◆ Human Brain Mapping 6:316–328(1998) ◆

A Sequence of Object-Processing Stages Revealed by fMRI in the Human Occipital Lobe

Kalanit Grill-Spector,[1] Tammar Kushnir,[2] Talma Hendler,[2] Shimon Edelman,[3] Yacov Itzchak,[2] and Rafael Malach[1]

[1]*Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel*
[2]*Diagnostic Imaging, Chaim Sheba Medical Center, Tel Hashomer 52621, Israel*
[3]*School of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton BN1 9RH, UK*

Abstract: Functional magnetic resonance imaging was used in combined functional selectivity and retinotopic mapping tests to reveal object-related visual areas in the human occpital lobe. Subjects were tested with right, left, up, or down hemivisual field stimuli which were composed of images of natural objects (faces, animals, man-made objects) or highly scrambled (1,024 elements) versions of the same images. In a similar fashion, the horizontal and vertical meridians were mapped to define the borders of these areas. Concurrently, the same cortical sites were tested for their sensitivity to image-scrambling by varying the number of scrambled picture fragments (from 16–1,024) while controlling for the Fourier power spectrum of the pictures and their order of presentation. Our results reveal a stagewise decrease in retinotopy and an increase in sensitivity to image-scrambling. Three main distinct foci were found in the human visual object recognition pathway (Ungerleider and Haxby [1994]: Curr Opin Neurobiol 4:157–165): 1) Retinotopic primary areas V1–3 did not exhibit significant reduction in activation to scrambled images. 2) Areas V4v (Sereno et al., [1995]: Science 268:889–893) and V3A (DeYoe et al., [1996]: Proc Natl Acad Sci USA 93:2382–2386; Tootell et al., [1997]: J Neurosci 71:7060–7078) manifested both retinotopy and decreased activation to highly scrambled images. 3) The essentially nonretinotopic lateral occipital complex (LO) (Malach et al., [1995]: Proc Natl Acad Sci USA 92:8135–8139; Tootell et al., [1996]: Trends Neurosci 19:481–489) exhibited the highest sensitivity to image scrambling, and appears to be homologous to macaque the infero-temporal (IT) cortex (Tanaka [1996]: Curr Opin Neurobiol 523–529). Breaking the images into 64, 256, or 1,024 randomly scrambled blocks reduced activation in LO voxels. However, many LO voxels remained significantly activated by mildly scrambled images (16 blocks). These results suggest the existence of object-fragment representation in LO. *Hum. Brain Mapping 6:316–328, 1998.*

Key words: object recognition; visual cortex; visual areas; brain mapping; form perception

INTRODUCTION

Recent functional imaging studies of the human object-recognition pathway have focused on its initial stage (primary visual cortex) and on its final stage (category-selective regions), while the intermediate links between these stages are less understood. On the one hand, extensive research was aimed at a detailed retinotopic mapping of human striate and extrastriate

Contract grant sponsor: ISF; Contract grant number: 1713/97-1.

This work was performed at the Weizmann Institute of Science, Rehovot, Israel, and at the Chaim Sheba Medical Center, Tel Hashomer, Israel.

*Correspondence to: R. Malach, Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: bnmalach@wis.weizmann.ac.il

Received for publication 18 September 1998; accepted 7 April 1998

areas using simple visual stimuli [Engel et al., 1994; Fox et al., 1987; DeYoe et al., 1996; Boynton et al., 1996; Sereno et al., 1995; Schneider et al., 1993], without addressing their potential involvement in object representation. On the other hand, several groups concentrated on detecting regions activated by viewing, categorizing, or naming whole objects [Martin et al., 1996; Kanwisher et al., 1996; Malach et al., 1995; Ishai et al., 1997], and in particular, faces [Sergent et al., 1992; Kanwisher et al., 1997a; Allison et al., 1994; Puce et al., 1995; Haxby et al., 1994, 1996], but did not investigate their retinotopic properties.

At present, the role of the various visual areas in the process of object recognition is still unclear [e.g., (Kanwisher et al., 1997b]. It is yet unknown what type of transformations or calculations are performed in each area and in particular whether representations in high areas pertain to complete objects [Edelman, 1995] or their component parts [Biederman, 1987]. Our aim in the present study was to functionally distinguish areas involved in the process of visual object recognition by concurrently measuring retinotopy and sensitivity to various levels of controlled object fragmentation and scrambling.

METHODS

Twelve healthy volunteers (ages 22–46, 9 male, 3 female) who gave written informed consent participated in the study. The experimental protocol was approved by the Sheba Medical Center Ethics Committee. The subjects were scanned with a whole body MRI ommitsystem (2T Prestige, Elscint, Ltd.) using multislice echo planar imaging (EPI) and a standard birdcage head coil. Susceptibility sensitive pulse sequence (a T2*-weighted multislice EPI gradient echo sequence: TR/TE/flip angle = 2,000/45/90) was used. Six slices, 6 mm thick, were oriented approximately perpendicular to the calcarine fissure; in-plane resolution was 3×2.7 mm.

Experimental protocol

Visual stimulation was generated on a SGI workstation and was projected via an LCD projector onto a screen located at the back of the scanner. Subjects viewed the stimuli through a mirror attached to the head coil, providing maximal 40° horizontal × 30° vertical visual angle. The average luminance of the images was 80 cd/m^2, and the luminance of blank epochs was 20 cd/m^2. A typical scanning session consisted of 6–8 scans, each scan comprised of a different experiment; in each scan, 960 images were collected over 320 sec, from six contiguous slices of the brain. Three scans in which head motion exceeded several voxels were discarded. The data were subjected to principal component analysis (PCA) [Reyment and Joreskog, 1993], a preprocessing stage that removed spatiotemporally correlated noise artifacts [for details see Grill-Spector et al., 1998]. Typically, the first two components were removed: the first component depicted a linear temporal drift (possibly the drift of the magnetic field), and the second component contained high temporal frequency-correlated noise (spatially distributed in a homogeneous manner within the slice). Subsequently, the experiments were analyzed using either a Kolmogorov-Smirnov (K-S) statistical test [Baker et al., 1993] (visual field experiments), or regression analysis to an ideal paradigm [Friston et al., 1995] (scrambling experiments). Time courses were derived from the statistically significant voxels.

Visual stimuli

Twelve subjects participated, within a single scanning session, in the visual field left-right, in the visual field up-down, and in scrambling experiments. The details of these experiments are described below. In the same scanning session, we mapped in 7 subjects the representation of the visual meridians in order to determine the borders of retinotopic visual areas. Five subjects participated in scrambling experiment ''c'' during a separate scanning session.

Visual field experiments

Visual field left-right

The experiment was divided into 12 epochs. The first and last epochs were blank epochs, except for a fixation cross. In each visual epoch, 14 different images were presented at a rate of 0.5 Hz (see Fig. 1a). Epochs were divided into right vs. left hemifield stimulation (epochs 6–9 vs. epochs 4, 5, 10, and 11, while epochs 8–11 excluded foveal stimulation). In the same experiment, gray-level pictures of objects (faces, animals, manmade) and images obtained by scrambling the original pictures into 1,024 blocks were alternated (even vs. odd epochs, respectively). At the center of the screen, a fixation cross was presented. Subjects were instructed to fixate upon the cross and covertly name its color, which was randomly changed.

Visual field up-down

This experiment was similar to the visual-field left-right experiment (see Fig. 3a). The first and last

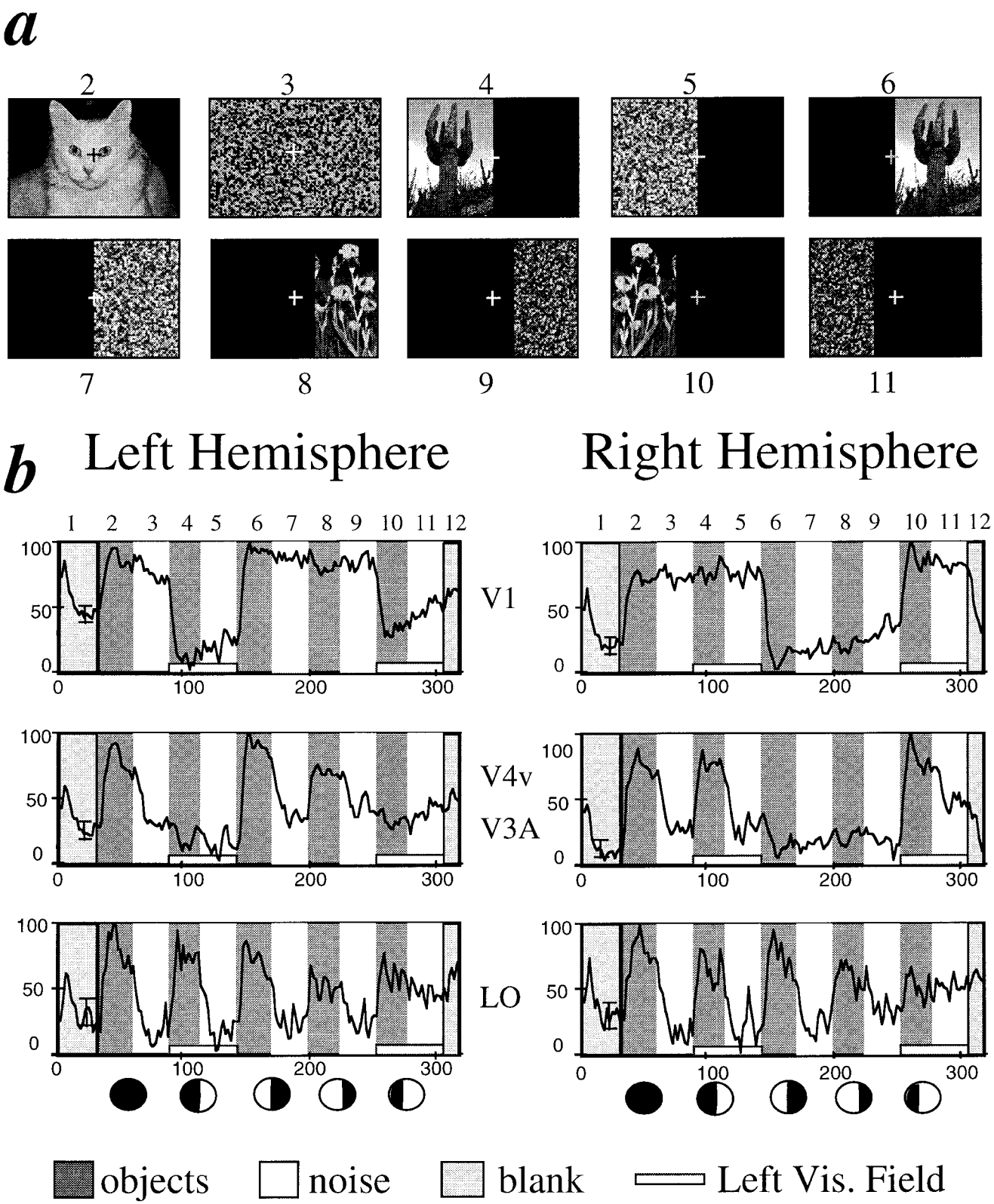
a
2
3
4
5
6
7
8
9
10
11
b
Left Hemisphere
Right Hemisphere
1 2 3 4 5 6 7 8 9 10 11 12
100
50
0
0 100 200 300
V1
V4v
V3A
LO
objects
noise
blank
Left Vis. Field

Figure 1.

epochs were blank, except for a fixation cross. In each visual epoch, 14 different images were presented at a rate of 0.5 Hz. Epochs were divided into upper vs. lower hemifield stimulation (epochs 6, 7, 10, and 11 vs. epochs 4, 5, 8, and 9, while epochs 8–11 excluded foveal stimulation). In the same experiment, gray-level pictures of man-made objects (vehicles) and images obtained by scrambling the original pictures into 1,024 blocks were alternated (even vs. odd epochs, respectively). A fixation cross whose color was randomly changed was presented in the center of the screen. Subjects were instructed to fixate upon the cross and covertly name its color.

Image-scrambling experiments

In the following experiments, gray-level images of natural objects (faces and animals) were randomly scrambled into an increasing number of squares (16, 64, 256, and 1,024, for visual epochs 2–5, respectively); see Figure 5a. Epochs of visual stimulation (40 sec long, 1 image every 2 sec) were alternated with blank epochs (20 sec long). Subjects were instructed to covertly name the visual stimuli, including the scrambled images. Three variations of this experiment were designed in order to control for various parameters, which might influence the results. In experiment a, we low-pass filtered all images subsequent to the scrambling process with a finite impulse response spatial filter (cutoff frequency = 15 cycles per image, size = 21 × 21 pixels) in order to cancel out changes in

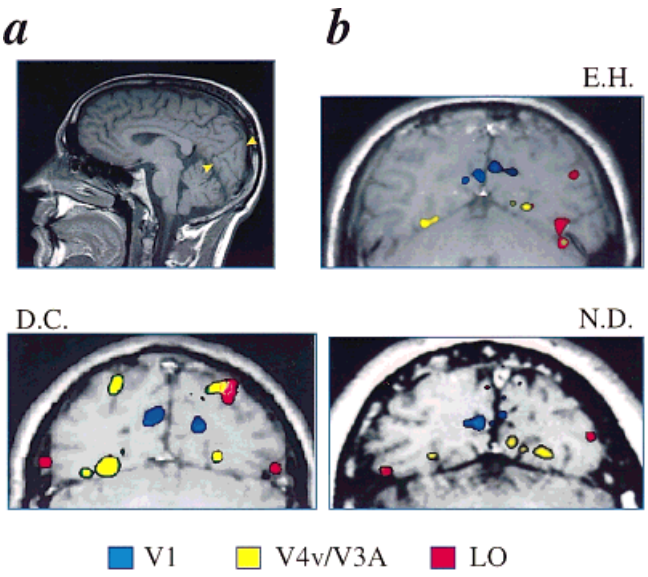

Figure 2.

Visual field left-right (L/R) experiment. Activation maps. a: A midsaggital view of the brain. Arrows indicate plane of sections (perpendicular to the calcarine sulcus) shown in b. Note that "dorsal" in these cross sections corresponds to "dorsal-posterior" in coronal cross sections, and that "ventral" corresponds to "ventral-anterior" in coronal cross sections. b: Superposition of the K-S activation maps of the three foci of activation in the visual field left-right experiment overlaid upon the T1-weighted high-resolution scans of 3 different subjects. The lightness of each color corresponds to the statistical significance (darkest $P < 1e\text{-}3$ and lightest $P < 1e\text{-}8$). Spatial Guassian smoothing (kernel = 3 × 3, sigma = 1) was used to reduce noncorrelated noise artifacts prior to the statistical tests. Note that the three foci are consistently arranged mediolaterally.

Figure 1.

Visual field left-right (L/R) experiment. a: An example of the pictures that were used in the experiment. Numbers denote consecutive epoch indexing. Each epoch consisted of 14 different images of the type depicted. The first and last epochs were blank epochs, except for a fixation cross. In the experiment, epochs containing images of natural objects alternated with highly scrambled (1,024 blocks) versions of the same images, and right visual field stimulation (epochs 6–9) alternated with left visual field stimulation (epochs 4, 5, 10, and 11). In the last four visual epochs (8–11), foveal stimulation was excluded. Subjects were instructed to covertly name the color of the central fixation cross (see Methods for details), which varied in color during the experiment. b: Activation time courses. Average time course of 9 subjects derived from each of the three distinct functional foci (shown in Fig. 2b). The x-axis denotes time in seconds and the y-axis shows normalized fMRI signal amplitude. Error bars indicate ± 1 averaged standard error of the mean (SEM). The icons beneath the bottom time courses illustrate the stimulated visual field during each epoch. Shaded regions indicate whole-object epochs, while the unshaded regions correspond to highly-scrambled-object epochs. Numbers at top correspond to the epoch number as denoted in a. Note the distinct differences in object selectivity and retinotopy manifested by the three foci (V1, V4v/V3A, and LO, respectively). Top: V1 voxels (blue in Fig. 2) showed enhanced activation to the contralateral visual field (fMRI signal increase defined as: (contralateral-ipsilateral)/ipsilateral: 4.05 ± 0.7% SEM, n = 9), but only slight enhancement of the activation by images of objects compared to scrambled images (0.18 ± 0.18% SEM, n = 9). Note that this enhancement is one order of magnitude smaller than the activation by the contralateral visual field). Middle: V4v and V3A voxels (yellow in Fig. 2) displayed both enhanced activation by contralateral visual field stimulation (fMRI absolute signal increase: 3.25 ± 0.6% SEM, n = 9) and amplified activation by objects compared to scrambled objects (fMRI signal increase: 1.99 ± 0.7% SEM, n = 9). Bottom: LO voxels (red in Fig. 2) exhibited increased activation by images of objects compared to highly scrambled versions of these objects (fMRI signal increase: 1.53 ± 0.54% SEM, n = 9) and activation by both contralateral and ipsilateral visual fields. Note that the fMRI signal induced by contralateral-field stimulation was only slightly increased (0.08 ± 0.1% SEM, n = 9) compared to ipsilateral-field stimulation.

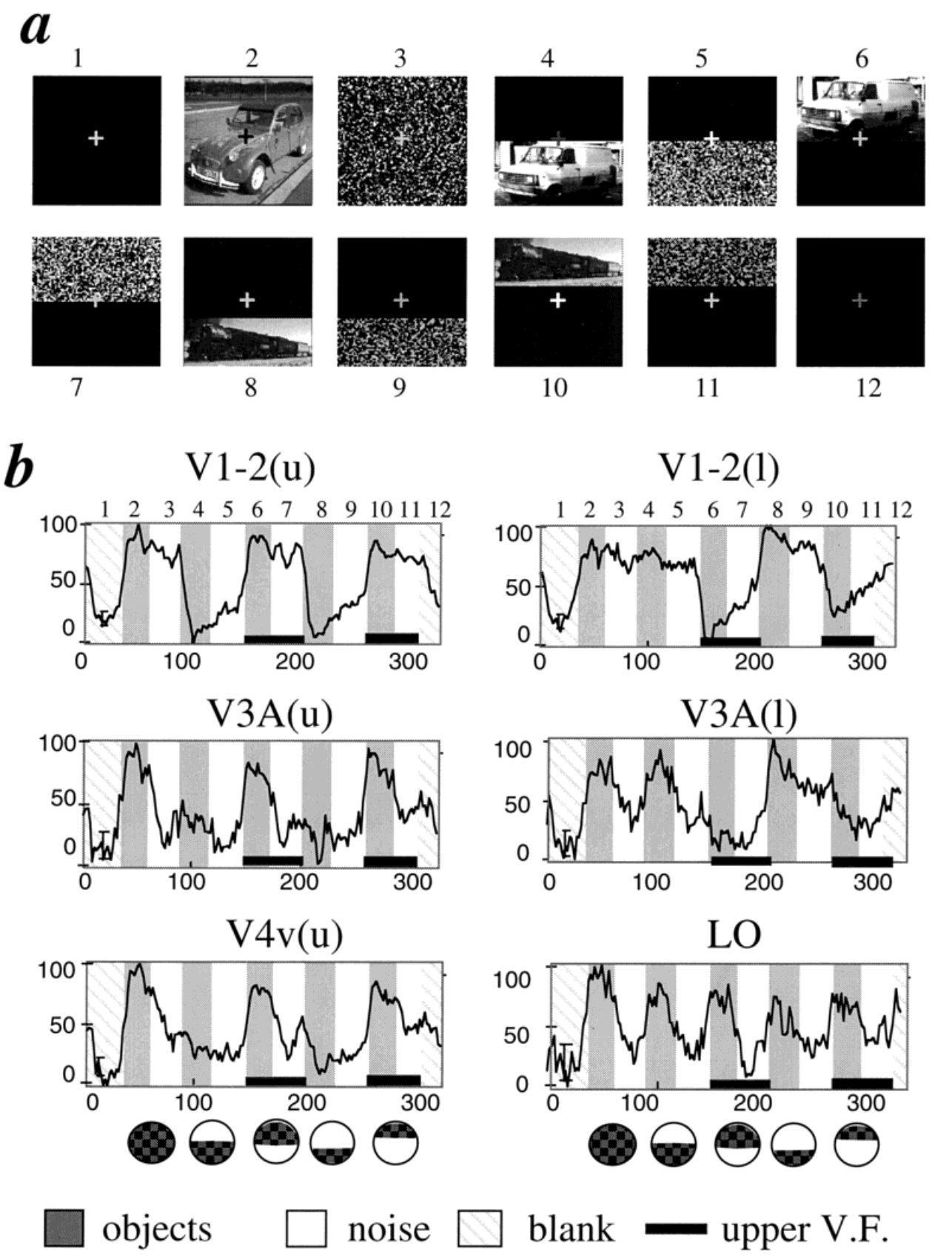
a
1
2
3
4
5
6
7
8
9
10
11
12
b
V1-2(u)
V1-2(l)
V3A(u)
V3A(l)
V4v(u)
LO
1 2 3 4 5 6 7 8 9 10 11 12
100
50
0
0 100 200 300
objects
noise
blank
upper V.F.

Figure 3.

the spatial frequency spectrum caused by the scrambling process (in particular, high-spatial frequencies). In experiment b, we also controlled the spatial frequency content of the various images, this time by introducing high-spatial frequencies to images of whole objects by superimposing a 4 × 4 grid upon these images. Another parameter we controlled for was the order of the visual epochs. In experiment c, we presented the same images that were used in b, with the order of epochs permuted. Thus, the first visual epoch contained images of the most scrambled pictures (1,024 blocks), the second epoch consisted of images fragmented into 16 blocks, the third consisted of 256 blocks, the fourth was of 64 blocks, and the last consisted of whole objects on which a 4 × 4 grid was superimposed.

Mapping borders of visual areas

In order to delineate the borders of retinotopic areas, the representations of the vertical and horizontal visual field meridians were mapped [Sereno et al., 1995; Engel et al., 1994; DeYoe et al., 1996]. Visual stimuli, presented at a rate of 2 Hz, consisted of horizontal or vertical wedge-shapes to compensate for the expanded foveal representation. A small fixation cross (length, 0.56°; width, 0.19°), whose color changed randomly, appeared in the center of the image. Subjects were instructed to fixate upon the cross and covertly name its color. Half of the wedges were made of gray-level natural images cropped to fit the wedge-shape, and the other half consisted of flickering black and white random dots. The flickering-dot stimuli were effective in mapping the borders of V1 and V2, while the natural images were optimal in revealing higher-order areas, i.e., the vertical meridians at the border between areas VP and V4v (ventrally) and V3 and V3A (dorsally), and the horizontal meridians defining the borders of V4v and V3A [Sereno et al., 1995; DeYoe et al., 1996].

RESULTS

In order to directly compare retinotopy and object sensitivity across the occipital lobe, a special visual mapping paradigm was used (12 subjects, as described in Methods) in which two parameters were concurrently measured: the spread of activation from the ipsilateral visual field, and the preferential activation to images of natural objects compared to highly scrambled versions of the same images (see Fig. 1a). This compound visual field test (visual field L/R) revealed three distinct foci of activation in the occipital lobe, arranged mediolaterally in both hemispheres (see Fig. 2). The medial focus (blue in Fig. 2), located over

Figure 3.

Visual field up-down (U/D) experiment. a: Visual stimuli. An example of the sequence of images presented in the experiment, where numbers denote consecutive epoch indexing. Each epoch consisted of 14 different images of the type depicted. The first and last epochs were blank epochs, except for a fixation cross. In the experiment, epochs containing images of vehicles were alternated with highly scrambled (1,024 blocks) versions of the same images, and lower visual field stimulation (epochs 4, 5, 8, and 9) alternated with upper visual field stimulation (epochs 6, 7, 10, and 11). In the last four visual epochs (8–11), foveal stimulation was excluded. Subjects were instructed to covertly name the color of a central fixation cross (see Methods for details), which varied in color during the experiment. b: Activation time courses. Average time course of 9 subjects derived from each of the distinct functional foci (depicted in Fig. 4). The x-axis denotes time in seconds and the y-axis shows normalized fMRI signal amplitude. Error bars indicate ± 1 averaged standard error of the mean (SEM). The icons beneath the time course illustrate the stimulated visual field during each epoch. Shaded regions indicate whole-object epochs, while unshaded regions correspond to highly-scrambled-object epochs. Numbers at top correspond to the epoch number as denoted in a. Letters in parentheses indicate preferred visual field. l, lower; u, upper. Top: V1–2(l) dorsal voxels (Fig. 4, purple) were activated by lower visual-field stimulation (fMRI signal increase: 2.04 ± 0.43% SEM, n = 9), and V1–2(u) ventral voxels (Fig. 4, blue) were activated by the upper visual-field stimulation (fMRI signal increase: 1.91 ± 0.23% SEM, n = 9), but were only slightly selective to images of objects compared to scrambled images (fMRI signal increase by objects compared to noise: 0.35 ± 0.11% SEM, n = 9). Middle: V3A exhibited a distinct functional profile. Unlike areas V1–2, V3A responded preferentially to images of objects compared to noise images (fMRI signal increase by objects compared to noise: 1.27 ± 0.27% SEM, n = 9). The posterior part V3A(l) (Fig. 4, light blue) was activated by lower visual-field stimulation (fMRI signal increase (lower-field objects—upper-field objects)/(upper-field objects): 0.78 ± 0.14% SEM, n = 9) and the anterior part V3A(u) (Fig. 4, green) was activated by upper visual-field stimulation (fMRI signal increase (upper-field objects—lower-field objects)/(lower-field objects): 1.10 ± 0.17% SEM, n = 9). Bottom, left: V4v voxels (Fig. 4, yellow) exhibited both preference to upper visual field (fMRI signal increase: 1.27 ± 0.31% SEM, n = 9) and images of objects (fMRI signal increase: 1.26 ± 0.20% SEM, n = 9). Bottom, right: LO voxels (Fig. 4, red) displayed preferential activation to images of objects compared to scrambled images of objects (fMRI signal increase: 1.07 ± 0.35% SEM, n = 9), but responded both to upper and lower visual field stimulation.

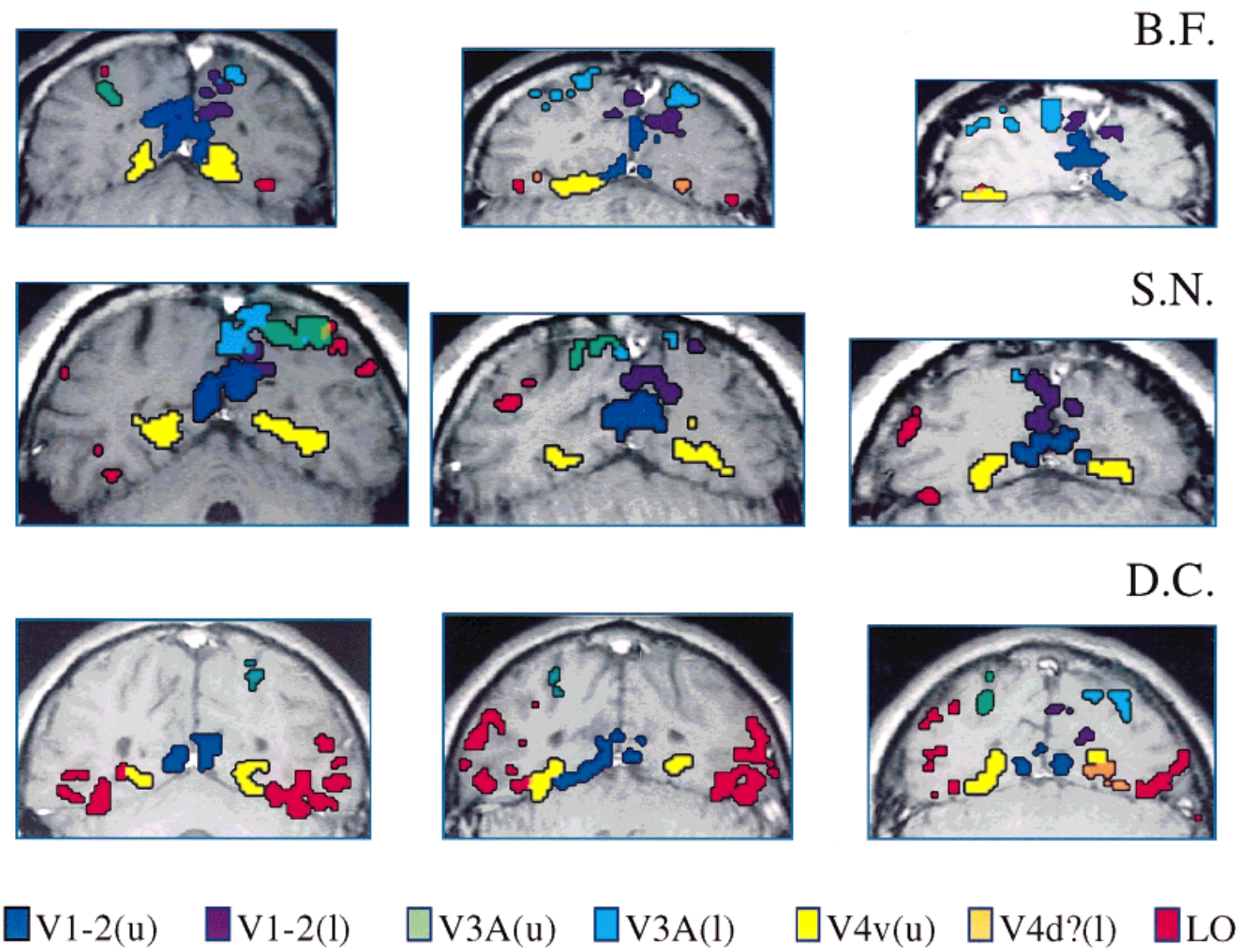

Figure 4.

Visual field up-down experiment. Activation maps. Superposition of the K-S activation maps of the different foci of activation obtained in the visual field up-down experiment have been overlaid upon the T1-weighted high-resolution scans of 3 different subjects. Each row depicts three consecutive slices of a single subject, with the middle slice located in a plane similar to the one illustrated in Figure 2a. Each functional profile ($P < 1e\text{-}4$; corresponding time courses are given in Fig. 3b) was assigned a different color. The only exceptions were V3A(u) and V4v(u), which exhibited a similar functional profile in this experiment (see Fig. 3b), but were colored differently because of their different anatomical locations. Letter in parentheses indicates preferred visual field. u, upper; l, lower. Note that these functional foci are similar to the foci of the visual field left-right experiment, which are depicted in Figure 2b.

the calcarine sulcus and the medial surface of the occipital lobe, was essentially shut off when the visual stimulus was confined to the ipsilateral field, but showed only weak reduction when objects were highly scrambled (see Fig. 1b, top, for corresponding time courses). More laterally, a small dorsal focus and a ventral focus that was located mainly within the fusiform gyrus (yellow in Fig. 2), showed both significant reduction in activation to the scrambled images and preferential activation to the contralateral visual field (see Fig. 1b, middle). Finally, voxels overlapping the lateral aspect of the fusiform gyrus (red in Fig. 2) showed a clear sensitivity to scrambling but an essentially complete spread of activation from the ipsilateral visual field (see Fig. 1b, bottom). It should be noted that in this population of voxels, there was some reduction in activation when object stimuli excluded the fovea (reduction, 0.71 ± 0.15, SEM, $n = 9$; cf. time courses in Fig. 1b, bottom, epochs 8–11 compared to epochs 4–7).

In a similar experiment (12 subjects), the sensitivity to upper vs. lower visual field activation was tested

(see Fig. 3a for stimuli used). Overall, this experiment again differentiated the same three functional foci; however, the foci which showed both retinotopy and sensitivity to object scrambling exhibited a dorsal-ventral heterogeneity (see below). The medial focus, similar to the visual-field L/R experiment, was sensitive to the stimulated visual field, but exhibited only a slight reduction in activation by scrambled images compared to whole images of objects. Dorsal-medial regions (Fig. 4, purple) were essentially shut off by upper visual field stimulation, and ventral-medial regions (Fig. 4, blue) were essentially shut off by lower field stimulation (see Fig. 3b, top, for corresponding time courses). The foci which displayed both retinotopy and sensitivity to image scrambling (Fig. 3b, middle) manifested a different mapping dorsally and ventrally. The dorsal focus contained a separate mapping of the lower and upper visual fields. The lower-field representation was typically located in the posterior slices (Fig. 4, light blue), while the map of the upper visual field (Fig. 4, green) was adjacent to the lower visual field map, but emerged more anteriorly. The ventral focus, located within the fusiform gyrus (Fig. 4, yellow) manifested a preferential activation to the upper visual field (see Fig. 3b, bottom left) in all subjects. Some voxels detected on the lateral aspect of this focus, exhibited both preference to images of whole objects and lower visual field stimulation (Fig. 4, orange). The activation time course of these voxels (not shown) showed increased activation by images containing objects, but a general decrease in activation by peripheral stimulation. Therefore, it is possible that they included the horizontal visual meridian representation. Moreover, the size of this lower field representation was smaller (on average, 18 ± 5 voxels) compared to the upper visual field representation (185 ± 21 voxels).

Finally, voxels overlapping the lateral aspect of the fusiform gyrus (Fig. 4, red) showed preferential activation to whole images compared to highly scrambled versions of these images, but responded similarly to upper and lower visual field stimulation (see Fig. 3b, bottom right).

In order to relate these foci of activation to established human visual areas, we mapped in 7 subjects, during the same experimental sessions, the vertical and horizontal meridians [Sereno et al., 1995; DeYoe et al., 1996], using polar-map sectors of either natural objects or texture stimuli (see Methods for details). Comparing the meridian mapping to the activation foci obtained in the visual field experiments indicated that the medial focus which exhibited retinotopy, but was not significantly affected by the type of visual stimulation, was confined to areas V1–V3. The more lateral ventral focus, which exhibited preference to images of objects compared to highly scrambled versions of these images and preference to both the upper and contralateral visual fields, overlapped mostly with area V4v. The more lateral dorsal focus, which showed preference to images of objects compared to scrambled images and contained a hemifield contralateral map, partially overlapped the second vertical meridian representation and probably corresponds to area V3A.

Finally, the most lateral focus, that was significantly activated by images of objects compared to noise patterns, but was essentially not retinotopic, was located outside the meridian mapping and corresponded anatomically to the lateral occipital (LO) complex [Malach et al., 1995; Tootell et al., 1996]. It should be noted that in 5 subjects some voxels anterior to area V3A exhibited a similar functional profile to LO, but with the peripheral representation significantly reduced.

To explore in detail the sensitivity to image-scrambling in the human object processing stream [Ungerleider and Haxby, 1994], we designed an experiment in which pictures of natural objects were scrambled in a gradual manner (see Fig. 5a). The pictures in all the epochs were spatially low-pass filtered (see Methods for details) in order to avoid confounding spatial frequency-related effects. This experiment again differentiated three main occipital regions, as depicted in Figure 6. V1–3 voxels (Fig. 6, blue) showed no reduction in activity and even slight enhancement with mild picture scrambling. V4v (Fig. 6, yellow) showed substantial reduction in the two highly scrambled epochs (256 and 1,024 blocks). LO voxels (Fig. 6, red) had the highest sensitivity to image-scrambling, showing reduced activation following the intermediate scrambling into 64 blocks (cf. the corresponding time courses of areas V1, V4v, and LO in Fig. 5b).

Analyzing the behavior of LO voxels in detail, we found that in the majority of LO voxels, breaking the pictures into 16 scrambled squares did not cause a drastic reduction in activation (see Fig. 5b, bottom). Thus, the overall activation in LO to this level of scrambling was $82 \pm 18\%$ (standard error of the mean (SEM), $n = 9$) of the activation to whole images. However, in a minority ($28 \pm 9\%$ SEM, $n = 9$) of LO voxels there was a larger degree of reduction to the

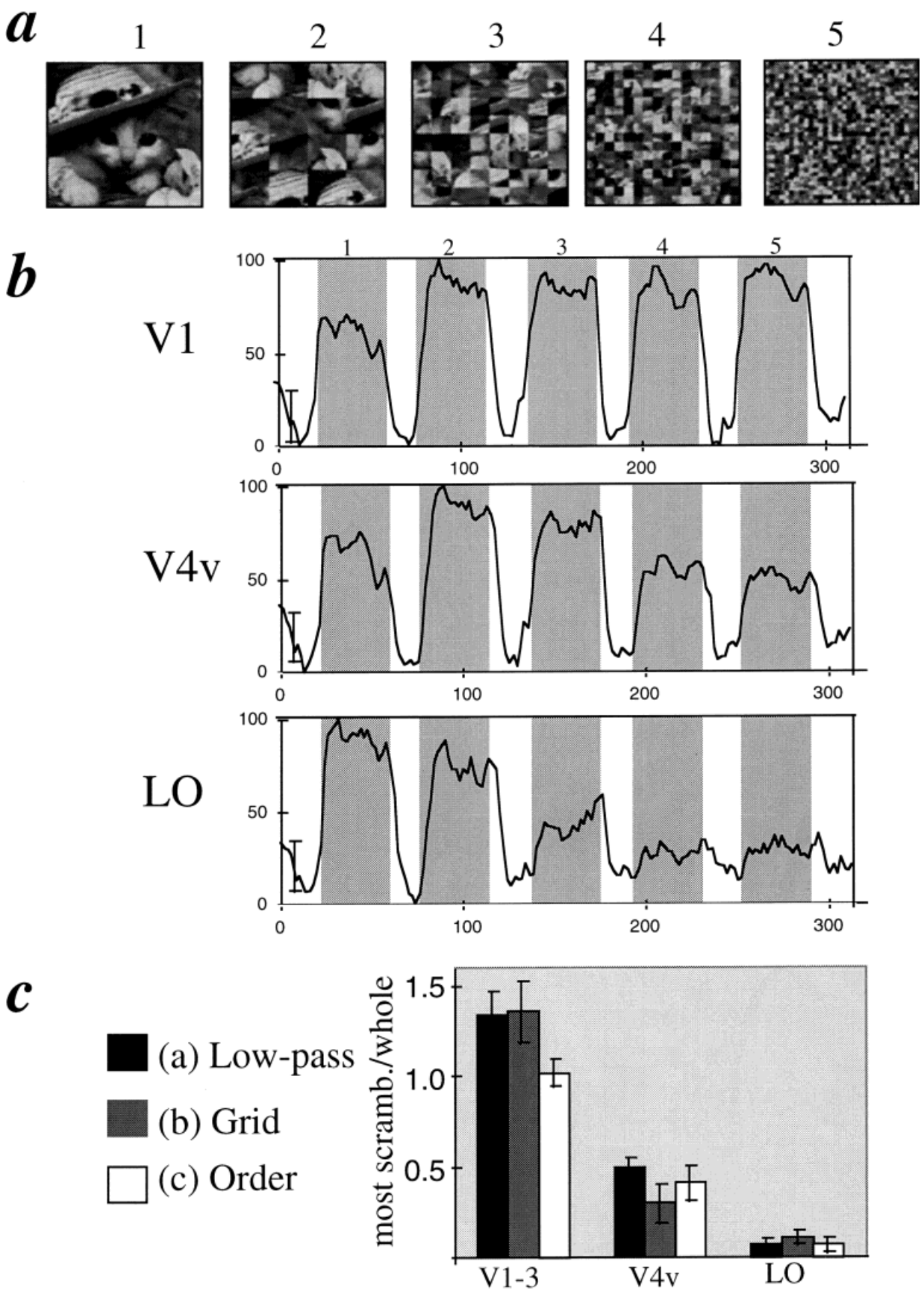
a
1
2
3
4
5
b
V1
V4v
LO
100
50
0
0
100
200
300
c
(a) Low-pass
(b) Grid
(c) Order
most scramb./whole
1.5
1.0
0.5
V1-3
V4v
LO

Figure 5.

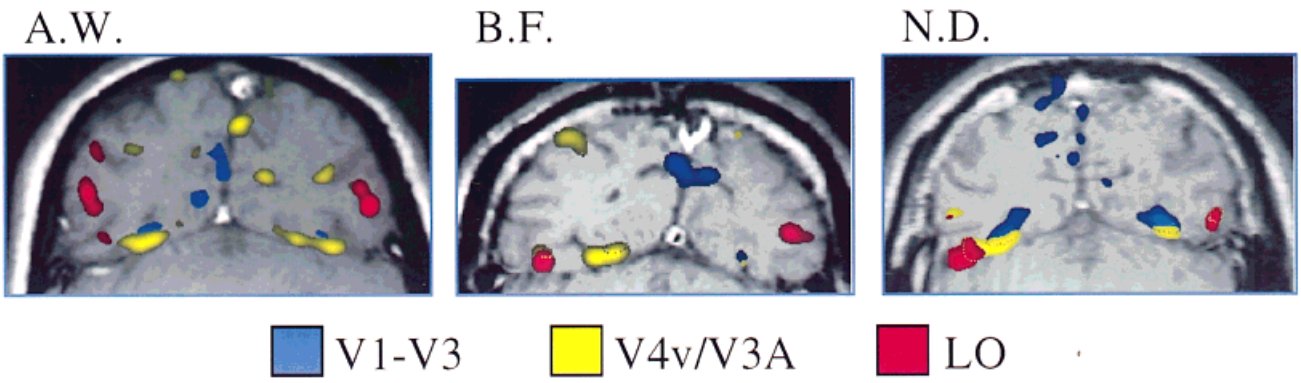

Figure 6.

Image-scrambling experiment. Activation maps. Superposition of the activation maps of the three functional foci obtained by regression analysis of the scrambling experiment shown in Figure 5, overlaid upon the T1-weighted high-resolution scans of 3 different subjects. The lightness of each color corresponds to the level of the regression coefficient (statistical significance $P < 1e\text{-}6$). Voxels below the threshold are not colored. Note the overall similarity in location and mediolateral arrangement of the three foci here and in Figure 2b.

same stimuli (32 ± 15% SEM, n = 9, of the maximal activation, time course not shown).

To further control for spatial frequency or edge effects, a second scrambling experiment (experiment b; 12 subjects) was conducted, in which images were similarly scrambled, but instead of filtering them with a low-pass filter, a 4 × 4 grid was overlaid upon the unscrambled images (see Methods for details). In another control experiment (experiment c; 5 subjects), the order of the epochs was permuted so that the highly scrambled epochs were presented first. The results of all these experiments were similar (see Fig. 5c), indicating that spatial frequency, number of edges, fatigue, or adaptation effects could not account for the scrambling results.

Combining the visual field and scrambling experiments

Since different criteria were used to delineate the three functional foci in the various experiments, we superimposed the maps obtained for object-related areas in the visual field and scrambling experiments. An example of this superposition for one subject in the three experiments is shown in Figure 7. As is evident, there was substantial overlap both in the bilateral LO region (Fig. 7, red) and in middle-tier object-related areas V4v and V3A (Fig. 7, yellow). Considering the low signal-to-noise ratios of fMRI and the variability encountered between experiments, it is most likely that these comparable regions represent the same cortical areas.

Figure 5.

a: Image-scrambling experiment (low-pass). Visual stimuli. An example of the pictures used in the experiment. Epochs of visual stimulation, which contained 20 different images (40 sec long), were alternated with blank epochs (20 sec long). Pictures were randomly scrambled into 16, 64, 256, and 1,024 blocks (epochs 2–5, respectively). Visual stimuli in all epochs were low-pass filtered (see Methods), but this is not apparent in these reduced images. Subjects were instructed to covertly name the images, including the scrambled images. b: Image-scrambling experiment (low-pass) activation time courses. Average time courses of 9 subjects derived from each of the distinct functional foci depicted in Figure 6. The x-axis denotes time in seconds and the y-axis denotes normalized fMRI signal strength. Error bars denote ± averaged SEM (n = 9). Note the increased scrambling sensitivity from V1 to LO. c: Histogram showing the scrambling index which equals: (average fMRI signal during most scrambled epoch—blank)/(average fMRI signal during unscrambled epoch—blank) for areas V1–3, V4v, and LO and for the different scrambling experiments (see Methods for details). Black, experiment a: low-pass filtered pictures; gray, experiment b: whole pictures including a grid; white, experiment c: same as experiment b, but order of presentation permuted. Note that LO is the most sensitive to image-scrambling, and that variability between experiments is much smaller compared to the different profiles of activation in each of the functionally defined areas.

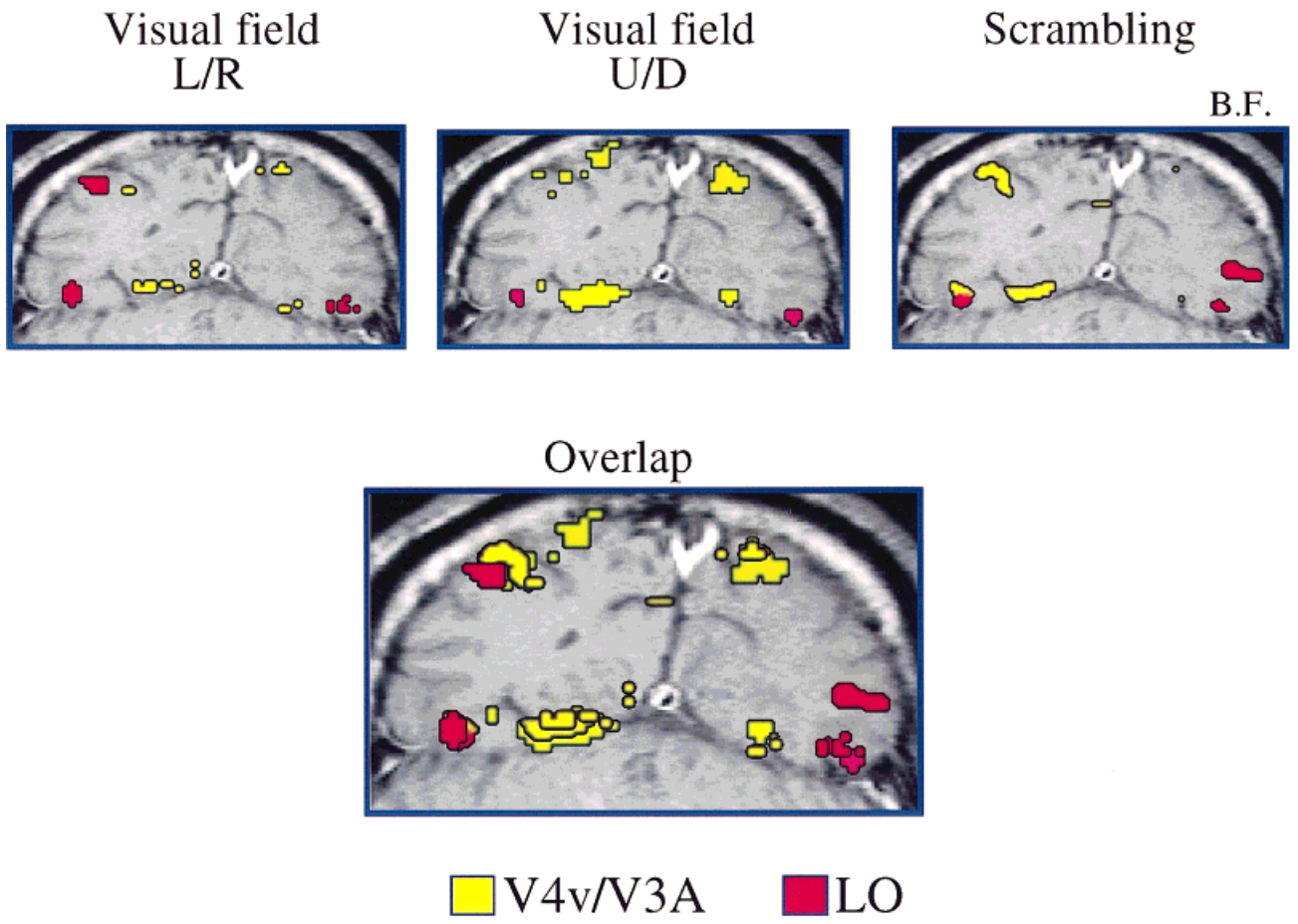

Figure 7.

Comparison of the foci of activation of object-related areas, in the different experiments. Top: Three activation maps of the same slice of a single subject in three different experiments ($P < 1e\text{-}4$). In the visual field L/R and U/D experiments, yellow denotes voxels which displayed both retinotopy and enhanced activation by images of objects compared to highly scrambled (1,024 blocks) versions of these images; red denotes voxels that exhibited preferential activation to images of objects but essentially no retinotopy. At left, results of the scrambling experiment. Yellow corresponds to voxels in which the activation declined only by a high degree of scrambling (256 and 1,024 blocks), and red corresponds to voxels in which the activation was decreased by an intermediate degree of scrambling (64 blocks). Bottom superposition of the activation foci from the three top images. Note that despite the fact that different criteria were used to delineate the foci of activation in the different experiments, the results are consistent. The activation foci defined in the different experiments are clustered together such that the yellow voxels corresponding to V3A/V4v are always located more medially with respect to the red LO voxels.

Anatomically, the LO complex is folded in a curved strip that can be bounded by three vertices. The corresponding Talairach coordinates [Talairach and Tournoux, 1988] of these three vertices are:

Left: Dorsal-posterior vertex: lateral -45 ± 2 mm, posterior: -77 ± 6 mm, inferior: 0 ± 4 mm; ventral-posterior vertex: -42 ± 7 mm, -71 ± 8 mm, -17 ± 3 mm; ventral-anterior vertex: -39 ± 7 mm, -52 ± 7 mm, -23 ± 5 mm (mean $\pm$ std; n = 9).
Right: Dorsal-posterior vertex: 41 ± 8 mm, -73 ± 7 mm, 5 ± 4 mm; ventral-posterior vertex: 41 ± 6 mm, -70 ± 7 mm, -18 ± 5 mm; ventral-anterior vertex: 41 ± 7 mm, -49 ± 4 mm, -20 ± 6 mm (mean $\pm$ 1std; n = 9).

It should be noted that the LO complex may extend even further anterior-ventrally, but due to susceptibility artifacts around the temporal lobes we were unable to image more anterior lateral regions.

One parameter that did exhibit some variability between the different experiments was the number of significant LO voxels that were detected. Typically, in

the visual field experiments, fewer voxels (L/R, 70 ± 31 voxels; U/D, 91 ± 45 voxels) were identified compared to the scrambling experiments (155 ± 54 voxels). The number of detected V4v voxels exhibited less variability (L/R, 131 ± 72; U/D, 186 ± 64 voxels; scrambling, 171 ± 67 voxels).

DISCUSSION

The results of the present study are compatible with a stagewise hierarchical scheme of object processing [e.g., Tanaka, 1996] leading from retinotopically organized local feature representations on the medial wall of the occipital pole, to a natural-object selective region which generalizes across visual field position, on the lateral aspect of the hemisphere.

The use of a combined visual field mapping paradigm allowed us to distinguish middle-tier areas V4v and V3A from LO. These areas showed enhanced activation to whole images of natural objects compared to highly scrambled images. While the former areas displayed clear retinotopy the latter did not. Moreover, the use of a gradual scrambling paradigm supports the functional distinction between areas V4v and LO: V4v was highly activated even when the image of the object was fragmented into 64 blocks, while activation in LO voxels was significantly reduced by this degree of image-scrambling. Thus, the present results suggest that human area V4v plays a role in intermediate shape representation and is activated by simpler and more localized visual features than LO.

The fact that most LO voxels remained active after the first scrambling stage indicates the domination of object fragments rather than whole-object representation in LO. This result is in line with several computational models [Biederman, 1987; Fukushima, 1988; Cave and Kosslyn, 1993] and physiological findings [Kobatake and Tanaka, 1994; Wachsmuth et al., 1994] emphasizing object-part representation as an important stage in object recognition. However, it should be noted that we did find a smaller subpopulation of LO voxels in which the activation was reduced with the first image-scrambling, suggesting a representation of entire objects, which can support shape representation as discussed by Edelman [1995, 1998].

Our experimental paradigm, which employed images of mixed object categories, did not directly explore the issue of category-specific representations [Martin et al., 1996; Ishai et al., 1997] such as face-selective areas [Allison et al., 1994; Puce et al., 1995; Kanwisher et al., 1997a]. However, our results are not necessarily in conflict with category-specific representations. It is possible that object fragments that are common to faces, inanimate objects, etc., are clustered in distinct anatomical subdivisions of the LO complex. Whether such clustering is based strictly on general object categories or follows more complex organizational rules remains to be studied.

It is important to emphasize that the relatively coarse spatial resolution inherent in our methodology may mask additional and more subtle subdivisions, both on the border of V4v and LO and within the LO complex itself. Indeed, some heterogeneity was previously observed in the functional selectivity within the LO complex [Malach et al., 1995].

Although there was an overall agreement in the definition of LO through the various experiments, a subtle effect was noted in which the activation was weaker during the visual field experiments compared to the scrambling experiments. One source for this variability might be the different tasks that the subject performed, i.e., naming the color of the fixation cross in the visual-field experiments, compared to naming the visual stimuli in the scrambling experiments. Alternatively, it may be that the constantly changing fixation cross that was present throughout the experiment introduced a constant activation that masked the object-selective responses. This issue requires further investigation.

The present results also help to clarify potential homologies between human and macaque ventral stream areas [Tootell et al., 1996; Van Essen and Drury, 1997; Tanaka, 1997], particularly the homology of LO. The bilateral field representation in LO argues against its homology to macaque area V4, which shows clear hemifield representation. Moreover, many LO voxels show selectivity to object fragments which appear compatible with the ''moderately complex'' receptive field properties found in the macaque infero-temporal (IT) cortex by Tanaka [1996]. Therefore, LO is likely to correspond to some subdivisions of posteriorly shifted macaque IT. Areas V4v and V3A then become likely human homologues of macaque areas V4 and V3A, respectively. More anterior regions, presumably containing whole-object representations [e.g., Martin et al., 1996; Kanwisher et al., 1997a; Ishai et al., 1997], may correspond to anterior subdivisions of macaque IT.

ACKNOWLEDGMENTS

We thank P. Harvey for his crucial contribution to the development of the EPI sequences, E. Okon for constructing the optical system, and G. Avidan, O. Scwartz, and I. Kahn for their help in conducting the

experiments. This work was supported by ISF grant 17131/97-1 (to R.M. and T.K.).

REFERENCES

Allison T, Ginter H, McCarthy G, Nobre AC, Puce A, Luby M, Spencer DD (1994): Face recognition in human extrastriate cortex. J Neurophysiol 71:821–825.

Baker JR, Hoppel BE, Stern CE (1993): Dynamic functional imaging of the complete human cortex using gradient-echo and asymmetric spin-echo echo planar magnetic resonance imaging. Soc Magn Reson Med Abstr 1400.

Biederman I (1987): Recognition-by-components: A theory of human image understanding. Psychol Rev 94:115–147.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996): Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Cave CB, Kosslyn SM (1993): The role of parts and spatial relations in object identification. Perception 22:229–248.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996): Mapping striate and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci USA 93:2382–2386.

Edelman S (1995): Representation, similarity, and the chorus of prototypes. Minds Machines 5:45–68.

Edelman S (1998): Representation is a representation of similarity. Behav Brain Sci (in press).

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994): fMRI of human visual cortex. Nature 369:525.

Fox PT, Miezin FM, Allman JM, Van Essen DC, Raichle ME (1987): Retinotopic organization of human visual cortex mapped with positron-emission tomography. J Neurosci 7:913–922.

Friston KJ, Homes A, Worsley K, Frith C, Frackwowiak R (1995): Statistical parametric maps in functional imaging, a general linear approach. Hum Brain Mapp 2:189–210.

Fukushima K (1988): Neocognitron: A hierarchical neural network capable of visual pattern recognition. Neural Networks 1:119–130.

Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R (1998): Convergence of visual cues in object related areas of the human occipital lobe. Neuron (in press).

Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994): The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. J Neurosci 14:6336–6353.

Haxby JV, Ungerleider LG, Horwitz B, Maisog JM, Rapoport SI, Grady CL (1996): Face encoding and recognition in the human brain. Proc Natl Acad Sci USA 93:922–927.

Ishai A, Ungerleider LG, Martin A, Maisog JM, Haxby JV (1997): fMRI reveals differential activation in the ventral object recognition pathway during the perception of faces, houses and chairs. Neuroimage 5:149.

Kanwisher N, Chun MM, Ledden P (1996): Functional imaging of human visual recognition. Cogn Brain Res 5:55–67.

Kanwisher N, McDermott J, Chun MM (1997a): The fusiform face area: A module in human extra-striate cortex specialized for face perception. J Neurosci 17:1–10.

Kanwisher N, Woods RP, Iacoboni M, Mazziotta JC (1997b): A locus in human extrastriate cortex for visual shape analysis. J Cogn Neurosci 9:133–142.

Kobatake E, Tanaka K (1994): Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol 71:856–867.

Malach R, Reppas JB, Benson R, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RBH (1995): Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA 92:8135–8139.

Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996): Neural correlates of category-specific knowledge. Nature 379:649–652.

Puce A, Allison T, Gore JC, McCarthy G (1995): Face-sensitive regions in human extrastriate cortex studied by functional MRI. J Neurophysiol 74:1192–1199.

Reyment R, Joreskog K (1993): Applied Factor Analysis in the Natural Sciences, Cambridge, MA: Cambridge University Press.

Schneider W, Noll DC, Cohen JD (1993): Functional topographic mapping of the cortical ribbon in human vision with conventional MRI scanners. Nature 365:150–153.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995): Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Sergent J, Ohta S, Macdonald B (1992): Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain 115:15–36.

Talairach J, Tournoux P (1988): Co-Planar Stereotaxic Atlas of the Human Brain, New York: Thieme Medical Publishers.

Tanaka K (1996): Inferotemporal cortex and object recognition. Annu Rev Neurosci 19:109–139.

Tanaka K (1997): Mechanisms of visual object recognition: Monkey and human studies. Curr Opin Neurobiol 523–529.

Tootell RBH, Dale AM, Sereno MI, Malach R (1996): New images from human visual cortex. Trends Neurosci 19:481–489.

Tootell RBH, Mendola J, Hadjikhani N, Ledden P, Liu A, Reppas J, Sereno M, Kwong K, Dale AM (1997): Functional analysis of V3a and related visual areas in human visual cortex. J Neurosci 71:7060–7078.

Ungerleider LG, Haxby JV (1994): "What" and "where" in the human brain. Curr Opin Neurobiol 4:157–165.

Van Essen DC, Drury H (1997): Structural and functional analyses of human cerebral cortex using a surface based atlas. J Neurosci 17:7079–7102.

Wachsmuth E, Oram MW, Perrett DI (1994): Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque. Cereb Cortex 4:509–522.