A Hierarchical Axis of Object Processing Stages in the Human Visual Cortex

Yulia Lerner, Talma Hendler1, Dafna Ben-Bashat1, Michal Harel and Rafael Malach

Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100 and 1Sourasky Medical Center, Tel Aviv 64239, Israel

How are objects represented in the human visual cortex? Two conflicting theories suggest either a holistic representation, in which objects are represented by a collection of object templates, or a part-based representation, in which objects are represented as collections of features or object parts. We studied this question using a gradual object-scrambling paradigm in which pictures of objects (faces and cars) were broken in a stepwise manner into an increasing number of blocks. Our results reveal a hierarchical axis oriented anterior–posteriorly in the organization of ventral object-areas. Along this axis, representations are arranged in bands of increasing sensitivity to image scrambling. The axis starts in early visual areas through retinotopic areas V4/V8 and continues into the lateral-occipital sulcus dorsally and the posterior fusiform girus ventrally, corresponding together to the previously described object-related lateral occipital complex (LOC). Regions showing the highest sensitivity to scrambling tended to be located at the most anterior-lateral regions of the complex. In these more anterior regions, breaking the images into 16 parts produced a significant reduction in activation. Interestingly, activation was not affected when images were cut in two halves, either horizontally or vertically. Car images generally produced a weaker activation compared to faces in the lateral occipital complex but showed the same tendency of increased scrambling sensitivity along the anterior–posterior axis. These results suggest the existence of a hierarchical axis along ventral occipito-temporal object-areas, in which the neuronal properties shift from sensitivity to local object features to a more global and holistic representation.

Introduction

The way in which objects are represented in high-order visual areas has been a point of controversy despite a large number of experimental and modeling studies (Edelman, 1999; Tarr and Büthoff, 1999). One central issue that remains unresolved is to what degree objects are represented holistically in a manner that captures the global, configurational aspects of the object, and to what extent such a representation is dominated by object parts or fragments (Perrett et al., 1982). Another related question is whether face representation may have uniquely holistic characteristics compared to other object classes (Toovey, 1998).

Single neuron recordings in inferior-temporal cortex of the macaque provide support for both types of representations. Using the reduction method it has been shown that a large proportion of neurons in the inferior-temporal cortex of the macaque are sensitive to ‘moderately complex’ object features (Tanaka et al., 1991). On the other hand, other reports point to the existence of neurons sensitive to a more holistic representation – particularly of faces (Desimone, 1991). Further support for a more abstract object representation comes from the finding of neurons sensitive to object-centered representations in frontal areas (Olson and Gettner, 1995). The delineation of object-related areas in the human visual system (Malach et al., 1995; Kanwisher et al., 1996) made it possible to characterize the nature of object representations using neuroimaging methods. Several studies reported localized selectivity to specific object categories, particularly faces (Clark et al., 1996; Kanwisher et al., 1997; McCarthy et al., 1997; Epstein and Kanwisher, 1998; Ishai et al., 1999). Such localization might be interpreted as indicating a holistic representation; however, this is not necessarily the case, e.g. it may well be that a localized face representation consists of aggregates of neurons sensitive to various moderately complex facial features.

Indeed, in a recent experiment we have found that the majority of voxels in non-retinotopic object areas located anterior-laterally in the occipital lobe (the lateral occipital complex, LOC) remain active when pictures of a variety of objects are broken into scrambled blocks, provided that the blocks are sufficiently large (Grill-Spector et al., 1998). This finding suggests that a part of the object representations involves neurons, which can be activated by object features. However, the lateral occipital complex is a large and heterogeneous (Grill-Spector et al., 1999) region and it will be of interest to explore to what extent sensitivity to image scrambling varies in a systematic way within subregions of LOC and in relation to neighboring visual areas.

In the present experiment we have found evidence for such organizing principle in that sensitivity to image scrambling significantly increases along an anterior–posterior axis – this trend begins in retinotopic areas and proceeds into the LOC. We have documented this increase for two object types – faces and cars – indicating that this phenomenon generalizes across different object categories. Thus, our results provide evidence for a hierarchical organization within non-retinotopic object areas of the human visual cortex.

Materials and Methods

MRI Acquisition

Blood oxygenation level dependent (BOLD) contrast was obtained with gradient-echo planar imaging (EPI) sequence (TR = 3000, TE = 55, FA = 90, imaging matrix = 80 × 80, FOV = 24 cm) on a 1.5 T Signa Horizon LX 8.25 GE echo-speed scanner. The MRI system was equipped with 22 mT/m field gradients with a slew rate of 120 T/m/s. 17 functional slices of 4 mm thickness and 1 mm gap were localized in an oblique-coronal orientation, perpendicular to the calcaneous sulcus, and covered the occipital, posterior parietal and posterior temporal lobes. Spin-echo TE weighted anatomical images were acquired correspondingly. Three-dimensional spoiled-gradient echo sequence was acquired on each subject, in order to allow volume statistical analyses of signal changes during the experiment.

Experimental Procedure

Subjects

Nine healthy volunteers (four women and five men, ages 27–49) participated in the experiments (seven subjects in the face-scrambling experiment and six subjects in the car-scrambling experiment, four of
them took part in both). To avoid repetition effects each experiment was presented once per subject. All subjects were given detailed instructions of the experiment both outside and inside the scanner before the actual scan and provided written informed consent approved by the Tel Aviv Sourasky Medical Center.

**Stimuli**

Visual stimuli were generated by a dedicated PC located in the control room and projected via a LCD projector (Epson MP 7200) onto a transparent tangent screen positioned over the subject’s forehead, and viewed through a tilted mirror fixed to the head coil. Stimuli consisted of achromatic photographs of faces or cars presented in a frontal view or progressively broken into 2–256 blocks and randomly scrambled. Figure 1a illustrates the eight types of image-epochs used in the face-scrambling experiment and their number of repetitions (×2, ×3 or ×4) during an experimental run. In the ‘face-scrambling’ experiment face images were broken into 2 vertical parts, 2 horizontal parts, 4, 16, 64 and 256 scrambled parts. In addition, an epoch containing car images was included for comparison. In the car-scrambling experiment (Fig. 8a) car images were broken into 2 vertical parts and into 4, 16, 64 and 256 scrambled parts, and epochs of intact face images were added for comparison. High frequency edges were smoothed using a two-dimensional Gaussian function, which was calculated as follows. First, the horizontal stripes of a scrambled image (near the edges) were convolved with Gaussian (σ = 0.05). The width of those stripes was calculated by the formula: width = 2√V, where V = 64/[number of parts in the scrambled image]1/2. The constant and formula were obtained empirically. A red fixation point was positioned centrally in each picture and blank. The average luminance of the images was 55 cd/m², and of the blanks was 40 cd/m².

An experimental run was composed of 27·8 s long image-epochs, each containing 16 pictures of one category. The image-epochs were alternated with 6 s long blank-epochs, except for the first and last blanks, which lasted 32 and 10 s respectively. Each image was presented for 300 ms interspersed with 200 ms blanks. Figure 1b illustrates a schematic presentation of the temporal sequence of events during an experiment. In order to control for ordering effects, epoch types (i.e. full images or broken images into 2, 4, 16, 64, 256 parts) were presented in a pseudo-random order. The subjects were asked to categorize all face images, including the scrambled ones, as male or female, and the car images as private or public transport.

**Data Analysis**

All fMRI data were processed using the BrainVoyager 3.9 software package (Goebel et al., 1998a,b) and complementary in-house software. Data analysis was performed separately for each subject. The two-dimensional functional images were superimposed on two-dimensional anatomical images and incorporated into the three-dimensional data sets through trilinear interpolation. The complete data set was transformed into Talairach space (Talairach and Tournoux, 1988). Preprocessing of functional scans included head movement measurements, high-frequency temporal filtering and removal of linear trends. Three-dimensional statistical maps were calculated by correlating the signal time course with a reference (box-car) function for each voxel, with a lag of 9 s (to account for the hemodynamic response delay).

Statistical evaluations were based on an average time course obtained from all voxels within a priori defined regions of interest (ROI). Only voxels whose correlation to the reference function had a value equal or greater than 0.2 were included in the statistical (parametric) map. Early visual areas were defined based on functional and anatomical properties (DeYoe et al., 1996; Tootell et al., 1996). In order to establish the borders between the different early visual areas, functional maps from each experiment were overlapped with a visual field map of the vertical and horizontal meridians for each subject (Sereno et al., 1995). The retinotopic map was obtained in a separate scan while the subject viewed either natural grayscale images or drifting texture-defined shapes. These stimuli proved particularly effective in activating intermediate-level visual areas (Grill-Spector et al., 1999).

Functional maps were superimposed on corresponding anatomical slices and on a flattened cortical surface reconstruction of the subject’s brain. Cortical surface was reconstructed from a three-dimensional spoiled-gradient echo scan. Surface reconstruction was done by segmentation of white matter using a grow-region function and outlining the gray–white matter boundary. The reconstructed surface was ‘inflated’, unfolded, cut along the calcarine sulcus and flattened.

**Measuring recognition performance**

Recognition performance was measured in seven subjects for the face-scrambling experiment. The measurements were taken outside the

---

**Figure 1.** Face-scrambling experimental paradigm. (A) Types of the visual stimuli that were used in the experiment. Numbers under the pictures denote the number of repetitions of the same epoch during a run. Each visual epoch consisted of 16 different images of the depicted type presented at a rate of 0.5 Hz. A red fixation point was presented centrally in all images. Subjects were instructed to categorize the face images as female or male, and the car images as private or public transport. (B) A segment from the time axis of an experimental run of an overall duration of 414 s. The presented sequence indicates the length of different epochs. During a run visual epochs lasted for 8 s and alternated with blanks of 6 s. Each run includes a pre- and post-stimulus blank epochs of 32 and 10 s respectively.

**Figure 2.** Distribution of object areas relative to visual meridian representations. Functional activation maps of objects (faces and cars) are superimposed on the inflated (top in each frame) and unfolded cortical surfaces of the right and left hemispheres from a single subject. Sulcal cortex is coded darker than gyral cortex, the superior temporal sulcus (STS) is indicated to assist in orientation. The borders of retinotopic visual areas are delineated by light green (blue and green) and horizontal (yellow) meridian representations. The non-retinotopic face- and object-selective areas (orange; faces > 256 blocks; red: cars and faces > 256) are labeled LO (lateral occipital) and pFs (posterior fusiform). The orientation of maps relative to brain is denoted by A (anterior), P (posterior), D (dorsal) and V (ventral).

**Figure 3.** Sensitivity to face-image scrambling along the occipito-temporal cortex. Colors represent the sensitivity of object areas to degree of face scrambling. Maps were acquired based on GLM (general linear model) statistical analyses, with activation to full face shown in green, and activation to faces broken into four blocks in orange. Blue and yellow colors indicate more balanced activation between predictors. The main sulci are labeled as COS (collateral sulcus), STS (superior temporal sulcus), ITS (inferior temporal sulcus), OTS (occipito-temporal sulcus). Anterior–posterior direction is denoted by A and P. (a) The functional activation map is superimposed on an inflated brain (showing right hemisphere). (b) The functional activation map is presented on an axial slice. On the right, activation profiles from clusters showing different levels of scrambling. Colored arrows point to specific clusters.
magnet, more than 2 months after the scan. Visual stimuli and task (gender identification) were similar to the fMRI scan with the exception that the interstimulus intervals were extended by 1 s to allow subjects time to respond.

Results

Mapping the Sensitivity to Face scrambling

To quantitatively document the sensitivity to face fragmentation we conducted the ‘face-scrambling’ experiment. In this experiment, pictures of faces were broken into smaller and smaller blocks, and the blocks were randomly scrambled. The cleavage lines separating the scrambled blocks were smoothed using a spatial Gaussian filter to reduce the high spatial frequencies associated with the sharp boundaries (see Materials and Methods and Fig. 1). We also included an epoch of car images to provide control objects for localizing the face-selective regions. The subjects’ task was to identify the face’s gender, and the car’s category (see methods).

Object-related voxels were identified by looking for preferential activation for both cars and faces compared to highly scrambled (256 blocks) cars and face images (full faces and full cars > 256). To localize specific face-related voxels we searched (using linear correlation) for voxels that were preferentially activated to full faces compared to cars (Faces > Cars). The statistical tests highlighted a large object-selective region located at the lateral ventral aspect of the occipito-temporal junction, which corresponded to the previously defined LOC. In previous studies (Grill-Spector et al., 1998; Halgren et al., 1999) this region was divided into two putative subdivisions: a posterior-dorsal one, termed lateral occipital (LO), and an anterior-ventral region associated with the posterior fusiform gyrus (pFs). Figure 2 shows the location of face- and object- (faces and cars) selective voxels relative to these subdivisions, and also relative to retinotopic regions defined by visual meridian mapping done on the same subject. As can be seen in the unfolded cortical format, the face- and object-related voxels were essentially non-retinotopic, and were located anterior to retinotopic regions. As we reported previously, LO was located consistently close to a lower visual meridian representation (Grill-Spector et al., 1998). The pFs face activation was situated anterior and lateral to the V4/V8 complex (Hadjikhani et al., 1998) and appears to correspond to the region termed the fusiform face area by Kanwisher et al. (Kanwisher et al., 1997). Note that preferential activation to faces can be found both in the LO and pFs subdivisions.

An Anterior–Posterior Gradient of Sensitivity to Image Scrambling

To map the sensitivity of object areas to face scrambling we used a modified general linear model (GLM) analysis tool (Goebel et al., 1998a). Figure 3a shows such map on an inflated right hemisphere of one subject. In this map, voxels are color coded according to the relative contribution of two predictors: full-face activation versus faces broken into four scrambled blocks. Green and red colors show selective activation to full faces and four-block faces respectively, with blue and yellow indicating a more balanced activation. A tendency can be seen for green voxels to be located more anteriorly, both on the lateral surface in LO and on the ventral aspect of the temporal lobe, along the fusiform gyrus. Figure 3b demonstrates a similar GLM map in the same subject but from a top view of a horizontal section taken along the ventral floor of the occipito-temporal lobe. Examples of activation profiles taken from regions showing different levels of scrambling sensitivity are shown on the right. A gradual increase in activation to the highly scrambled images can be observed, as one moves posteriorly.

The tendency shown in Figure 3a,b for occipito-temporal regions was part of a more global trend that extends from early retinotopic-areas. This global trend is shown in Figure 4, which presents examples from three subjects. Here the GLM map contrasted a 256-block predictor (red–yellow) with a full-face predictor (green–blue). In one subject these maps are shown relative to the borders of retinotopic areas. It is quite clear that the sensitivity to scrambling – i.e. the reduction in fMRI activation as images are broken, was not distributed randomly across the visual cortex but showed a consistent anterior to posterior organization. Thus, regions overlapping retinotopic areas, e.g. V4/V8, showed relatively stronger activation by the scrambled images (as reflected in yellow-colored voxels) compared to more anterior, non-retinotopic regions in the LOC.

Within the LOC, the regions most highly sensitive to scrambling tended to be located more anteriorly and laterally. Figure 5 illustrates this point by contrasting the GLM predictors to full and two-face blocks with 16 and 64 blocks. Note that within the LOC, regions showing the highest sensitivity to scrambling (green) appeared in two main clusters, corresponding to two previously defined subdivisions: a ventral aspect (pFs) extending into the temporal lobe along the posterior-fusiform gyrus and the occipito-temporal sulcus, and a more

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The main centers of activation in Talairach coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Right hemisphere</td>
</tr>
<tr>
<td>Green</td>
<td>Blue</td>
</tr>
<tr>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>LO</td>
<td>42 ± 4.11</td>
</tr>
<tr>
<td>pFs</td>
<td>−39 ± 4.68</td>
</tr>
<tr>
<td>Left hemisphere</td>
<td></td>
</tr>
<tr>
<td>LO</td>
<td>−45.7 ± 3.78</td>
</tr>
<tr>
<td>pFs</td>
<td>−38.1 ± 4.26</td>
</tr>
</tbody>
</table>

Talairach coordinates (Talairach and Tommoux, 1998) of seven subjects who participated in the face-scrambling experiment. The coordinates were derived for the green and blue patches in LOC and pFs. Values represent the mean ± SD in mm.
Figure 6

Figure 9
Figure 7. Activation profiles of face-selective and retinotopic regions. Histograms show the average activation (n = 7) derived from regions with the highest level of scrambling sensitivity (a,b), and retinotopic regions V1–V2 (c,d). The x-axis denotes type of stimuli and the y-axis shows the percentage signal change calculated relative to blank epochs. Asterisk indicates significant difference from the full-face epoch. The icons under the histograms illustrate the stimuli presented to the subjects during each epoch. Average recognition performance (n = 7) is presented as histograms for each stimulus type (e). Error bars indicate ± averaged SEM.

Figure 6. Relationship of face-selective regions to regions sensitive to face scrambling. Superposition of activation maps on unfolded cortical surfaces of right hemisphere from two subjects. Retinotopic borders are indicated by white dotted lines on subject's SN map. The maps show the location of regions most sensitive to face scrambling (pink) relative to regions showing face versus car selectivity (red). Orange corresponds to overlap regions. Note the large overlap between these two maps.

Figure 9. Sensitivity to car image scrambling in the visual cortex. Unfolded hemispheres from two subjects showing a GLM map contrasting predictors of full and half cars (green) versus cars scrambled into 64 and 256 parts (orange). Retinotopic border conventions as in Figure 4. Note a similar, albeit more disorderly, anterior–posterior trend of increased scrambling sensitivity as in the face-scrambling experiment.
to intact faces and faces broken into halves (Fig. 7).

Voxels showing the highest sensitivity to scrambling (green) could be observed in both subdivisions, and seemed to belong to a single, albeit variable, arc-like complex. Table 1 summarizes these results in Talairach coordinates (Talairach and Tournoux, 1988).

To relate the location of regions sensitive to face scrambling with face-selective regions we compared the location of voxels preferentially activated by full faces compared to four parts of a face (Full > 4) with voxels preferentially activated by faces compared to car images (Full > Cars). Figure 6 shows examples of the right hemispheres of two subjects. As can be seen there is a large overlap between the two sets, indicating that regions which show high sensitivity to scrambling also show high object selectivity.

Change in Scrambling Sensitivity as Reflected in the Time Course of Activation

Figure 7 contrasts the time course of activation obtained from regions showing the highest level of scrambling sensitivity (colored green in Fig. 4), with early retinotopic areas. At the highest level of sensitivity, anterior voxels in the LOC showed some reduction even when images were broken into four quadrants. Interestingly, the voxels most sensitive to image scrambling did not show a reduced activation when images were broken into halves, either horizontal or vertical (Fig. 7a,b). In contrast, posterior retinotopic areas V1–V2, showed an opposite trend: enhanced activation to highly scrambled faces compared to intact faces and faces broken into halves (Fig. 7c,d,e).

Thus, the functional profile of ventral stream areas changes drastically in terms of sensitivity to image scrambling – from sensitivity to local image features in early retinotopic areas to more global image elements in non-retinotopic object related areas.

We compared these profiles of activation to subjects’ performance which was measured outside the magnet but using the same visual stimuli. Figure 7e shows subjects’ percentage correct responses on a gender identification task of face scrambling. As was expected, performance for this task declined with increasing scrambling level, although for the gender identification task this decline was rather gradual.

The ‘Car-scrambling’ Experiment

Are the effects of scrambling specific to faces or do they generalize to other objects as well? To address this issue we repeated the same scrambling experiment, but using another object category – car images. Figure 8 shows examples of the pictures used in the different car-scrambling epochs and their sequence. As in the face-scrambling experiment, images of cars were broken successively into a larger number of blocks (see Materials and Methods). We used also full-face images to allow localization of object related voxels in an identical fashion as in the face scrambling experiment.

Figure 9 shows the GLM maps, similar to those constructed from the face-scrambling experiment – but this time for the car-scrambling experiment. Here, green colors indicate dominance for full and half car images while red colors indicate dominance of 64 and 256 block images, with blue and yellow indicating intermediate levels. Similar to the face-scrambling experiment an anterior–posterior axis is evident here as well, along which there is a trend for increased sensitivity to car-image scrambling. This axis extends from retinotopic visual areas showing the least sensitivity to car scrambling (yellow colors) to non-retinotopic anterior regions showing highest sensitivity to car scrambling (green colors).

Figure 10 shows a finer analysis of this trend, contrasting GLM predictors of full car images and two scrambled halves (green) with 16 and 64 blocks (orange). The LOC showed heterogeneous mapping for the different levels of scrambling sensitivity, with some, albeit less orderly, tendency for increased sensitivity in anterior regions.

We wanted to explore the anatomical relationship between the regions most sensitive to face scrambling and those most sensitive to car scrambling. To reveal this relationship we superimposed these regions from the car-scrambling experiment onto the most sensitive regions from the face-scrambling experiment. Figure 11 illustrates this relationship. It can be seen that regions highlighted in the face-scrambling experiment and the car-scrambling experiment (red and yellow colors respectively) are both located in the LOC, albeit at some displacement. These results are compatible with the notion that similar, although not necessarily overlapping, brain regions are involved in the more holistic representations of faces and cars.
However, whether this conclusion can be generalized to other object categories remains to be seen.

Discussion

The Involvement of Low-level Features

It may be argued that the change in activation with object scrambling is not due to the loss of complex object features but rather due to changes in lower-level aspects. Specifically it may be that the sharp edges that are introduced with each level of scrambling inhibit the response in high-order object areas and lead to the observed reduction in activity. Several lines of evidence argue against this interpretation. First, in the present study, the edges within each image were smoothed using a Gaussian function (see Materials and Methods). Second, in a previous study using scrambled images we directly checked the level of line-inhibition by superimposing a grid on the full images. This procedure did not affect the activation, indicating that the lines and edges by themselves have no inhibitory role in the LOC (Grill-Spector et al., 1998, Kourtzi and Kanwisher, 2000). Smoothing the images with a low-pass filter also did not affect the results. Finally, we have demonstrated that substantial changes in the spatial frequency Fourier power spectrum of images do not affect LOC activation. In particular, high- and low-pass filtering of object images, as well as random blocks, did not affect the activation, and could not account for the preferential activation to objects relative to scrambled images (Malach et al., 1995).

Evidence for a Hierarchical Organization in the Human Ventral Stream

The organizational principle that underlies the relationship of the LOC to neighboring areas is still unclear. Here we provide evidence of a systematic axis that runs from retinotopic areas, into the anterior aspect of the occipito-temporal LOC. This axis is defined by the sensitivity to image scrambling and runs in two anatomical branches – a more dorsal one extends in a purely anterior lateral direction into the vicinity of the lateral-occipital sulcus and posterior aspect of the inferior-temporal sulcus. A more ventral branch runs into the ventral aspect of the temporal lobe along the fusiform gyrus and the occipito-temporal sulcus.

Several reports have identified the fusiform gyrus as a site of highly selective object processing and, in particular, face processing (Clark et al., 1996; Kanwisher et al., 1997; McCarthy et al., 1997). However, face-selective activation has also been found in the dorsal branch as was defined above (Clark et al., 1996; Kanwisher et al., 1997; Grill-Spector et al., 1998). Inspecting the face-scrambling maps in the present study, in the unfolded format, it appears that the two subdivisions may actually be a part of a single superstructure, that in some extent extends further into the fusiform gyrus, while in others extends dorsally into the lateral-occipital sulcus. The functional relationship between the dorsal and ventral parts of this superstructure is not clear at present.

This notion of an anterior–posterior axis is compatible with the concept of a feature-based hierarchical organization within ventral-stream, object-related, human visual areas, leading from part- or fragment-based coding to a more holistic representation. The idea of a hierarchical organization within the visual system has been a central theme in primate visual cortex study, but surprisingly few studies, if any, have demonstrated how it may be reflected across the entire ensemble of human visual areas. The gradual object-scrambling approach combined with the ability to map visual areas in an unfolded surface format provided an opportunity to map in detail such hierarchical transformation through the sequential change in scrambling sensitivity along the ventral stream. Thus, lower-tier regions that contain neurons sensitive to local, isolated object features did not show a reduction with image scrambling, since, presumably, when images are broken into large blocks most of the local feature structure from the image is preserved. In contrast, higher-order representations, which presumably contain neurons sensitive to complex object fragments, showed a more rapid decline even with a minor level of object scrambling.

In our previous study (Grill-Spector et al., 1998) we have demonstrated a similar trend of increased sensitivity to object scrambling in high-order visual areas. The present work confirms these early findings, however, the use of the unfolded cortical format allowed us to reveal a hierarchical trend within the non-retinotopic LOC itself. Thus, more posterior LOC regions show less sensitivity to image scrambling compared to the most anterior regions (e.g. Fig. 5).

The notion of an anterior–posterior hierarchical gradient is fully compatible with results from monkey object-related areas. However, in monkeys, the sequence of object-areas is displaced anteriorly along the temporal lobe (Kobatake and Tanaka, 1994). Although searching for direct homologies between high-order visual areas in humans and monkeys is tentative at this stage, it is interesting to note that in monkeys face-selective neurons can be found in an extended complex of high-order areas including the different parts of the superior temporal sulcus, and the lateral and ventral banks of the inferior temporal cortex (Desimone et al., 1991). This is somewhat analogous to the extent of human face-related voxels reaching from posterior-ITS dorsally to posterior fusiform ventrally.

Recently, the image-scrambling approach has been applied directly to single neurons of the behaving monkey (Vogels, 1999). The results are qualitatively similar to the profile of activation found in the most sensitive human sites; however, the methodological gap between these studies is too large to permit a quantitative and direct comparison.

It is interesting that cutting objects into halves did not result in a substantial reduction in activation even in regions, which are presumably at the highest level of the object hierarchy. The reason for this phenomenon is still unclear: an appealing explanation is that it may reflect a powerful object completion effect, whereby showing a large object part is sufficient to activate the entire network representing this object, but further experiments are needed to clarify this result.

Semantic-based Organization in Human Object-areas

An alternative to the feature-based hierarchical scheme proposed here is the suggestion for separate and parallel representations in high-order object-areas; one for general object recognition located in dorsal-posterior regions, and another specialized for faces in anterior-ventral regions (Haxby et al., 1994; Kanwisher et al., 1997; Halgren et al., 1999; Kourtzi and Kanwisher, 2000). The basis for this subdivision is the observation that anterior regions show preferential activation to faces compared to common objects, while posterior regions are more equally activated by both categories. Although not immediately obvious, the serial, hierarchical scheme proposed here is also compatible with these experimental findings. It is expected from the hierarchical scheme that posterior regions in LOC, which contain a representation of intermediate-level features should be activated more equally by a variety of objects,
since these relatively low-level features are shared by many different objects. In contrast, anterior LOC regions, which contain more complex/holistic features, may manifest a higher selectivity to a specific class of objects such as faces, since the statistical distribution of these features is more unique to each object class.

Are there separate hierarchies for different object classes? Our results comparing cars and faces point to a single overall hierarchical organization subserving different categories. Thus, cortical regions showing the highest selectivity to face scrambling were located adjacent to regions showing highest selectivity to car scrambling. However, we did not find a precise overlap in these maps, which may account for neuropsychological findings of double-dissociation between face and object recognition (Moscovitch et al., 1997). In addition, we can not rule out that other object categories, e.g., architectural landmarks (Aguirre et al., 1998; Epstein et al., 1999), may be represented in a separate hierarchical stream.

Our results confirm our previous finding (Grill-Spector et al., 1998) that activation in high-order object-areas is quite insensitive to substantial changes in object configurations, provided the main feature structure is preserved. Thus, a substantial part of the LOC did not show a significant decline in activation even as the images were broken into four scrambled blocks. Even in the most face-selective voxels, there was only a minor reduction in activation when face pictures were cut into halves. Thus, one has to conclude that the integrity of the entire object and face template is not a critical factor in activation of high-order object-areas.

It should be noted that for the task employed in the present study (gender identification) the overall face ‘template’ was unnecessary for successful performance, as shown in Figure 7e. However, the recognition performance is undoubtedly extremely sensitive to the task requirement, so under different task demands (e.g. specific individual identification) a much higher sensitivity to scrambling may be revealed.

A hypothetical representation consisting of extremely local and independent feature detectors is also incompatible with the results, since image quarters preserve a substantial structure of such local features, and yet in many voxels, and particularly at the most anterior-ventral subregions, this level of image scrambling produced a decline in activation. Finally, due to the limited resolution of fMRI we certainly cannot exclude the possibility that a small subset of neurons intermixed within the imaged voxels require the entire image template for activation.

Notes
We thank E. Okon and I. Levy for software development. We thank G. Avidan-Carmel for fruitful discussions and comments. This study was funded by ISF 644/99 and by JSMF grant 99–28 CN-QUA.05 grants.

Address to correspondence to R. Malach, Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel. Email: rafii.malach@weizmann.ac.il.

References


