Object Representations in the Temporal Cortex of Monkeys and Humans as Revealed by Functional Magnetic Resonance Imaging

Andrew H. Bell,1 Fadila Hadj-Bouziane,1 Jennifer B. Frihauf,1 Roger B. H. Tootell,1,3 and Leslie G. Ungerleider1
1Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland; 2Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts; and 3Department of Radiology, Harvard Medical School, Boston, Massachusetts

Submitted 10 June 2008; accepted in final form 25 November 2008

Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RB, Ungerleider LG. Object representations in the temporal cortex of monkeys and humans: representations based on animacy, semantic categories, or visual features. J Neurophysiol 101: 688–700, 2009. First published December 3, 2008; doi:10.1152/jn.90657.2008. Increasing evidence suggests that the neural processes associated with identifying everyday stimuli include the classification of those stimuli into a limited number of semantic categories. How the neural representations of these stimuli are organized in the temporal lobe remains under debate. Here we used functional magnetic resonance imaging (fMRI) to identify correlates for three current hypotheses concerning object representations in the inferior temporal (IT) cortex of monkeys and humans: representations based on animacy, semantic categories, or visual features. Subjects were presented with blocked images of faces, body parts (animate stimuli), objects, and places (inanimate stimuli), and multiple overlapping contrasts were used to identify the voxels most selective for each category. Stimulus representations appeared to segregate according to semantic relationships. Discrete regions selective for animate and inanimate stimuli were found in both species. These regions could be further subdivided into regions selective for individual categories. Notably, face-selective regions were contiguous with body-part-selective regions, and object-selective regions were contiguous with place-selective regions. When category-selective regions in monkeys were tested with blocks of single exemplars, individual voxels showed preferences for visually dissimilar exemplars from the same category and voxels with similar preferences tended to cluster together. Our results provide some novel observations with respect to how stimulus representations are organized in IT cortex. In addition, they further support the idea that representations of complex stimuli in IT cortex are organized into multiple hierarchical tiers, encompassing both semantic and physical properties.

INTRODUCTION

The inferior temporal (IT) cortex has been repeatedly implicated in the visual perception and identification of complex stimuli (e.g., Logothetis and Sheinberg 1996; Tanaka 2000; Ungerleider and Mishkin 1982). However, the question of how individual stimulus representations are organized remains under vigorous debate. Evidence from behavioral studies has shown that subjects can make categorical judgments about such stimuli (e.g., a face vs. a car) more readily than they can name its specific identity (e.g., my neighbor vs. my co-worker, a Toyota vs. a Honda) (Grill-Spector and Kanwisher 2005; Murphy and Brownell 1985; Rosch et al. 1976). This finding suggests that object recognition proceeds through multiple levels of classification before a stimulus is fully characterized (see Op de Beeck et al. 2008; Ullman 2007).

There are several hypotheses about how IT cortex is organized. One hypothesis proposes that IT cortex is divided into separate regions coding for animate and inanimate stimuli. This idea arose from the demonstration of specific visual deficits related to the recognition and naming of living versus nonliving stimuli (e.g., Farah et al. 1991; Warrington and Shallice 1984). More recently, several imaging studies have found cortical regions selectively activated by animate versus inanimate stimuli (e.g., Beauchamp et al. 2002; Mahon et al. 2007; Mitchell et al. 2002; Schultz et al. 2005; Wheatley et al. 2007), and a recent neurophysiological study reported that responses of monkey IT neurons are segregated primarily according to this distinction (Kiani et al. 2007).

A second hypothesis proposes that IT cortex contains a number of modules specialized for individual “semantic” categories (i.e., groups of objects that share similar functions but not necessarily similar visual features). This idea is supported by neuroimaging studies that have reported discrete regions specialized for faces (e.g., Sergent et al. 1992; Haxby et al. 1994; Puce et al. 1996; Kanwisher et al. 1997; Hadjikhani and de Gelder 2003), body parts (e.g., Downing et al. 2001; Peelen and Downing 2005; Schwarzlose et al. 2005), and places (e.g., Epstein and Kanwisher 1998). Damage that includes such regions often leads to specific perceptual deficits related to these categories, such as prosopagnosia following damage to the fusiform gyrus (Bodamer 1947; Damasio et al. 1982; De Renzi et al. 1994; Meadows 1974) and topographical disorganization following damage to the parahippocampal gyrus (Barth 1998; Epstein et al. 2001; Luzzi et al. 2000). More recently, a small number of neuroimaging studies have reported regions responsive to faces and body-parts in monkey IT cortex (Pink et al. 2005; Tsao et al. 2003a, 2006).

A third hypothesis assumes a fine-grained, columnar map of visual features across IT cortex. This model is based largely on electrophysiological and optical imaging studies (Brincat et al. 2004, 2006; Fujita et al. 1992; Tanaka 2003; Tanaka et al. 1991). According to this hypothesis, an object would be represented by the correlated activation of neuronal populations in different columns. A recent fMRI study in monkeys provided support for this hypothesis by reporting different spatially
distributed patterns of functional magnetic resonance imaging (fMRI) activity in response to each of three computer-generated shapes (Op de Beeck et al. 2007).

Attempts to dissociate these three hypotheses are complicated by the fact that the evidence in support of each derives from different species using a variety of experimental techniques. Here we attempt to relate these three models using a single experimental approach to facilitate the investigation of object representations in IT cortex. Using fMRI in both monkeys and humans, we first identified regions in both species selective for animate and inanimate stimuli as well as for individual categories. We then calculated the relative selectivity of the category-selective regions to determine if they showed any tuning for their nonpreferred categories. Finally to assess the degree to which individual category-selective regions show feature selectivity, we examined whether individual voxels within these regions showed preferences for individual exemplar images.

METHODS

Animal experiments

Scanning of nonhuman primates was conducted at the Martinos Center for Biomedical Imaging at the Massachusetts General Hospital (MGH). All procedures conformed to MGH guidelines and were approved by the MGH Center for Comparative Medicine. Four male rhesus macaques (3–10 kg) were implanted with a MR-compatible head post during a single aseptic surgical session. After a recovery period of ≥3 wk, monkeys were trained to maintain fixation within a 2° window surrounding a central fixation point (FP) while blocks of stimuli were presented using procedures similar to those described previously (Tsao et al. 2003a,b; Vanduffel et al. 2001).

Visual stimuli were divided into two conceptual classes, animate and inanimate, each containing two categories of stimuli (Supplementary Fig. S1A).1 Animate stimuli included monkey faces and monkey body-parts (torsos, paws, etc.). Inanimate stimuli included objects and places, both of which were taken from the monkeys’ surroundings. We chose these categories as representative members of animate and inanimate stimuli because they are reportedly reliable in producing category-selective regions in human neuroimaging studies (see Downing et al. 2006). Similarly in monkey physiology, when the neuronal responses to a large number of different stimuli (>1,000) have been organized into a category-based structure, the animate category most readily segregated into faces, bodies, and hands (see Kiani et al. 2007).

Each of the four categories (faces, body parts, objects, and places) was presented in separate blocks lasting 32 s. Each block included 16 intensity-matched grayscale images (presentation time of 2 s each) randomly selected from a set of 48 images/category. Images were resized so that their maximum dimension was set to ~22° and were presented on a background of random grayscale dots for a total uniform stimulus size of 40° across. Individual scanning runs began and ended with blocks of baseline fixation (neutral gray background, intensity-matched to the average level of all other blocks). The remainder of each run was composed of a random ordering of the following conditions: upright monkey faces, inverted monkey faces, upright places, inverted places, objects, monkey body parts, and a random dot pattern condition. In the case of monkey S, the inverted conditions were omitted and the stimulus blocks were interleaved with blocks of baseline fixation (30-s blocks). This monkey’s data were qualitatively and quantitatively similar to the datasets of the other three subjects, so all data are considered together. Runs lasted ~6 min, and monkeys completed 18–32 runs/session. Although monkeys were required to fixate for a minimum of 85% of the duration of the run for the data to be included, their average performance was consistently much higher (95 ± 1%). Short breaks in fixation lasting <200 ms (e.g., due to blinks) were ignored. Eye position was monitored using a MR-compatible pupil-reflection system (ISCAN-RK-726 PCI; Cambridge, MA) with a spatial resolution of 0.25° and a temporal resolution of 120 Hz.

In two of the animals (monkeys J and R), we conducted a second experiment in which we mapped the activity produced by individual exemplars (i.e., single images). For each category, two visually dissimilar exemplars, selected from the original stimulus sets, were tested, in independent blocks lasting 32 s each. These stimuli were prepared in an identical manner as in the previous experiment, randomly shifted 1–2° every 2 s to prevent habituation, and selected on the basis of their uniqueness from one another (in terms of shape, texture, etc. as defined by visual inspection). To approximate the degree of visual dissimilarity across and within each category, we first identified the images for each category to produce an “average” face, place, etc. We then subtracted this from each of the exemplars to calculate the variance in pixel intensity associated with each category (Supplementary Table S1). Greater values indicate greater variability among the images. A similar analysis was conducted to approximate the visual dissimilarity between the two exemplars used in the second experiment. The two images were subtracted from one another and the average intensity difference was calculated for each category, on a pixel-by-pixel basis (Supplementary Table S1).

Human experiments

All human experimental procedures were reviewed and approved by the National Institute of Mental Health’s (NIMH) Institutional Review Board and conducted at the NIMH. Nine subjects (mean age: 25 ± 4; 4 females) provided informed consent, had normal or corrected-to-normal vision, had no history of neurologic or psychiatric illness, and reported no difficulties in distinguishing basic colors. To facilitate comparisons across species, the human subjects performed a foveal discrimination task. This task was irrelevant to the study, but it ensured that the subject’s attention was directed toward the central FP. This task required them to report, with a button press, when the color of the FP changed from red to green or vice versa. Subjects were instructed to maintain fixation on the FP but eye position was not recorded.

Stimulus blocks lasted 16 s (8 images/block, 1-s presentation time followed by 1 s of blank interstimulus interval) and each block included the following conditions: human faces, human body parts, objects, places, and a random grayscale dot pattern condition identical to that used in the monkey experiments. Stimuli were randomly selected from a set of 32 possible images/category and were prepared in an identical fashion to those used for the monkey experiments. Stimulus blocks were interleaved with blocks of baseline fixation (neutral gray background). Each run consisted of 11 blocks, lasting a total of ~3 min. Each subject completed between six and eight runs over a single scanning session.

Scanning procedures

All monkey experiments took place in either a Siemens Allegra 3T or TimTrio 3T scanner. To improve the contrast-to-noise ratio, monkeys were injected with an iron-based contrast agent (mono-crystalline iron nanoparticle, MIcron) prior to each scan session (see Leite et al. 2002; Vanduffel et al. 2001 for details). Functional scans for monkeys E, J, and R were acquired in the Allegra scanner at 1.25-mm isotropic resolution with a 12-cm send/receive surface coil and a multi-echo sequence (2 echo/TR with alternating phase-encoding direction; 28 coronal slices, TR = 4 s, TE = 30 and 71 ms, FOV =
80 mm, flip angle = 90°, 64 × 64). Periodically throughout these scan sessions, 2D gradient echo field maps were acquired (28 slices, 1.25-mm isotropic resolution, TR = 4 s, flip angle = 30°, 64 × 64) for use in the reconstruction of the multi-echo data. Functional scans for monkey S were acquired in the TimTrio scanner at 1.25 × 1.25 × 1.90-mm resolution using a single-echo gradient echo sequence (12-cm send/receive coil, 33 coronal slices, TR = 2.5 s, TE = 28 ms, FOV = 80 mm, flip angle = 90°, 64 × 64). High-resolution anatomic scans for monkeys E, J, and R were acquired using an MP-RAGE sequence at 3T (0.35-mm isotropic resolution, TR = 2.5 s, TE = 4.35 ms, flip angle = 8°, 384 × 384). High-resolution anatomical scans for monkey S were acquired in a Bruker scanner using an MDEFT sequence at 4.7T (0.5-mm isotropic resolution, TR = 2.6 s, TE = 4.6 ms, flip angle = 14°, 256 × 256).

Human functional data were collected in a GE 3T scanner. Single-echo functional images were acquired at a resolution of 3.5 × 3.5 × 5.0 mm (3.75 × 3.75 × 5.00 mm for 1 subject; 28 coronal slices, TR = 2 s, TE = 30 ms, FOV = 220 mm, flip angle = 90°, 64 × 64). Anatomical scans were collected from each subject during the same session as the functional data (3D MP-RAGE sequence; 1–2 scans, 0.85 × 0.85 × 1.20 mm; TR = 7.5 s, TE = 3–16 ms, flip angle = 6°, 224 × 192). Anatomical scans for monkeys and humans were segmented into inflated and flattened cortical surface maps using FreeSurfer (Dale et al. 1999; Fischl et al. 1999, 2001; Segonne et al. 2004). Aligning the monkey functional data to the surface maps was done by first running an automated alignment program (FLIRT/FSL) (Jenkinson and Smith 2001; Smith et al. 2004), which included a step that resampled the functional data to the resolution of the anatomical scans, followed by a series of manual corrections. Note, however, that all quantitative and qualitative analyses are based on the original, unresampled, data. Alignment of the human functional data were done by registering each subject’s data to a Talairach atlas (Talairach and Tournoux 1988) using automated procedures (AFNI) (Cox 1996). Functional overlays for the inflated and flattened cortical surface maps, illustrated in the figures, were generated using AFNI/SUMA.

Data analysis
Analysis of both monkey and human data was done using a combination of AFNI (Cox 1996) and user-generated programs written in Matlab (Mathworks). Prior to data analysis, all functional data obtained with the multi-echo sequence were reconstructed off-line into individual volumes for each TR. This process included a static field correction and a dynamic B0 correction that partially compensated for distortions in the field caused by the monkey repositioning itself during a session.

A total of 40,334 functional volumes were collected from four monkeys during 17 sessions. A total of 5,280 volumes were collected from nine human subjects (6–8 run/subject). Functional datasets for both monkeys and humans were de-spiked (removal of data points >2 SD from the mean), motion-corrected, and spatially smoothed with a Gaussian kernel (2-mm full-width half-max for monkeys; 6 mm for humans). Signal intensity was normalized to the overall mean intensity for each run, and time series data were grouped into a single dataset for each subject. The human blood-oxygen-level-dependent (BOLD) response was modeled using a gamma-variate response function (Boynton et al. 1996; Friston et al. 1995). The monkey MION response was modeled using a modified gamma-variate function approximating changes in cerebral blood volume (Leite et al. 2002; Mandeville et al. 1999). For ease of presentation, data derived with this function have been inverted so that an increase in signal intensity indicates an increase in activation—as in BOLD results. Statistical overlays were generated from multiple regression analyses using the convolved functions for each stimulus condition as regressors of interest. Regressors of no interest, including motion-correction factors, baseline condition, and linear drifts, were also included in these analyses. After the individual regression analyses, data from the human subjects were compiled into a single dataset to improve statistical power. Unless otherwise stated, all human-derived results are based on this grouped dataset.

Conjunction mapping and region of interest analysis
Many of the analyses here are based on conjunction maps combining three individual contrasts (e.g., upright faces vs. places + upright faces vs. objects + upright faces vs. body parts; see Supplementary Fig. S1B). In the monkey data, each contrast was first thresholded to $P < 0.05$ such that the final statistical threshold (after all 3 contrasts were overlapped) approximated $P < 0.05$ (0.363 = 0.047). In the human data, thresholding the individual contrasts prior to combining them resulted in a significant drop in the number of activated voxels. Therefore in the human experiments, voxels were not thresholded prior to their combination. As a control, we repeated all the analyses on the monkey data using this less conservative thresholding technique and found the results to be virtually identical. In both cases, the output of the combination was masked with a contrast of all four conditions (voxels activated by any of the 4 conditions) versus the random dot pattern (to control for low-level influences caused by simple visual stimulation), thresholded to at least $P < 10^{-15}$ uncorrected for the monkeys and $P < 0.01$ uncorrected for the humans (see Supplementary Fig. S1, C and D). Any regions of contiguous voxels <20 mm$^3$ (monkeys) or 414 mm$^3$ (humans) in volume were excluded.

Quantitative analyses of the functional data were performed using receiver-operating characteristic (ROC) curves based on regions of interest (ROIs). To account for the large differences in volume for the various category-selective regions that could potentially affect the results of the ROC analyses, we normalized the size of each ROI to a volume approximating that of the average size of a face-selective region (~40–100 mm$^3$ for the monkey data and ~1,000–2,000 mm$^3$ for the human data) using the following procedure: within each individual category-selective region, the voxel with the greatest signal intensity was identified (i.e., seed voxel). All voxels within a radius of 7–8 mm (monkeys) or 10–15 mm (humans) of the seed voxel were added to yield a “spherical” mask. Any voxels within the sphere that were not identified as being selective for that particular category were removed to yield the final ROI. In the case of the monkey subjects, voxels located posterior to area TEO, defined in each subject as voxels caudal to the posterior edge of the posterior middle temporal sulcus (Boussaoud et al. 1991), were also removed. In all analyses, the magnitude of the response was defined as the peak-normalized signal intensity over each block, ignoring the first two TRs of each block to allow for a steady-state response. All results are expressed as means ± SE.

RESULTS

Representations of animate versus inanimate stimuli in IT cortex

To investigate the degree to which IT cortex in monkeys and humans might be organized according to the distinction between animate and inanimate stimuli, we constructed statistical maps contrasting upright faces and body parts (i.e., animate stimuli) versus objects and upright places (i.e., inanimate stimuli). Figure 1 reveals a striking consistency across subjects for voxels preferring animate versus inanimate stimuli in IT cortex. In monkeys, voxels preferentially activated by animate stimuli were located in two regions along the inferior bank of the superior temporal sulcus (sts); one anteriorly in area TE and another posteriorly near area TEO. On the other hand, voxels preferentially activated by inanimate stimuli occupied regions
elsewhere within the sts and along the ventral surface of IT cortex, extending medially to the occipitotemporal sulcus (ots). In humans, voxels preferentially activated by animate stimuli were found along the fusiform gyrus, lateral to those preferentially activated by inanimate stimuli, which were concentrated near/within the parahippocampal gyrus. Notably, the two species showed a similar topographic correspondence between these two classes of stimuli: regions activated by animate stimuli were located more dorsolaterally, and regions activated by inanimate stimuli were located more ventromedially.

**Representations of semantic categories in IT cortex**

The preceding analysis revealed that IT cortex in monkeys and humans shows a robust and consistent segregation for the animate and inanimate stimuli tested. To determine if these regions could be further subdivided into regions selective for more specific semantic concepts, we constructed individual statistical maps for each possible contrast among the four categories (e.g., faces vs. places, faces vs. objects, etc.). Examples of these individual contrasts for both monkeys and humans are shown in Supplementary Fig. S2. Contrasting faces versus places, for instance, revealed two bilateral face-responsive regions: one anteriorly in area TE and one posteriorly near area TEO. Place-responsive regions were observed along the ventral inferior temporal and parahippocampal gyri. The same contrast in humans showed an elongated bilateral face-responsive region along the fusiform gyrus and a second, smaller, bilateral region in the inferior occipital gyrus (not visible in
Supplementary Fig. S2). Place-responsive regions in humans were located near/within the parahippocampal gyrus. However, by definition, these contrasts only highlight voxels that are more responsive to one of the four categories versus another. To improve the specificity of this analysis, we contrasted each category with the remaining three using a conjunction analysis approach (e.g., faces vs. objects/places/body parts; see METHODS for details). Figure 2 shows the results of this analysis for monkeys E and J and the grouped human dataset (examples of individual human subjects are shown in Supplementary Fig. S3). Despite some idiosyncratic differences with respect to the location and relative size of category-selective regions across subjects, there were several notable consistencies. In monkeys, face-selective regions were found anteriorly in area TE and posteriorly near area TEO, as described previously (Pinsk et al. 2005; Tsao et al. 2003a, 2006). Body-part-selective regions were located adjacent to face-selective regions within IT cortex, both anteriorly and posteriorly (Pinsk et al. 2005). Object-selective regions were found throughout the sts and the IT gyrus. Place-selective regions were found in the anterior portion of the sts (within area TE), along the ventral surface of IT cortex (within or surrounding the ots), and within the parahippocampal gyrus in a subset of monkeys (3/4 subjects; not shown).

In humans, category-selective regions were restricted to the ventral surface of the temporal lobes. Face-selective regions were found bilaterally in the fusiform gyrus and, as in monkeys, body-part selective regions were located in close proximity, also along the fusiform gyrus. Place- and object-selective regions were located adjacent to one another near/within the parahippocampal gyrus with the former located more medially.

Examining the intermediate stages of the conjunction maps also revealed some intriguing trends. Figure 3 plots the size (in number of voxels) of the four most consistent face-selective regions for monkey E as a function of the contrast used to define them. In all four locations, faces versus places produced the largest ROIs, and faces versus body parts produced the smallest. Thus although many voxels showed greater responses to faces compared with places, most of these same voxels did not show greater responses to faces compared with body parts. This finding was consistent across all four monkeys (Fig. 3B). In fact, the reduction in size with the faces versus body parts contrast was so pronounced that a number of regions visible with other contrasts disappeared when the faces versus body-parts contrast was used. Of the 16 initial face-selective ROIs in IT cortex revealed by the faces versus places contrast, only 9 remained when all contrasts were used. A similar but weaker
effect was observed in monkeys for place-selective regions, such that places versus faces produced the largest ROIs and places versus objects produced the smallest (Fig. 3C). Thus contrasting an animate category with an inanimate one (and vice versa) produced more selective voxels compared with contrasting an animate category with another animate category. When we conducted the same analysis on the face-selective regions for the individual human subjects’ datasets, a similar albeit weaker trend emerged (Fig. 3D). Thus in a face-selective ROI, it was significantly easier to distinguish a face from a place response, compared with distinguishing a face from a body-part response. We performed the same analysis on ROIs selective for the other three categories (Fig. 4, B–D, F–H, and J–L). ROIs selective for faces showed the strongest dissociation from faces and the weakest from objects (Fig. 4, D, H, and L). ROIs selective for objects showed the strongest dissociation from faces and the weakest from places (Fig. 4, C, G, and K). Finally, ROIs selective for body parts showed the strongest dissociation from faces and places and the weakest from objects (Fig. 4, B, F, and J). These results highlight that faces are the most distinguishable, in terms of their response in IT cortex, compared to the other three categories.

When we compared these findings from monkeys to the results of a similar analysis done on category-selective ROIs identified in the nine individual human subjects. Data obtained from each subject’s ROIs (e.g., Supplementary Fig. S3) were grouped to produce a large enough dataset for an effective ROC analysis. Qualitatively, category-selective regions in humans showed stronger selectivity for their preferred category compared with monkeys (Fig. 5). Specifically, responses to the nonpreferred categories in humans were much weaker than the responses to the preferred categories. Furthermore, compared with monkeys, the responses to the nonpreferred categories in humans were more similar to one another. This possible species difference was discussed in the following text.

**Responses within category-selective regions to their nonpreferred categories**

The preceding data show that regions selective for animate versus inanimate stimuli can be further segregated into regions coding for faces, body parts, objects, and places in both monkeys and humans, consistent with previous reports (e.g., Downing et al. 2001; Epstein and Kanwisher 1998; Kanwisher et al. 1997; Tsao et al. 2003a). To what degree are these category-selective regions specialized or “tuned” for their particular categories? We addressed this question by assessing the relative selectivity of these regions for their nonpreferred categories.

Figure 4, A–H, shows the normalized responses and corresponding ROC curves for representative ROIs from monkey E. If we examine the face-selective ROI (Fig. 4, A and E), we see that the largest area under the ROC curve is for places and the smallest is for body parts, indicating that the responses for faces are most dissociable from those for places and least dissociable from those for body parts. The identical trend was observed in the remaining face-selective ROIs of monkey E and the majority of face-selective ROIs across the four monkeys (Fig. 4I)—although one difference (the responses to faces vs. objects and faces vs. body parts) did not reach statistical significance. Thus in a face-selective ROI, it was significantly easier to distinguish a face from a place response, compared with distinguishing a face from a body-part response. We performed the same analysis on ROIs selective for the other three categories (Fig. 4, B–D, F–H, and J–L). ROIs selective for places showed the strongest dissociation from faces and the weakest from objects (Fig. 4, D, H, and L). ROIs selective for objects showed the strongest dissociation from faces and the weakest from places (Fig. 4, C, G, and K). Finally, ROIs selective for body parts showed the strongest dissociation from faces and places and the weakest from objects (Fig. 4, B, F, and J). These results highlight that faces are the most distinguishable, in terms of their response in IT cortex, compared to the other three categories.

**Representations of individual exemplars within category-selective regions**

The existence of anatomically distinct, functional modules selective for semantic categories does not, by itself, explain how the brain can distinguish between individual exemplars from a given category. Neurophysiological experiments have
proposed the existence of sub-regions or “columns” in IT cortex selective for visual features the activity of which is presumably combined to represent an individual stimulus (see Tanaka 2003). What is the relationship between these putative feature columns and the category-selective regions revealed by fMRI? To investigate this question, we looked within category-selective regions for individual voxels that showed significant differences in their responses to two visually dissimilar exemplars from the same category.

Figure 6 shows flattened views of the left sts of monkeys J and R, respectively (right sts shown in Supplementary Fig. S4). Individual panels show the category-selective regions for the four categories identified in the previous experiment, now colored according to their preference for the two exemplars (shown on the right). If a category-selective region was sensitive only to a stimulus’ membership in a particular category (i.e., no feature selectivity), one might predict that the two exemplars would produce similar responses across the region. On the other hand, if a category-selective region was sensitive to a stimulus’ visual features, one might predict that visually distinct exemplars from the same category would produce topographically disparate responses within the region.

Both monkeys showed several examples of voxels within category-selective regions that preferred one exemplar to the other. Furthermore, voxels with the same preference tended to cluster together—suggesting that there is some topographical organization based on visual features and/or retinotopic differences within the category-selective regions. To measure the reliability of this effect, we divided the data into even and odd

---

**FIG. 4.** Relative selectivity of category-selective regions to their nonpreferred categories, in monkeys. A–D: normalized responses to the 4 categories within representative category-selective regions of interest (ROIs) for monkey E. E–H: corresponding receiver-operating characteristic (ROC) curves for the category-selective ROIs shown in A–D. An area under the curve of 1.0 indicates that the response distributions for the 2 categories are completely dissociable. A value of 0.5 (diagonal line) indicates that they are overlapping, and thus the ability to assign a given response to the correct category would be at chance (50%). Inset values indicate area under the ROC curve for the given contrast. I–L: average area under the ROC curves for all category-selective regions identified in the 4 monkeys. Statistical comparisons between pairs of contrasted categories for the population performed using paired Wilcoxon rank sum tests (*, P < 0.05, **, P < 0.01, SE indicated by the error bars). ---, the 0.50 (i.e., chance) level.
runs and repeated the analysis (Supplementary Fig. S5). Correlating the responses on a voxel-by-voxel basis for the even and odd runs revealed a strong spatial correspondence between voxels showing a significant difference to exemplar 1 versus exemplar 2 (see Supplementary Fig. S6D), further reinforcing the strength of this observation.

**Face-selective regions in IT cortex**

Given the robust nature of face-selective regions in both monkeys and humans, it has been suggested that faces might represent a “special” category in terms of visual processing. In our monkeys, faces consistently produced two bilateral face-selective regions: one anteriorly in area TE, and another posteriorly near area TEO (Fig. 2), a finding consistent with previous reports (Pinsk et al. 2005; Tsao et al. 2003a). Additional subdivisions of these two regions have recently been described (Moeller et al. 2008). Do the anterior and posterior face-selective regions play different roles in face processing? To begin addressing this question, we further characterized these two regions by examining both their relative selectivity and their sensitivity to face inversion. The presence of an inversion effect would imply that these regions are sensitive to the holistic quality of a face and not just the presence of a collection of specific visual features (Epstein et al. 2006; Passarotti et al. 2007; Rossion and Gauthier 2002; Yovel and Kanwisher 2005).

Figure 7A shows the inflated surfaces of the right and left hemispheres of *monkey E*, colored according to the selectivity index (SI) of each voxel. Greater SI values indicate voxels that showed a larger difference in response magnitude between their preferred category relative to the average of the responses to the three remaining categories (i.e., “more selective”). The individual face-selective ROIs are indicated by the red outlines. Do face-selective regions in area TE include voxels with greater selectivity indices compared with face-selective regions near area TEO? When the average selectivity indices for all face-selective regions were calculated for all four monkeys, the anterior face-selective regions were more selective (i.e., showed sharper tuning) than did the posterior face-selective regions (Wilcoxon signed rank test; \( P < 0.05 \)). This is consistent with the general observation that as one moves anterior along the ventral pathway, neuronal selectivity increases (Boussaoud et al. 1991; Gross et al. 1972; Kobatake and Tanaka 1994).

A significant inversion effect (defined as the difference in response magnitude between upright and inverted stimuli) was observed in 4/6 face-selective regions tested with face and face-inverted stimuli. Notably, only one of these six face-selective regions showed a significant inversion effect for place stimuli, demonstrating specificity for the particular category of visual stimulus. Furthermore, the magnitude of the face inversion effect tended to be larger in face-selective regions located anteriorly in area TE (Fig. 7C, left; mean effect: 0.60% signal change—22% decrease from maximal response) compared with face-selective regions located posteriorly near area TEO (Fig. 7C, right; mean effect: 0.22% signal change—11% decrease from maximal response; sample too small for statistical analysis). No significant inversion effect was observed in any of the place-selective regions, using either face or place stimuli (not shown). Therefore it appears that sensitivity to stimulus inversion is not a ubiquitous feature of all category-selective regions but is instead restricted to certain categories.

**DISCUSSION**

Here we used a single experimental approach (fMRI) to investigate three current models of the organization and selec-
tivity in ventral temporal areas thought to be involved in object recognition and compared those results in human and non-human primates. fMRI correlates were found for each model that when taken in the context of previously published studies suggest that representations of complex visual stimuli in IT cortex may be organized into multiple hierarchical tiers encompassing both semantic and physical properties. Specifically, we found that macaque IT cortex, like human IT cortex, can be segregated into regions coding for animate versus inanimate stimuli, which themselves can be further subdivided into regions coding for individual semantic categories. Finally, we found that voxels selective for visually dissimilar but semantically related exemplars can be identified within category-selective regions in monkey IT cortex. In the following discussion, we describe how our results contribute to an understanding of how stimulus representations in IT cortex are organized.

**Animate versus inanimate stimulus representations in IT cortex**

The first hypothesis examined was whether humans and monkeys show distinct regions in IT cortex selective for animate versus inanimate stimuli. This hypothesis has been tested using neuroimaging techniques in humans (e.g., Beauchamp et al. 2002; Mahon et al. 2007; Mitchell et al. 2009...
additional neural analyses. Of others and are perhaps more likely to require different/
uli, animate stimuli, such as facial expressions or body surroundings. For example, compared with inanimate stim-
benefit of allowing one to rapidly categorize stimuli in one’s pathway according to this distinction has the presumed
organization, our data suggest that the relative selectivity of individual voxels for their preferred versus nonpreferred categories. The selectivity index (SI) was calculated using the following formula: $SI = \frac{R_p - \frac{1}{3}(R_{np1} + R_{np2} + R_{np3})}{[R_p + \frac{1}{3}(R_{np1} + R_{np2} + R_{np3})]}$, where $R_p$ is the response to the preferred category and $R_{np1,3}$ are the responses to the other 3 categories. All responses were first normalized to the lowest value to avoid confounds with negative signal intensities. Greater indices indicate voxels that show a larger difference in response magnitude between their preferred cate-
gory and the average of the responses to the three remaining categories (i.e., “more selective”). Face-selective regions (see Fig. 2) are indicated by the red outlines. B: average selectivity indices for the face-selective regions in near areas TE and TEO for all 4 monkeys. C: inversion effect for face (red bars) and place (blue bars) stimuli in face-selective regions in area TE and near area TEO. Significance not assessed due to inadequate sample size ($n < 5$).

Category-selective regions in IT cortex

The second hypothesis examined concerned the distribution and properties of category-selective regions in humans and monkeys. We contrasted each of four different categories against the other three to reveal regions of strong category-selectivity in both species. Consistent with previous monkey fMRI studies, two bilateral face-selective regions were found: one anteriorly in area TE and another posteriorly near area TEO (Pinsk et al. 2005; Tsao et al. 2003a, 2006). We also found body-part-selective regions located adjacent to these as described earlier (Pinsk et al. 2005). In humans, we found face-selective regions in the fusiform gyrus adjacent to body-part selective regions, consistent with previous reports (Hadjikhan and de Gelder 2003; Haxby et al. 1994; Kanwisher et al. 1997; Peelen and Downing 2005; Puce et al. 1996; Schwarzlose et al. 2005; Sergent et al. 1992). We also found segregation in monkeys for an additional category, places, which we and others have found to reliably activate voxels in/near the parahippocampal gyrus in humans (Fig. 2) (Epstein and Kanwisher 1998). In monkeys, place-selective voxels were found bordering object-selective voxels in the occipitotemporal sulcus and/or inferior temporal gyrus (Fig. 2) with an additional place-selective region in the parahippocampal gyrus. This suggests a topographic organization in IT cortex reflecting the semantic relationships between categories.

In addition to the topographic evidence for a semantic-based organization, our data suggest that the relative selectivity of voxels might also reflect this organization. Contrasting face- and place-selective voxels with semantically related categories (i.e., faces vs. body parts, places vs. objects) produced the greatest decreases in the number of selective voxels (Fig. 3). Furthermore, face- and place-selective voxels showed greater differences in response magnitude to stimuli from semantically related categories, compared with stimuli from semantically related categories (Fig. 4). This indicates that dissociating the responses in these voxels evoked by two unrelated stimuli was much easier than dissociating between the responses evoked by two related stimuli.
Qualitatively, this “categorical tuning” was sharper for humans compared with monkeys (compare Figs. 4 and 5). However, this result could reflect methodological differences rather than (or in addition to) interspecies differences. For one thing, we measured blood volume (i.e., MION) in monkeys, and blood oxygenation (i.e., BOLD) in humans. Second, the voxel size used for monkeys was smaller than that used for humans. Third, we analyzed single subject effects in monkeys, and group effects in humans. Thus although the experimental distinction was robust, it is not yet clear whether this reflects a difference between species, or technical details.

Visually distinct stimuli activate clusters within category-selective regions

The third hypothesis examined was whether category-selective regions in monkey IT cortex show evidence of a feature-based organization. This hypothesis has been investigated extensively in monkeys using traditional neurophysiological techniques as well as optical imaging techniques. We explored this hypothesis by examining the selectivity of category-selective regions for visually distinct exemplars. Our results (Fig. 6) support this model by demonstrating that a number of individual category-selective voxels show statistically significant preferences for different exemplars from the same category and voxels with similar preferences tend to cluster together.

How can such stimulus-selectivity be reconciled with the evidence for category-selective regions? Conceivably, category-selective regions would be observed at the level of fMRI because semantically related stimuli tend to share similar visual features (Rosch and Mervis 1975; Tversky and Hemenway 1984). For example, faces, which yielded the most discrete and consistent category-selective regions in our study, are very stereotypical in terms of their physical appearance. On the other hand, the object and body-part categories, which show greater variance in their appearance and included stimuli with a wide variety of shapes and features (see Supplementary Table S1), activated larger and more varied (in terms of location from subject to subject) regions (Fig. 2). The latter categories also produced selective regions with broader tuning (Fig. 4), which may further reflect the range in visual configurations/features among the nonface stimuli. Additional evidence that category-selective regions are directly linked to visual features comes from a recent fMRI study that employed continuous morphs between faces and houses (Tootell et al. 2008). More “face-like” stimuli yielded greater responses in the fusiform gyrus and weaker responses in the parahippocampal gyrus, whereas more “place-like” stimuli produced the opposite effect. Thus voxels in face- and place-selective areas showed progressively less and less activation as the stimuli ceased to include the “canonical” features of a face or a place, respectively.

It is possible that these effects might be driven by retinotopic and/or low-level differences between the stimuli, at least in part. However, the relatively weak retinotopic differences observed in IT cortex (Ito et al. 1995) makes it unlikely that retinotopy is all that is driving these effects.

Role of face-selective regions in face processing

Face-selective regions were, by far, the most discrete and consistent of all category-selective regions identified in both humans and monkeys. Face-selective regions were also the most selective relative to the nonpreferred categories. This is not surprising considering the particularly important role that faces play in our social interactions. Both species show multiple face-selective regions in IT cortex (Kanwisher et al. 1997; Kriegeskorte et al. 2007; Pinsk et al. 2005; Sergent et al. 1992; Tsao et al. 2003a, 2006). Furthermore, in the monkey, it has been shown that such face-selective “patches” are interconnected (Moeller et al. 2008). Additional face-selective regions have recently been reported in monkey prefrontal cortex (Tsao et al. 2008), suggesting a widespread face-processing network.

In monkeys, it is not yet clear whether the face-selective regions within IT cortex serve distinct roles in face perception. However, there are several intriguing points to consider. First, we found that face-selective regions located anteriorly in area TE tend to be more sharply tuned (i.e., greater differences between responses evoked by the different categories) compared with those located posteriorly near area TEO (Fig. 7B). Second, face-selective regions in area TE were almost twice as sensitive to face inversion compared with those near area TEO (Fig. 7C). However, neither region showed a place inversion effect analogous to the inversion effect for faces. Face inversion severely compromises face recognition in humans (e.g., Yin 1969) and monkeys (Parr et al. 1999; Tomonaga 1994; Vermeire and Hamilton 1998). However, attempts to identify a neural correlate in face-selective regions in humans have led to conflicting results: some labs report a decrease in activation in response to inverted faces (Passarotti et al. 2007; Rossion and Gauthier 2002; Yovel and Kanwisher 2005), while others report no change (Aguirre et al. 1999; Epstein et al. 2006; Haxby et al. 1999). In the current study, while we did not specifically test whether the monkeys show a behavioral face inversion effect, the neural effect here suggests that these two regions do process face stimuli in different ways.

Finally, studies have shown that experimental lesions of anterior IT cortex of monkeys cause greater deficits in stimulus recognition, whereas lesions of posterior IT cortex cause greater deficits in visual perception (e.g., Ettlinger et al. 1968; Iwai and Mishkin 1969; Merigan and Saunders 2004). Therefore, in the context of face processing, the posterior face-selective region near area TEO may be well suited to discriminate between faces versus non faces, whereas the anterior face-selective region in area TE may play a role in the recognition of individual faces. Similar theories concerning differential roles of face-selective regions have been suggested for human temporal cortex (e.g., Sergent et al. 1992; Haxby et al. 2000; Kriegeskorte et al. 2007; Rossion et al. 2003; Schiltz et al. 2006; Sorger et al. 2007). Functional specializations for anterior and posterior IT regions in face processing could be probed directly through neurophysiological comparisons and/or more detailed lesion experiments.

ACKNOWLEDGMENTS

The authors thank W. Vanduffel for invaluable contributions to the establishment of primate imaging at Massachusetts General Hospital. They also thank H. Deng, H. Kolster, A. van der Kouwe, A. Dale, Z. Saud, L. Ekstrom, and L. Guillory for assistance.

GRANTS

This work was supported by grants from the National Institutes of Health (R01MH-67529 and R01 EY-017081 to R.B.H. Tootell, the Athinoula A. Martinos Center for Biomedical Imaging, the National Center for Research
RESOURCES AND THE INSTITUTE OF MENTAL HEALTH INTRAMURAL RESEARCH PROGRAM TO A. H. Bell, F. Hadji-Bouziane, J. B. Frihauf, and L. G. Ungerleider.

REFERENCES


