object activation has been extensively studied in the context of categorization (Freedman, Riesenhuber, Poggio, & Miller, 2003) as well as image understanding (Palmeri & Gauthier, 2004) and that the preferential activation found in the present study may be related to such high-level cortical functions which go beyond the merely perceptual stages.

Acknowledgments

R. M. was funded by Benoziyo Center for Neurological Disorders, The Dominique Center, The Minerva Foundation, and The Israeli Science Foundation. S. U. was supported by Grant 7-0369 from the Israeli Science Foundation and EU Daisy Grant 015803. Y. L. was funded by Charles Clore Fellowships. We thank the Wohl Institute for Advanced Imaging in the Tel Aviv Sourasky Medical Center. We thank S. Gilad, S. Gilaie-Dotan, and M. Katkov for fruitful discussions and comments on the manuscript. We thank M. Harel for help with the brain reconstructions, O. Levin for help with eye movement measurements, and E. Okon for technical assistance.

Reprint requests should be sent to Rafael Malach, Weizmann Institute of Science, Rehovot, 76100 Israel, or via e-mail: rafi.malach@weizmann.ac.il.

REFERENCES

Agarwal, S., & Roth, D. (2002). *Learning a sparse representation for object recognition*. Paper presented at the 7th ECCV.

Altmann, C. F., Bulthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology, 13,* 342–349.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Cover, T., & Thomas, J. (1991). *Elements of information theory*. New York: Wiley.

Culham, J. C., & Valyear, K. F. (2006). Human parietal cortex in action. *Current Opinion in Neurobiology, 16,* 205–212.

Epshtein, B., & Ullman, S. (2005). *Feature hierarchies for object classification*. Paper presented at the ICCV.

Fergus, R., Perona, P., & Zisserman, A. (2003). *Object class recognition by unsupervised scale-invariant learning.* Paper presented at the Proc. of the IEEE Conf on Computer Vision and Pattern Recognition.

Fleuret, F. (2004). Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research, 5,* 1531–1555.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic-resonance-imaging (fMRI)—Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33,* 636–647.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience, 23,* 5235–5246.

Friston, J., Homes, A., Worsley, K., Poline, J., Frith, C., & Frackwowiak, R. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping, 2,* 189–210.

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *Neuroimage, 10,* 385–396.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360,* 343–346.

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science, 259,* 100–103.

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., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping, 6,* 316–328.

Harris, C., & Stephens, M. (1988). *A combined corner and edge detector*. Paper presented at the Fourth Alvey Vision Conference, Manchester.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron, 37,* 1027–1041.

Hasson, U., Hendler, T., Ben Bashat, D., & Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. *Journal of Cognitive Neuroscience, 13,* 744–753.

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 ventral temporal cortex. *Science, 293,* 2425–2430.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London), 195,* 215–243.

Imber, M. L., Shapley, R. M., & Rubin, N. (2005). Differences in real and illusory shape perception revealed by backward masking. *Vision Research, 45,* 91–102.

Kanwisher, N., Chun, M. M., McDermott, J., & Ledden, P. J. (1996). Functional imagining of human visual recognition. *Brain Research, Cognitive Brain Research, 5,* 55–67.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kourtzi, Z., Erb, M., Grodd, W., & Bulthoff, H. H. (2003). Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cerebral Cortex, 13,* 911–920.
Lerner, Y., Harel, M., & Malach, R. (2004). Rapid completion effects in human high-order visual areas. *Neuroimage, 21,* 516–526.
Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex, 11,* 287–297.
Lerner, Y., Hendler, T., & Malach, R. (2002). Object-completion effects in the human lateral occipital complex. *Cerebral Cortex, 12,* 163–177.
Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature Neuroscience, 4,* 533–539.
Levy, I., Hasson, U., Harel, M., & Malach, R. (2004). Functional analysis of the periphery effect in human building related areas. *Human Brain Mapping, 22,* 15–26.
Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60,* 91–100.
Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences, 6,* 176–184.
Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A., 92,* 8135–8139.
Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, Series B, Biological Sciences, 200,* 269–294.
Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5,* 58–70.
Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5,* 291–303.
Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.
Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21,* 409–418.
Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *Journal of Neuroscience, 10,* 3369–3389.
Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268,* 889–893.
Shmuelof, L., & Zohary, E. (2005). Dissociation between ventral and dorsal fMRI activation during object and action recognition. *Neuron, 47,* 457–470.
Stanley, D. A., & Rubin, N. (2003). fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron, 37,* 323–331.
Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71–86.
Ullman, S., & Sali, E. (2000). Object classification using a fragment-based representation. In S. Lee, H. Bűlthoff, & T. Poggio (Eds.), *Biologically motivated computer vision* (pp. 73–87). First IEEE International Workshop, BMVC 2000, Seoul, Korea. Proceedings Lecture Notes in Computer Science 1811, Springer.
Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687.
Wertheimer, M. (1938). *A source book of Gestalt psychology*. London: Routledge & Kegan Paul.
Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology, 15,* 2256–2262.