

Neuroimaging evidence for object model verification theory: Role of prefrontal control in visual object categorization

Giorgio Ganis,^{a,b,c,*} Haline E. Schendan,^{b,d,e} and Stephen M. Kosslyn^{c,f}

^aDepartment of Radiology, Harvard Medical School, Boston, MA 02115, USA

^bMassachusetts General Hospital, Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA 02129, USA

^cDepartment of Psychology, Harvard University, Cambridge, MA 02138, USA

^dDepartment of Psychology, Tufts University, Medford, MA 02155, USA

^eDepartment of Psychology, Boston University, Boston MA 02215 USA

^fDepartment of Neurology, Massachusetts General Hospital, Boston, MA 02142, USA

Received 26 April 2006; revised 10 July 2006; accepted 2 September 2006
Available online 27 October 2006

Although the visual system rapidly categorizes objects seen under optimal viewing conditions, the categorization of objects seen under impoverished viewing conditions not only requires more time but may also depend more on top-down processing, as hypothesized by object model verification theory. Two studies, one with functional magnetic resonance imaging (fMRI) and one behavioral with the same stimuli, tested this hypothesis. fMRI data were acquired while people categorized more impoverished (MI) and less impoverished (LI) line drawings of objects. fMRI results revealed stronger activation during the MI than LI condition in brain regions involved in top-down control (inferior and medial prefrontal cortex, intraparietal sulcus), and in posterior, object-sensitive brain regions (ventral and dorsal occipitotemporal, and occipitoparietal cortex). The behavioral study indicated that taxing visuospatial working memory, a key component of top-down control processes during visual tasks, interferes more with the categorization of MI stimuli (but not LI stimuli) than does taxing verbal working memory. Together, these findings provide evidence for object model verification theory and implicate greater prefrontal cortex involvement in top-down control of posterior visual processes during the categorization of more impoverished images of objects.

© 2006 Elsevier Inc. All rights reserved.

Introduction

The visual system can rapidly categorize clearly perceived, single objects into a known class. For instance, the human brain responds differently to images of any common object vs. any face within 125 ms and to specific instances of correctly classified common objects vs. unidentified objects within 200–300 ms (Schendan et al., 1998; Schendan and Kutas, 2002). This remarkable speed of processing has led many theorists to focus

on fast bottom-up processes during object categorization, thought to be implemented in neural structures of the ventral visual pathway (Biederman, 1987; Grill-Spector and Malach, 2004; Perrett and Oram, 1993; Poggio and Edelman, 1990; Riesenhuber and Poggio, 2002; Rousselet et al., 2003; Wallis and Rolls, 1997). Nevertheless, in ordinary environments, objects are often not clearly perceived because of shadows, partial occlusion by other objects, poor lighting, and so forth. In these situations, object categorization is markedly slower (Schendan and Kutas, 2002). In two studies, we examined categorization of common objects at the basic-level (e.g., dog, cat, car, or chair), which is the main focus of most theories of visual object categorization (e.g., Biederman, 1987), as opposed to a broader superordinate level of categorization (e.g., mammals, vehicles, or furniture) or a more specific subordinate level (e.g., collie dog, Siamese cat, Prius car, or Windsor chair) or identification of a particular unique item (e.g., an individual person, my dog or cat, your car, or grandfather's chair) (Rosch et al., 1976; Smith and Medin, 1981).

To explain how objects can be categorized even when the image is impoverished, some theorists hypothesize that top-down control processes augment bottom-up processes. Top-down control processes direct a sequence of operations in other brain regions, such as during the engagement of voluntary attention or voluntary retrieval of stored information (Ganis and Kosslyn, in press; Miller and Cohen, 2001; Miller and D'Esposito, 2005). According to these theories, top-down control processes play a crucial role after an initial bottom-up pass through the visual system. If the input image is impoverished, this first pass may provide only weak candidate "object models" (i.e., structural representations stored in long-term memory) for the match with the input (Kosslyn et al., 1994; Lowe, 1985). Some theorists have proposed that top-down processes drive *object model verification* (Lowe, 2000), a process that determines which one of the object models stored in long-term memory best accounts for the visual input. This verification process is engaged during the categorization of any image. However, it only runs to completion when bottom-up processes

* Corresponding author. Martinos Center, Building 149, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA. Fax: +1 617 496 3122.

E-mail address: ganis@nmr.mgh.harvard.edu (G. Ganis).

Available online on ScienceDirect (www.sciencedirect.com).

produce partial or weak matches between the input and stored object models, which is more often the case with more impoverished images.

In our view, top-down processes are recruited to evaluate stored models (Kosslyn et al., 1994). To date, researchers have reported only sparse neurocognitive evidence for the role of top-down control processes in object categorization (for recent reviews see, Miller and D'Esposito, 2005; Ganis and Kosslyn, *in press*). This is surprising because the role of top-down processing is a core issue that must be addressed to develop a comprehensive theory of visual object categorization.

In the present article, we report two studies, one using neuroimaging and one using behavioral interference methods, to evaluate this class of accounts of how objects are categorized when seen under impoverished viewing conditions. To this end, we used line drawings of objects that were impoverished by removing blocks of pixels, henceforth referred to as *impoverished objects* (note, that the objects themselves were not impoverished, but rather it was the pictures of the objects that were impoverished—but the present notation is more concise than detailing the stimuli every time). In the first study, fMRI was used to test specific predictions about the categorization of more impoverished (MI) vs. less impoverished (LI) objects. Top-down control processes are thought to be implemented in a prefrontal and posterior parietal network (e.g., Corbetta et al., 1993; Hopfinger et al., 2000; Kastner and Ungerleider, 2000; Kosslyn, 1994; Miller and D'Esposito, 2005; Wager et al., 2004; Wager and Smith, 2003). Thus, one prediction is that categorizing MI objects should engage frontoparietal brain networks involved in top-down control more strongly than does categorizing LI objects. We propose that the specific processes that are engaged more by MI than LI objects include the following cognitive control processes (cf. Kosslyn et al., 1994, 1995): (a) *Retrieving Distinctive Perceptual Attributes*, which involves activating perceptual knowledge stored in long-term memory associated with a candidate object model, especially those perceptual attributes that are most distinctive for that object model (e.g., a section of a wing, if a candidate object is an airplane); (b) *Attribute Working Memory (WM)*, which maintains retrieved knowledge about distinctive visual attributes of the candidate object model and compares it with the visual input; (c) *Covert Attention Shifting*, which shifts attention to locations where such distinctive attributes are likely to be found; and (d) *Attribute Biasing*, which consists of biasing representations of the distinctive attributes to facilitate detecting and encoding them. Information obtained following top-down processing may reveal that expected attributes are indeed present at the expected locations. This would constitute evidence that the candidate object is the one being perceived.

We thus specifically predicted stronger activation in regions of prefrontal cortex (PFC) and inferior and superior parietal regions that are involved in knowledge retrieval, WM and attentional processes (e.g., Corbetta et al., 1993; de Fockert et al., 2001; Hopfinger et al., 2000; Kastner and Ungerleider, 2000; Kosslyn et al., 1995; Petrides, 2005; Smith and Jonides, 1999; Oliver and Thompson-Schill, 2003; Wager et al., 2004; Wager and Smith, 2003). In contrast, bottom-up processing accounts are essentially agnostic with regard to areas outside of the ventral stream; for instance, although PFC engagement is thought to be task dependent, any visual stimulus that is categorized (be it LI or MI) is assumed to be processed similarly at the PFC stage. Thus, these accounts would not predict differences in activation in

frontoparietal networks between successfully categorized MI and LI objects (Riesenhuber and Poggio, 2002).

We also predicted that categorizing MI objects, relative to LI, should more strongly activate regions of occipital, ventral temporal and inferior posterior parietal cortex that play key roles in representing or processing information about visually perceived objects (Hasson et al., 2003), hereafter referred to as *object-sensitive regions*. This is because the top-down control processes recruited during object model verification (i.e., retrieving distinctive attributes, holding them in working memory, shifting attention, and biasing relevant features) work in concert with these posterior regions. Therefore, for MI objects (relative to LI ones), top-down processing should recruit neuronal populations in posterior object-sensitive regions until categorization is achieved, which would thereby result in more overall activation of these regions for MI than LI objects.

In contrast, bottom-up processing accounts (Riesenhuber and Poggio, 2002) would predict no difference or the opposite effect (i.e., MI activation should be weaker than LI activation) because MI objects contain fewer visual features than LI objects. Bottom-up processing accounts postulate categorization via populations of feature detectors that are organized hierarchically along the ventral stream. On average, impoverished images with fewer visual features should result in fewer units being activated, and each unit may be activated more weakly. This would predict that categorizing MI objects, relative to LI, should activate these regions less strongly.

In addition, we used two independent localizer tasks. One localizer task defined object-sensitive regions, which allowed us to use our experimental results to test our hypotheses in these regions. A second localizer allowed us to remove the contributions of eye movement regions, adjacent to the prefrontal and parietal areas of interest, from the experimental analyses.

In the second study, we augmented the neuroimaging evidence, which is inherently correlational, with causal evidence. We used a behavioral interference paradigm to investigate an additional prediction: If the categorization of impoverished objects relies upon top-down control processes, then a concurrent task that engages some of these same processes should interfere with the categorization of MI objects—and should do so more than it interferes with the categorization of LI objects. The design and predictions of this study rest on the following assumptions: (a) Top-down control relies on WM processes (Smith and Jonides, 1999). (b) WM processes in dorsal vs. ventral lateral PFC, respectively, may be distinguished according to the content they operate upon, such as spatial vs. nonspatial (Romanski, 2004) or relations vs. single items (Ranganath and D'Esposito, 2005), or according to the processes performed, such as monitoring and manipulation vs. maintenance (Petrides, 2005). (c) WM and attentional processes share common neural resources (Awh and Jonides, 2001; de Fockert et al., 2001; Wager et al., 2004). (d) During top-down model verification, WM and attentional processes need to maintain, keep track of, and manipulate visual representations of distinctive attributes of the candidate objects and their probable spatial locations (Kosslyn et al., 1994, 1997). To our knowledge, this is the first study explicitly relating WM processes to processes involved in the categorization of impoverished objects.

Finally, we note that the use of top-down processing is not all-or-none, but rather falls along a continuum. Thus, our manipulation should vary only the degree to which such processing is used in the task.

Materials and methods

Experiment 1

Subjects

Twenty-one Harvard University undergraduates (12 females; mean age=20 years), volunteered to participate in the study for pay. All had normal or corrected-to-normal vision, no history of neurological disease, and were right-handed. All subjects gave written informed consent for the study according to the protocols approved by Harvard University and Massachusetts General Hospital Institutional Review Boards. We analyzed data from 17 subjects; data from 4 subjects were not analyzed because of uncorrectable motion artifacts (2 subjects) or because they did not complete the study; demographics of these 4 subjects were comparable to those of the entire group.

Stimuli

Line drawings of 200 objects from a standardized picture set (Snodgrass and Vanderwart, 1980) were impoverished by removing random blocks of pixels (Fig. 1), a method referred to as *fragmentation*. This method of impoverishing pictures is atheoretical; no assumptions are made about whether certain parts are more important than others to carry out object categorization. In the following, *fragmentation level* per se refers to the proportion of deleted pixel blocks in the image, regardless of how perceptual properties of each picture affect categorization. Eight levels of fragmentation (from 1 to 8, with 8 corresponding to the most fragmented version) were available for each picture, making up a *fragmentation series*. The formula that expresses the proportion of deleted pixel blocks (PD) as a function of fragmentation level is (modified from Snodgrass and Corwin, 1988):

$$PD(\text{level}) = 1 - 0.7^{(\text{level}-1)}$$

Using this formula, for instance, the proportion of deleted blocks at levels 4 and 6 is 66% and 83%, respectively. For 150 of the objects, the fragmentation series were from the Snodgrass and Corwin (1988) set. We used the same software algorithm originally used to produce that set (Snodgrass et al., 1987) to generate the fragmentation series for the remaining 50 objects.

We then tested a separate group of 16 subjects to obtain normative data used to select 128 objects that included two fragmentation levels (high vs. low) such that: (a) each picture was categorized correctly (i.e., defined by the acceptable names given in Snodgrass and Vanderwart, 1980) by at least 75% of people at the two levels; (b) for each picture, the RT for the low fragmentation level was numerically lower than for the high fragmentation level. For different objects, different fragmentation levels were used (which is a factor we later considered in our analysis).

As control stimuli, we used 64 pseudo-objects, that is, stimuli that could be real objects (in the sense that they were not impossible objects that cannot exist in the Euclidean three-dimensional world) but do not correspond to any known object and so are unidentifiable. The pseudo-objects were from a prior study (Schendan et al., 1998) and had been created by rearranging the parts of our object pictures. Objects and their corresponding pseudo-objects were fragmented using the same procedures and to the same degree, thereby equating this aspect of perceptual similarity (Fig. 1). Stimuli subtended 6 by 6° of visual angle, on average with a visual contrast of approximately 30% (dark pixels against a brighter background).

Procedure

The tasks were administered on a MacIntosh G3 Powerbook computer using Psyscope software (Macwhinney et al., 1997). Stimuli were projected via a magnetically shielded LCD video projector onto a translucent screen placed behind the head of each subject. A front-surface mirror mounted on the head coil allowed the subject to view the screen. Prior to the MRI session, general health history and Edinburgh Handedness (Oldfield, 1971) questionnaires were administered.

Before the magnetic resonance imaging (MRI) session, subjects read instructions on the computer screen and paraphrased them aloud. We corrected any misconceptions at this time. We then administered 10 practice trials. Subjects pressed one key if they could categorize the visual stimulus and another key if they could not. They were instructed to respond as quickly as possible without sacrificing accuracy. Furthermore, they were instructed to fixate their gaze on the center of the screen at all times, but eye movements were not otherwise controlled.

The MRI session consisted of 8 functional scans. During the first 4 scans, we presented the pictures of objects and pseudo-objects for

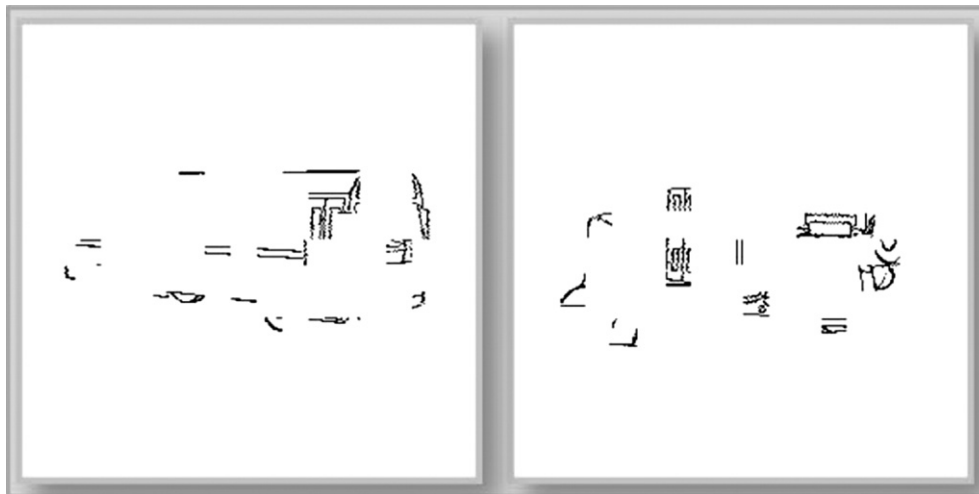


Fig. 1. Sample stimuli in Experiment 1; at the left is an impoverished image of a bus; at the right is an impoverished image of its pseudo-object version.

2.2 s in a fast event-related paradigm. The average stimulus onset asynchrony was 6.8 s, varying between 4 and 16 s from trial to trial, according to a random sequence optimized for deconvolution using program ‘optseq2’ (Dale, 1999). The order of conditions was randomized. Note, no mention was made of the existence of the pseudo-objects: from the standpoint of subjects, the pseudo-objects were simply objects that they could not categorize. A debriefing questionnaire at the end of the study revealed that none of the subjects realized some stimuli were pseudo-objects.

For the next 2 scans, we localized object-sensitive brain regions by alternating grayscale pictures of objects and textures in a blocked design (6 blocks, each lasting 60 s); the textures were created using the standard method of scrambling the phase information in the Fourier representation of the corresponding objects (Malach et al., 1995). For the last 2 scans, we localized regions involved in the generation of saccadic eye movements to eliminate from analyses any regions related to saccades per se. In the eye movement condition, a dot appeared at random locations on the circumference of an invisible circle (with a radius equal to 3° of visual angle) at a rate of 1 Hz. The area of the circle was the average area of the objects used in the object categorization task, which thus induced, as closely as possible, saccades with the same amplitude as those during that task. The control condition required fixating the same dot when it was stationary at the center of the screen, and the two conditions alternated every 30 s, with each cycle repeated 6 times.

MRI parameters

Using a 3 T Siemens Allegra scanner with whole head coil, we collected full-volume structural images using an MPRAGE sequence, before and after the functional scans (128, 1.33 mm thick sagittal slices, TR=2530 ms, TE=3.25 ms, TI=1100 ms, FOV=25.6 cm, flip angle=7°, 256×256 matrix). Functional scans assessed blood oxygenation changes, using a T2*-sensitive sequence (gradient echo, TR=2000 ms, TE=30 ms, TI=30 ms, FOV=20 cm, flip angle=90°, 64×64 matrix, voxel size=3.125×3.125×5 mm, 21 oblique slices).

Analyses

Data were pre-processed and analyzed with *AFNI* (Cox, 1996) using the following steps: (1) slice timing correction; (2) motion correction; (3) spatial smoothing with a Gaussian filter (full-width half-maximum=4 mm); (4) amplitude normalization, by scaling each time series to a mean of 100 and calculating the percent signal change about this mean; (5) spatial normalization to the MNI305 template (Collins et al., 1994); and (6) spatial resampling to a 3×3×3 mm grid. For the hemodynamic response function, we used a finite impulse response (FIR) model and estimated the fMRI response at each time point independently using multiple linear regression. The multiple regression model included an offset (i.e., mean), a linear, and a quadratic trend regressor for each scan. In addition, for each condition, there was one regressor for each time step in the window of interest (from -2 s before trial onset to 16 s after trial onset, for a total of 10 regressors).

We were primarily interested in data from correct trials, defined as those in which subjects appropriately categorized an object when an object was presented or responded they could not categorize the image when a pseudo-object was presented. Within each block, MI vs. LI objects were defined as those with mean RTs above vs. below the median RT, respectively, for correctly categorized objects. MI and LI pseudo-objects were defined

analogously. Note that, for the main analysis comparing activation to MI and LI objects, we used only 98 of the 128 object stimuli (those corresponding to fragmentation levels 3, 4, and 5, with the median RTs defining MI and LI conditions computed for these levels) to equate average visual complexity for the MI and LI sets, according to the normative data from Snodgrass and Corwin (1988), as described in Results. For correct trials, we used 4 sets of regressors: 2 for the MI and LI objects and 2 for the MI and LI pseudo-objects. For incorrect trials, we used 2 sets of regressors, one for objects that were not categorized and one for pseudo-objects that were incorrectly reported to be objects. The 30 trials (out of 128) from fragmentation levels 1, 2, and 6 (not included in the main analysis) were modeled with a separate set of regressors. Finally, we used another set of regressors for occasional trials during which the subject did not provide a response (no-response trials). We note that, although these last three groups of trials were not included in the main analysis, they contributed to the observed fMRI timeseries and so had to be included explicitly in the multiple regression model. We also note that we repeated all analyses on the entire set of 128 objects (encompassing all levels), and the pattern of results was unchanged. Furthermore, the same analyses repeated on MI and LI objects defined in terms of fragmentation level (4, 5, and 6 for MI and 1, 2 and 3 for LI) produced comparable results.

Maps of percent signal change for each subject and condition were obtained using the normalized regression coefficient at 6 s post-onset, at the peak of the hemodynamic response. The principal whole-brain analysis was a repeated-measures ANOVA on the maps for the MI and LI objects (5 voxel extent threshold, all significant at $p < .001$). These parameters provide a good compromise between sensitivity and protection against false positives (Xiong et al., 1995).

For the object-sensitive regions and eye-movement localizer tasks, we performed the same pre-processing steps (1 through 6). The multiple regression model included an offset (i.e., mean), a linear, and a quadratic trend coefficient for each scan. We created the regressor for these blocked designs by convolving the timecourse of the paradigm with an assumed hemodynamic response.

We use the term “activation” to refer to positive activations and “deactivation” to refer to negative activations.

Experiment 2

Subjects

Sixteen Harvard University undergraduates, not tested in Experiment 1, volunteered to participate in this study for pay (8 females, mean age=20 years). All gave written informed consent for the study according to the protocols approved by the Harvard University Institutional Review Board.

Stimuli

Of the 200 objects created for Experiment 1, the normative data described for Experiment 1 stimuli were used to select 100 objects that included two levels of fragmentation (high vs. low) that fulfilled both criteria (a) and (b) from Experiment 1 for being impoverished, and a third criterion (c) the MI version was exactly two levels more fragmented than the corresponding LI version, to keep the images as comparable as possible.

The WM materials were digits (1–9) and 9 non-verbalizable shapes. These shapes were selected from a set of 100 polygons

created by connecting 4 to 8 dots randomly placed on the vertices of a 10×10 grid. Eight pilot subjects were shown the entire set and asked to provide a list of names that could be used to refer to the shape (with a rating of how well the names fit the object, from 1, worst fit, to 5, best fit; 0 was used if no name was provided). For each shape, we calculated the average nameability, after adding the ratings for all names provided by each subject. The final set was selected by taking the 9 shapes with the lowest total nameability (.3, on average). We arranged the stimuli into strings, where each digit string had 5 distinct numbers, and each shape string had 2 distinct shapes. We determined the string lengths in a pilot study with a separate group of 8 subjects to equate difficulty for the two WM tasks. For this normative study, we used intact versions of the pictures. We equated task difficulty by ensuring that the WM task accuracy (as measured by the WM score, see Analyses section below) was equivalent in the two tasks (.92 for Digit WM compared to .91 for Shape WM, $t(7) = .58, p > .5$).

Procedure

We presented Shape WM and Digit WM trials in separate blocks. Block order, assignment of pictures to MI and LI conditions and Shape and Digit tasks were counterbalanced across subjects. Each trial began with a visual string of digits or shapes (Fig. 7a). Subjects were told to keep the string in mind throughout the trial. After studying the string for 5 s, 4 objects were presented in sequence, each one for 2 s, followed by a 1 s fixation cross. Subjects spoke the name of each object into a microphone as soon as they categorized it (or said “don't know,” if they could not do so). Voice onset response times (referred to as RTs) were recorded by the computer, and names were recorded by the investigator. Following the sequence of 4 pictures, subjects were to press the sequence of digits on a numerical keypad or the sequence of shapes (little pictures of the shapes were attached to the keys) on a second keypad. We ensured that subjects did not look at the keypad while studying the strings at the beginning of each trial to discourage explicit strategies based on the association between items and keypad locations. We repeated the experiment, with the same stimuli and order, to assess the possibility of ceiling effects on RTs.

Analyses

First, to quantify performance on each of the WM tasks, we assigned a normalized score to each WM trial. This was necessary to ensure that only trials were included during which subjects had actually performed the WM task. We assigned one point for each item reported correctly and an extra point for each item reported at its correct sequential position within the string, and divided the total by the maximum number of points (and thus the scores ranged from 0 to 1, perfect performance). The RTs and error rates (ERs; when subjects could not categorize the object) for trials where the WM score was .25 or higher (96% of the trials) were submitted to a 3-way repeated measures ANOVA, with factors of WM type (shape vs. digit), fragmentation level (MI vs. LI), and session (first vs. second). We checked for possible speed-accuracy trade-offs by correlating RTs and ERs across subjects.

Results

We here report the results from both Experiment 1 (with fMRI) and Experiment 2 (which used a behavioral paradigm to test the causal role of top-down processing during categorization).

Experiment 1

We used a median split of categorization times to operationally define the degree to which the image was *impoverished*, that is, MI vs. LI stimuli (Fig. 1). To demonstrate that the effects of image impoverishment on brain activation were not due to mere task difficulty (as indexed by RTs), unbeknownst to subjects, a set of pseudo-objects was included, formed by rearranging parts of the objects to produce a shape that does not embody any known object category.

Performance

On average, subjects correctly categorized 80% of the real objects and failed to categorize 91.7% of the pseudo-objects. The average fragmentation levels (defined in terms of the proportion of deleted pixel blocks in the image on a scale from 1, least fragmented, to 8, most fragmented, see Material and Methods) for the MI and LI objects were 4 and 3.5, $F(1,16) = 201.2, p < .0001$, respectively, which is consistent with the fact that subjects categorized objects more slowly when they were more fragmented than when they were less fragmented. Indeed, as expected, RTs and fragmentation levels were highly correlated, $r = -.65, p < .0001$. In addition, we did not find any systematic differences between MI vs. LI objects on various dimensions (Snodgrass and Vanderwart, 1980): name agreement (86% vs. 87%), image agreement (3.7 vs. 3.6), familiarity (3.4 vs. 3.2), visual complexity (2.9 vs. 2.9), name frequency (18 vs. 15), and age of acquisition (2.6 vs. 2.8). We found the opposite pattern for the pseudo-objects: Average fragmentation levels for the MI and LI pseudo-objects were 4.7 and 5.2, respectively, which show that subjects realized that they could not categorize a pseudo-object faster when it was more fragmented than when it was less fragmented, $F(1,16) = 175.2, p < .0001$.

Fig. 2a illustrates RTs for correctly categorized objects (MI and LI) and uncategorized pseudo-objects (MI and LI). A 2×2 repeated measures ANOVA with factors of stimulus type (objects vs. pseudo-objects) and impoverishment (MI vs. LI) showed that subjects categorized MI stimuli slower than LI stimuli, $F(1,16) = 184.4, p < .0001$. They also required more time for pseudo-objects than real objects, $F(1,16) = 20.3, p < .001$. The interaction of type by impoverishment was not significant, $F(1,16) = 2.5, p > .1$. The impoverishment effect was thus similar for real objects and for pseudo-objects.

Finally, the behavioral results were not contaminated by speed/accuracy trade-offs: RTs for categorized objects and ERs for objects were not correlated across subjects (all $r_s < .1, p > .5$).

Brain activation maps and time courses for the object categorization task

To test top-down control theories of object categorization, the crucial comparison is between the activation elicited when subjects successfully categorized MI vs. LI objects. Consistent with the predictions of these theories, categorizing MI objects elicited stronger activation (bilaterally, unless otherwise specified) than did categorizing LI objects in several regions implicated in top-down control and object processing (Table 1). First, activation was stronger in numerous frontal and parietal areas when subjects categorized MI than when they categorized LI objects, specifically the ventrolateral prefrontal cortex (VLPFC), anterior cingulate gyrus, extending into supplementary motor area, right precentral sulcus/gyrus, postero-lateral portions of left inferior frontal cortex,

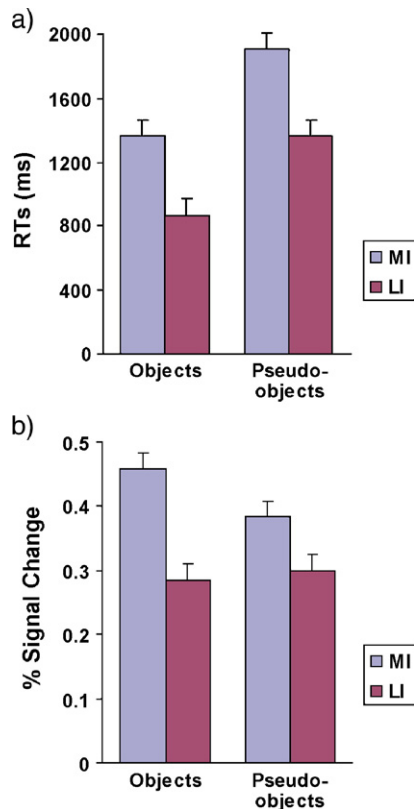


Fig. 2. (a) Mean RTs to MI and LI objects and pseudo-objects (Experiment 1); (b) Percent signal change (relative to the mean of the corresponding timeseries, see text for details) to the same stimuli, averaged across regions of interest defined by the contrast of MI and LI objects in the main analysis. The dissociation between the RT pattern and the brain activation pattern for objects versus pseudo-objects indicates that RTs alone are not responsible for the object impoverishment effects. Error bars represent the standard error across subjects.

and intraparietal sulcus, encompassing superior and inferior parietal lobule regions (Figs. 3 and 4). One region in the right angular gyrus showed instead deactivation to all stimuli, with MI objects eliciting more deactivation than LI objects (Fig. 4). In addition, activation was stronger for MI than LI objects in the fusiform gyrus, extending into inferior temporal sulcus, left lateral occipital cortex, and left middle occipital gyrus (Fig. 4).

Notice in Figs. 3–6 that the hemodynamic response for both MI and LI objects peaked at 6 s post-stimulus, but both (MI and LI) responses relative to baseline were already significant by 4 s post-stimulus in all regions, including posterior ones, $F(1,16)=57.7$, $p<.0001$. However, at this earlier time, the posterior regions did not yet show any activation difference between MI and LI objects, including in the fusiform gyrus, lateral occipital cortex, and middle occipital gyrus (neither the main effect of impoverishment, nor interactions with impoverishment were significant, all $ps>.5$). By contrast, in frontal top-down control areas, differences in activation evoked by MI and LI objects were already significant at this time, as shown by an ANOVA including the VLPFC, anterior cingulate/SMA, and left inferior frontal cortex, $F(1,16)=9.45$, $p<.01$.

Objects vs. pseudo-objects

To determine whether the effect of impoverishment is due to object categorization processes and not simply due to generic task

difficulty, as may be suggested by the RTs, we conducted an analysis of activation in all regions of interest that were found in the comparison between MI and LI conditions for the real objects (excluding the angular gyrus because it was the only region in which deactivation occurred). The ANOVA had 3 factors: stimulus type (objects vs. pseudo-objects), impoverishment (MI vs. LI), and region. If the activation pattern described earlier was due entirely to generic task difficulty, activation should be stronger overall for pseudo-objects than objects because the subjects required more time overall to respond to pseudo-objects than objects (Fig. 2a). Further, activation corresponding to the *impoverishment effect* (MI minus LI) should not differ between pseudo-objects and objects, because RTs had no such interaction (Fig. 2a). Note that although these regions of interest were defined by using the comparison between MI and LI conditions for real objects, here we focus on how these regions were activated when processing the pseudo-objects. The activation when objects were being processed served as an independent baseline against which to compare activation when pseudo-objects were being processed (which were not used to define these regions of interest).

The results were clear (Fig. 2b and Fig. 5): activation was higher when subjects processed MI stimuli than LI stimuli, $F(1,16)=38.72$, $p<.0001$, and different levels of activation were found in different regions, $F(10,160)=18.79$, $p<.0001$. Crucially, there was no sign of stronger activation for pseudo-objects than for objects. In fact, on average, activation elicited by objects was stronger than that elicited by pseudo-objects (.37 vs. .34, respectively), the opposite of the RT pattern, albeit only numerically (main effect of Stimulus type, $F[1,16]=1.27$, $p>.1$). Furthermore, categorizing objects elicited stronger activation in some regions than did attempting to categorize pseudo-objects; stimulus type interacted with region, $F(10,160)=3.8$, $p<.0001$. Follow-up analyses revealed that these regions were the fusiform gyrus/ITS and left intraparietal sulcus. Critically for the issue at hand, the impoverishment effect was larger for objects than for pseudo-objects (Fig. 2b): stimulus type interacted with impoverishment, $F(1,16)=7.93$, $p<.05$. Because we did not find this interaction with RTs, fMRI activation differences cannot be explained by appealing to differences in RT per se. No other effects or interactions were significant, $p>.1$ in all cases.

Object-sensitive regions

We also defined which regions were activated more strongly when subjects saw images of objects than when they viewed phase scrambled versions of those objects (which look like textures); this comparison identified a large section of occipital, posterior temporal, and posterior parietal cortex. These regions appeared as a continuous swath of cortex within each hemisphere, so we parceled them out using the Talairach coordinate ranges from the previous literature (Grill-Spector et al., 2000; Hasson et al., 2003) and referred to them using the corresponding nomenclature: VOT, DOT-ITS, DOT-LOS (corresponding to the classic LOC), DF1, and DF2 (Fig. 6). Within these regions, we defined ROIs by identifying voxels that were significant in the object-sensitive localizer and in the main analysis comparing activation for MI and LI objects. We used a more liberal statistical threshold ($p<.005$) than in the whole-brain analysis because the search for activated voxels was carried out within small, pre-defined regions (Xiong et al., 1995). Virtually all voxels in posterior regions that were significant in the main analysis fell within the object-sensitive regions that were identified by our localizer task.

Next, we performed an analysis using these ROIs to test our hypothesis that, for successfully categorized objects, categorizing

Table 1
Brain areas showing more activation or more deactivation for MI than LI objects

| Region | Volume | Talairach coordinates (center of mass) | | | Talairach coordinates (range) | | | | | |
|--------------------------|--------|---|-----|-----|-------------------------------|-------|-------|-------|-------|-------|
| | | x | y | z | Min x | Max x | Min y | Max y | Min z | Max z |
| <i>Activation</i> | | | | | | | | | | |
| Ventrolateral PFC | 2673 | 35 | 22 | 3 | 29 | 44 | 17 | 32 | -7 | 21 |
| Ventrolateral PFC | 2268 | -35 | 22 | -1 | -41 | -26 | 14 | 32 | -7 | 12 |
| ACG/SMA | 5238 | -3 | 15 | 39 | -14 | 11 | 2 | 29 | 21 | 54 |
| Precentral sulcus/gyrus | 540 | 34 | -7 | 47 | 29 | 44 | -11 | -5 | 45 | 51 |
| Inferior frontal cortex | 945 | -40 | 5 | 25 | -47 | -35 | 2 | 11 | 18 | 33 |
| Intraparietal sulcus | 864 | 23 | -70 | 36 | 20 | 29 | -77 | -62 | 24 | 48 |
| Intraparietal sulcus | 2376 | -25 | -61 | 43 | -38 | -20 | -74 | -47 | 30 | 51 |
| Fusiform gyrus/ITS | 4428 | 32 | -55 | -15 | 20 | 47 | -74 | -38 | -20 | -4 |
| Fusiform gyrus/ITS | 1080 | -38 | -57 | -16 | -50 | -32 | -71 | -44 | -18 | -7 |
| Lateral occipital cortex | 243 | -38 | -87 | 3 | -41 | -35 | -89 | -83 | -1 | 6 |
| Middle occipital gyrus | 2214 | -29 | -79 | 19 | -35 | -26 | -86 | -71 | 6 | 30 |
| <i>Deactivation</i> | | | | | | | | | | |
| Angular gyrus | 1674 | 41 | -65 | 34 | 35 | 47 | -77 | -59 | 27 | 42 |

Note. Abbreviations: PFC, prefrontal cortex; ACG, anterior cingulate gyrus; SMA, supplementary motor area; ITS, inferior temporal sulcus; Coordinates: x, left/right; y, posterior/anterior; z, inferior/superior. Voxel clusters were considered significant if they were composed of at least 5 voxels, all significant at $p < .001$.

MI objects would evoke stronger activation in posterior object-sensitive regions than would categorizing LI objects. At the same time, we also tested the opposite prediction made by bottom-up accounts, namely that there should be greater activation in posterior object-sensitive regions when subjects categorized LI than when they categorized MI objects. Finally, this analysis allowed us to assess whether the response to stimulus impoverishment differed among object-sensitive regions. Results from the repeated measures ANOVA on the 5 region pairs with factors of hemisphere (left vs. right), region (VOT, DOT-ITS, DOT-LOS, DF1, and DF2), stimulus type (object vs. pseudo-object), and impoverishment (MI vs. LI) revealed that the overall strength of activation was different among the 5 bilateral pairs of object-sensitive regions, as witnessed by a main effect of region, $F(4,64)=9.56$, $p < .0001$. Furthermore, categorizing the MI stimuli elicited stronger activation than did categorizing the LI stimuli, as shown in a main effect of impoverishment, $F(1,16)=24.59$, $p < .0001$. It is important to note that, as in the main analysis, impoverishment had a larger effect on objects than pseudo-objects, as shown by the interaction between impoverishment and stimulus type, $F(1,16)=5.43$, $p < .05$. We also found an interaction of region by hemisphere, $F(4,64)=4.26$, $p < .005$, which reflected the fact that activation was stronger in the left hemisphere for DOT-LOS, $F(1,16)=5.14$, $p < .05$, but stronger in the right hemisphere for all other regions. Finally, an interaction between stimulus type and region, $F(4,64)=3.52$, $p < .05$, arose because activation in response to categorizing objects was reliably stronger than that elicited when subjects tried to categorize pseudo-objects in regions VOT, DOT-ITS, and DF2, but not in DOT-LOS and DF1. The fact that we did not find an interaction between impoverishment and region is consistent with the idea that the critical impoverishment effect was uniform across object-sensitive regions.

Eye movement regions

Comparing the two conditions in the eye movement localizer scan revealed activation in previously known eye movement regions (Beauchamp et al., 2001) of the precentral sulcus, extending into

precentral gyrus, left posterior intraparietal sulcus, extending dorsally into the superior parietal lobule, lateral occipital sulcus, extending into middle occipital gyrus, and large parts of medial occipital cortex, extending dorsally into the parieto-occipital sulcus and inferiorly to ventral occipital cortex (Table 2). The precentral cortex activations correspond to the locations of the frontal eye fields (Beauchamp et al., 2001).

To assess whether any differences between MI vs. LI object activation could reflect differential eye movements, we defined voxels in the MI-LI object maps that overlapped with those found in the eye movement localizer. Almost no voxels activated during the eye movement task (compared to the fixation task) overlapped with those in the MI-LI maps. We found overlap only in 3 small clusters, which encompassed a total of 6 voxels: these voxels were located in the right precentral sulcus (2 voxels), left intraparietal sulcus (3 voxels), and left middle occipital gyrus (1 voxel). When these voxels were excluded from the analyses that led to the results summarized in Table 1, the results were unchanged.

Experiment 2

The fMRI results provided correlational evidence that top-down processes play a larger role when one categorizes MI objects than when one categorizes LI objects. To interpret these results, we conducted a second experiment that produced not correlational evidence, but rather evidence of causal relations. Specifically, we designed this behavioral study to demonstrate that interference with top-down control processes impairs performance when subjects must visually categorize MI objects more than when they must categorize LI objects.

Based on our assumptions relating top-down control, attention, and WM processes in lateral PFC (see Introduction), we devised an interference paradigm in which subjects categorized MI and LI objects while performing one of two types of WM tasks. One task was designed to recruit WM processes involved in object model verification: The Shape WM task used strings of 2 novel, non-verbalizable visual shapes. The other task was designed to rely minimally, if at all, on WM processes involved in model verification:

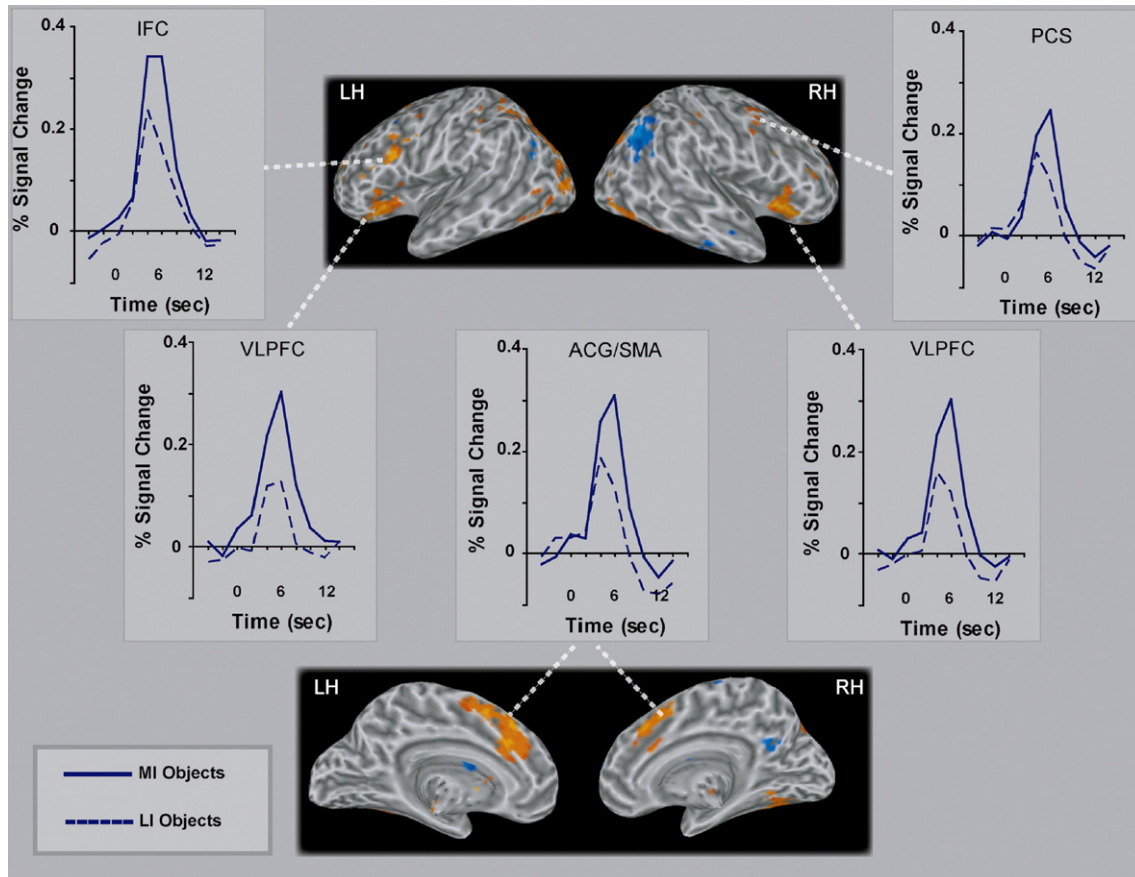


Fig. 3. Activation maps (6 s post-stimulus onset) and time courses of activation (percent signal change is relative to the mean of the corresponding time series, see text for details) showing frontal regions significantly active in contrasts of MI versus LI objects (IFC=inferior frontal cortex; PCS=precentral sulcus; VLPFC=ventrolateral prefrontal cortex; ACG/SMA=anterior cingulate gyrus/supplementary motor area; LH=left hemisphere; RH=right hemisphere). Activations are displayed on an average inflated brain [created with FreeSurfer (Dale et al., 1999; Fischl et al., 1999) from 27 anatomical scans obtained from MNI and UCLA (Holmes et al., 1998)]. Note, orange and blue regions indicate activations and deactivations, respectively.

The Digit WM task was a verbal WM task that required subjects to keep in mind strings of 5 digits. These two tasks were equated in difficulty (see Materials and methods for details). We hypothesized that the Shape WM task would interfere more than the Digit WM task with categorization of MI images, and such interference would not be as strong with categorization of LI images.

Analyses were limited to object stimuli that were categorized correctly (92%) and trials with a WM score higher than .25 (96%, see Materials and methods). Fig. 7b shows mean RTs. Most important, the Shape WM task interfered more with the categorization of MI objects than did the Digit WM task (1159 ms vs. 1103 ms, respectively), and this effect was absent for LI objects (958 ms vs. 956 ms, respectively), as witnessed by the interaction of WM task by impoverishment, $F(1,15)=5.8, p<.05$. Subjects also categorized MI objects slower than LI objects (1131 ms vs. 957 ms, respectively), $F(1,15)=70.7, p<.0001$. Finally, subjects were faster the second time they saw the stimuli (mean RTs of 1116 ms vs. 971 for the first and second presentations, respectively), $F(1,15)=31.1, p<.001$. No other effects or interactions were significant, all $p>.1$. Note that the lack of a difference between RTs for the two WM tasks for the LI objects confirms that the two WM conditions were equally difficult; moreover, these results are unlikely to reflect a ceiling effect because repeating the experiment further reduced the RTs, but the interaction of interest persisted nevertheless.

The ERs showed a similar pattern, being higher for MI than LI objects (13.9% vs 2.4%, respectively, $F(1,15)=64.1, p<.0001$), and for the first than the second block (9.6% vs 6.7%, respectively, $F[1,15]=27.9, p<.0001$). Furthermore, the impoverishment effect was smaller the second time subjects received the stimuli; block by impoverishment was significant, $F(1,15)=46.3, p<.0001$. In addition, the results were not due to speed-accuracy trade-offs because all RTs and ERs were not correlated across subjects (all $ps>.1$).

Finally, all average WM scores were above .80 and were affected neither by repetition nor by WM task (all $ps>.1$), showing that subjects performed the concurrent WM tasks well, and the difficulty of these tasks was constant across conditions.

Discussion

Our findings confirm three key predictions from theories that posit a role for top-down control during visual object categorization. First, the fMRI results provide direct evidence that the categorization of impoverished images of objects engages frontoparietal brain regions implicated in top-down control for long-term memory retrieval, WM and covert attention, regardless of eye movements (e.g., Kosslyn, 1994; Miller and Cohen, 2001). This finding lends support to the claim that the lateral PFC plays an

important role in object model verification. Second, using an independent localizer for object-sensitive regions, the fMRI results also suggest that these top-down control processes modulate activity in regions that represent and process object properties during the course of model verification. Third, the behavioral interference results demonstrate that a Shape WM task, designed to engage top-down control processes for model verification, interferes with categorizing MI objects more than does a Digit WM task that engages these verification processes minimally; by contrast, we did not find such selective interference for LI objects. This finding is as predicted if top-down control processes are required during the categorization of impoverished objects. We next consider the results in more detail, interpreting the specific findings reported earlier.

VLPFC-IPS regions for top-down control

VLPFC

In our view, these findings are best interpreted as implicating the anterior part of the VLPFC (BA 45, 47) in the top-down control of both WM and covert attention processes that subserve the categorization of impoverished stimuli. We draw this inference on the basis of converging evidence. First, the VLPFC, bilaterally, is one of the most reliably activated regions during the categorization of MI relative to LI objects in the present study, as well as in related neuroimaging research that compared pictures of objects degraded by masking, adding noise, or viewing from unusual angles (Bar et al., 2001; Heekeren et al., 2004; Kosslyn et al., 1994, 1997; Sugio et al., 1999).

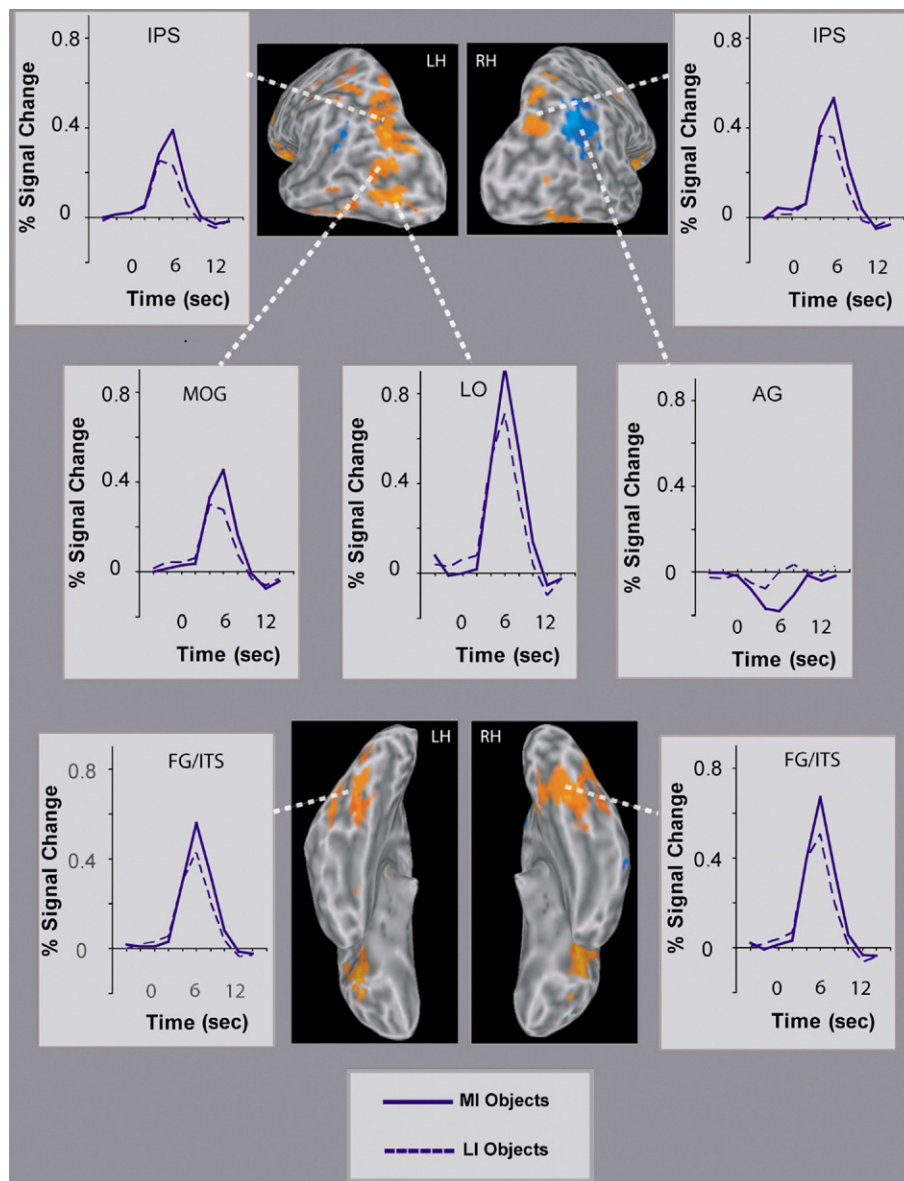


Fig. 4. Activation maps (6 s post-stimulus onset) and time courses of activation (percent signal change is relative to the mean of the corresponding time series, see text for details) showing occipital, temporal and parietal regions significantly active in contrasts of MI versus LI objects (IPS=intraparietal sulcus; MOG=middle occipital gyrus; LO=lateral occipital; AG=angular gyrus; FG/ITS=fusiform gyrus/inferotemporal sulcus). LH=left hemisphere; RH=right hemisphere. Activations are displayed on the same average inflated brain as in Fig. 3. Note, orange and blue regions indicate activations and deactivations, respectively.

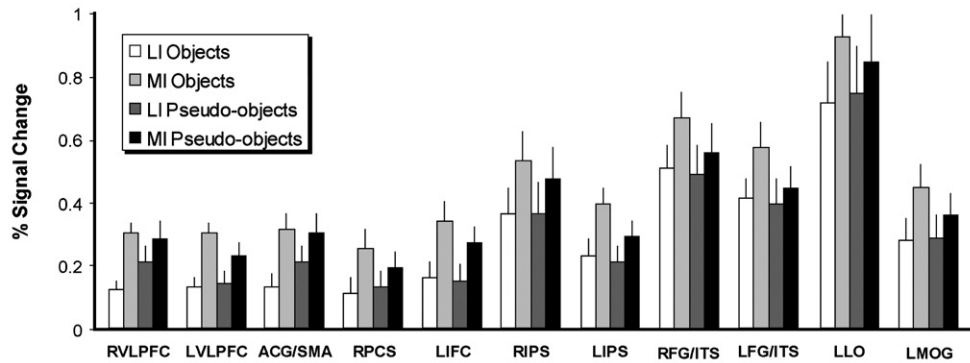


Fig. 5. Percent signal change (relative to the mean of the corresponding timeseries, at the peak occurring at 6 s post-stimulus onset, see text for details) to MI and LI stimuli in the regions listed in Table 1 (excluding the angular gyrus, which exhibited deactivation for all stimuli). Error bars represent the standard error across subjects. “L” means “left”, and “R” means “right”.

Second, VLPFC is part of a prefrontal network that is recruited during the control of attention by WM functions, such as the online monitoring of performance to achieve a goal, including object categorization (de Fockert et al., 2001; Freedman et al., 2003; Lavie et al., 2004; Miller and Cohen, 2001; Rainer and Miller, 2000; Tomita et al., 1999). VLPFC has been implicated not only in simple WM maintenance tasks (Smith and Jonides, 1999) but also in tasks requiring complex operations on information held in WM (Petrides, 2005), such as the inhibition and switching of attentional sets (Konishi et al., 1999; Monchi et al., 2001). VLPFC is thought to play a critical role in the dynamic interplay between facilitation of task-relevant information vs. the inhibition of irrelevant information in posterior cortex.

The output from these initial cognitive control processes in VLPFC can then be passed on to other PFC regions to achieve more complex goals, such as monitoring and manipulation functions in the dorsolateral PFC (Petrides, 2005). In fact, areas 44, 46, and 10 are connected bidirectionally with VLPFC and other PFC areas. We found that medial PFC, the anterior cingulate, the medial supplementary motor area, and a precentral gyrus region near the frontal eye fields, were more strongly activated during the categorization of MI objects than during the categorization of LI objects. These areas have been implicated in selective attention (i.e., directed to one specific attribute of a stimulus among others) and WM (Wager et al., 2004; Wager and Smith, 2003). In addition, medial PFC may be recruited to facilitate switching of selective attention (Beauchamp et al., 2001; Corbetta and Shulman, 2002; Wager et al., 2004).

Third, VLPFC has the functional and neuroanatomical properties expected for an area that computes the perceptual decisions required for object categorization. VLPFC has strong bidirectional connections with object-sensitive regions in the ventral stream, has been implicated in visual memory encoding (Brewer et al., 1998; Kirchoff et al., 2000; Wagner et al., 1998), and VLPFC neurons learn to code stimulus shape and feature information in terms of their relevance for categorization (Rainer and Miller, 2000). Intriguingly, VLPFC neuronal responses and human behavioral performance with impoverished images of objects both show learning curves that are maximal when just enough information is available in the learned item to support categorization (Rainer and Miller, 2000; Snodgrass and Feenan, 1990). In general, VLPFC seems to be a key node in a brain network that implements and acquires the associative mappings that relate patterns of visual cues to the decisions or actions required to perform a task (Bunge, 2004; Bussey et al., 2001; Passingham et al., 2000).

Intraparietal sulcus (IPS)

To categorize objects in MI images, subjects must search for visual attributes of objects that will allow them to verify that the input matches a specific stored object model. This consideration suggests an interpretation for our finding of robust bilateral activation in a swath of posterior parietal regions along the IPS when subjects processed MI relative to LI images. Most importantly, these regions have been implicated in diverse attention tasks, including those requiring covert shifts of attention, such as visual search (Beauchamp et al., 2001; Corbetta and Shulman, 2002; Downing et al., 2001; Hopfinger et al., 2000; Wager et al., 2004; Wojciulik and Kanwisher, 1999).

In addition, caudal parts of IPS, including dorsal object-sensitive region DF2 (Grill-Spector et al., 2000), may contribute other spatial, perceptual, and representational functions to the object model verification process. In the right hemisphere, this region has been implicated in object-oriented spatial tasks, including categorizing objects from unusual views (i.e., another type of impoverished image) (Kosslyn et al., 1994; Sugio et al., 1999; Turnbull et al., 1997), object mental rotation (Kosslyn et al., 1998), object-directed action (James et al., 2002), and binding visual object features (Friedman-Hill et al., 1995). After all, our MI images are composed of spatially disparate elements that may need to be bound together to create a spatially coherent image that can be matched to stored object representations. In the left hemisphere, caudal IPS is part of a frontoparietal network implicated in WM, grasping, perception, and action-related processes with objects, such as many of those used in our study (Chao and Martin, 2000; Grezes et al., 2003); the frontal component of this network, in inferior frontal cortex (BA 6 in ventral precentral sulcus; possible human homologue of monkey area F5), was also implicated herein. Finally, object-sensitive region DF2 also overlaps with parts of BA 19, which has been suggested to be a storage site for object and spatial perceptual knowledge that is activated during object model verification processes (Kosslyn et al., 1994).

Ventral extrastriate regions for object-processing and visual representation

Although the temporal smoothing in fMRI data does not allow us to demonstrate directly that the frontoparietal network modulates ventral posterior regions, studies in human and non-human primates provide convergent evidence that such top-down processing exists (Corbetta and Shulman, 2002; Fuster et al., 1985;

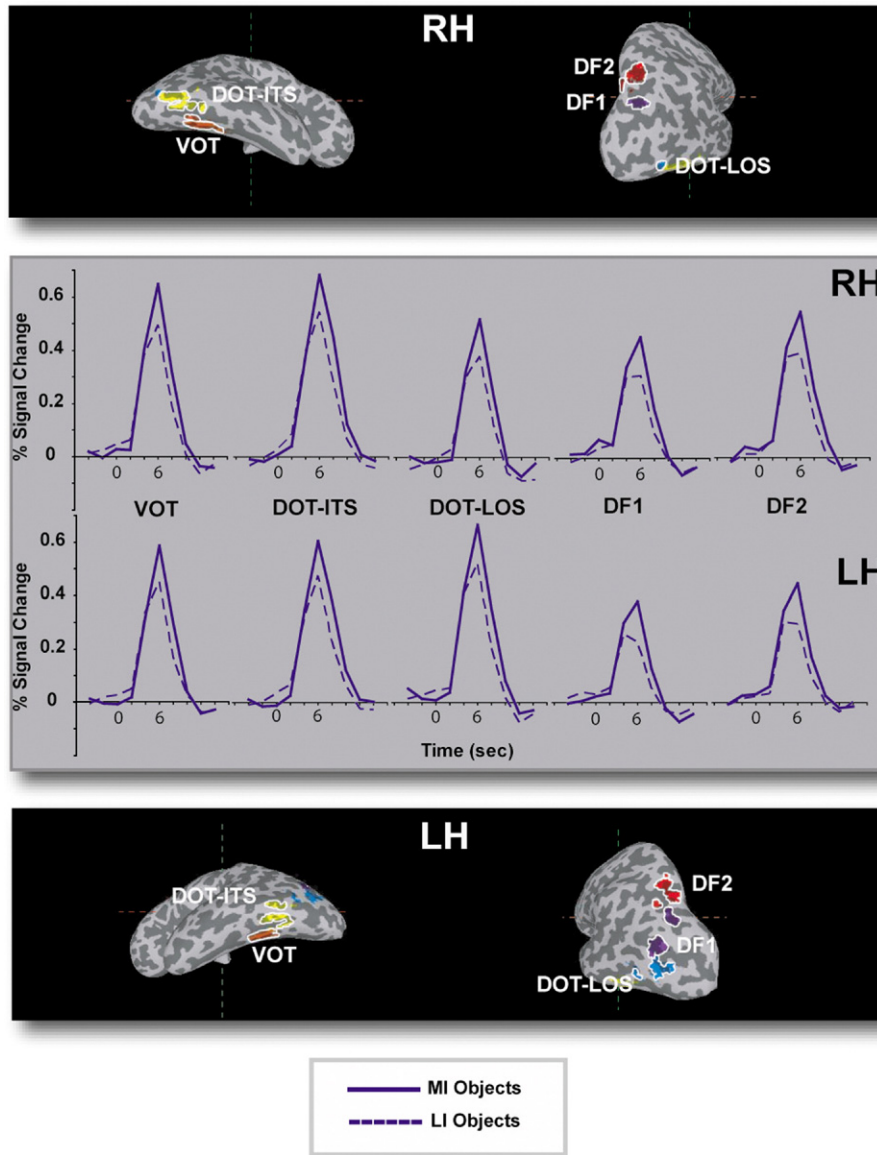


Fig. 6. Time courses of activation in object-sensitive regions activated in both the MI-LI objects comparison and the independent object-sensitive localizer task (percent signal change is relative to the mean of the corresponding time series, see text for details). The specific regions were defined using Talairach coordinate ranges from prior studies of object-sensitive regions: VOT (ventral occipitotemporal), DOT-ITS (dorsal OT-inferotemporal sulcus), DOT-LOS (DOT-lateral occipital sulcus), DF1 (dorsal focus in occipital cortex), and DF2 (dorsal focus in ventral caudal IPS). LH=left hemisphere; RH=right hemisphere. Activations are displayed on the same average inflated brain as in Fig. 3.

Hopfinger et al., 2000; Kastner and Ungerleider, 2000; Kosslyn, 1994; Mechelli et al., 2004; Miller and Cohen, 2001; Moore and Armstrong, 2003; Petrides, 2005; Ranganath and D'Esposito, 2005;

Tomita et al., 1999). For instance, Fuster et al. (1985) demonstrated that the neurons in monkey inferotemporal cortex are less selective during the delay period of a WM task when the prefrontal cortex is

Table 2
Brain areas identified by the eye movement localizer

| Region | Volume | Talairach coordinates (center of mass) | | | Talairach coordinates (range) | | | | | | |
|--------------------------|--------|--|-----|-----|-------------------------------|-------|-------|-------|-------|-------|----|
| | | x | y | z | Min x | Max x | Min y | Max y | Min z | Max z | |
| Precentral sulcus/gyrus | R | 507 | 26 | -5 | 48 | 22 | 30 | -9 | -1 | 42 | 53 |
| Precentral sulcus/gyrus | L | 1710 | -28 | -9 | 46 | -43 | -19 | -14 | -4 | 36 | 55 |
| Intraparietal sulcus | | 594 | -29 | -53 | 49 | -32 | -24 | -58 | -46 | 39 | 57 |
| Lateral occipital sulcus | R | 621 | 46 | -65 | -3 | 39 | 53 | -70 | -59 | -6 | 1 |
| Lateral occipital sulcus | L | 147 | -40 | -69 | -3 | -43 | -38 | -73 | -65 | -4 | 0 |
| Medial occipital | | 6428 | -1 | -91 | 0 | -26 | 30 | -102 | -71 | -10 | 12 |

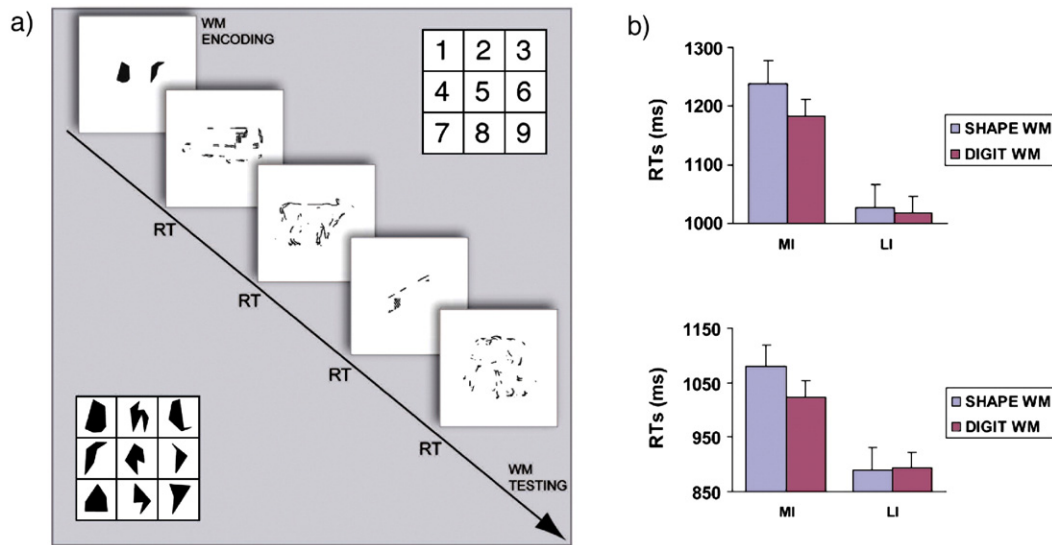


Fig. 7. (a) Diagram of a trial in Experiment 2 and illustration of the nine digits and shapes used in the WM tasks (RT=voice onset time); (b) Behavioral data from Experiment 2; average RTs for categorizing MI and LI objects in the *Shape WM* and *Digit WM* tasks during the first (top) and second (bottom) block of trials. Error bars represent the standard error across subjects.

deactivated bilaterally (by reversible cooling); Tomita et al. (1999) demonstrated that prefrontal cortex directly affects responses in inferotemporal cortex in monkeys during a cued recall task; and Moore and Armstrong (2003) demonstrated that subthreshold electrical stimulation of monkey frontal eye fields enhances the responses of neurons in area V4 to preferred visual stimuli. In our fMRI study, we predicted that top-down modulation would cause object-sensitive areas to be more strongly activated during the categorization of MI than LI objects. Indeed, this was what we found in areas VOT, DOT-ITS, DOT-LOS, DF1, and DF2 (Grill-Spector et al., 2000; Hasson et al., 2003).

Purely bottom-up accounts of visual categorization cannot easily explain this finding. Bottom-up accounts are based on the well-established, hierarchical organization of visual cortex: More posterior regions respond well to elementary parts, whereas more anterior regions respond best to higher-order parts or whole objects (e.g., Lerner et al., 2004; Tanaka, 1996). Such results predict the opposite of our findings: The reduced visual information available in MI objects should result in reduced activation for MI relative to LI objects in more posterior areas, as well as successive processing areas along the visual hierarchy. This is because, on average, MI objects have fewer visual features than LI objects and so they should activate a smaller neural population and perhaps should activate these neurons more weakly. Although this was the observed pattern in prior neuroimaging studies based on a bottom-up framework and aimed at other issues, the studies were not designed specifically to reveal top-down processing in the ventral stream. For example, Lerner et al. (2004) superimposed opaque bars over line drawings, creating images of objects that were occluded to varying degrees (i.e., impoverished), but, crucially, all of the objects were actually easy to categorize via purely bottom-up processes. In other studies (Bar et al., 2001; Grill-Spector et al., 2000), presentation rates may have been too fast to allow top-down processes to gather additional information from the stimulus, and the backward masking used would further minimize feedback influences (Rolls et al., 1999). However, we note that Grill-Spector et al. (2000) did not collect data from prefrontal cortex, and Bar et al. (2001) did, in fact, find

increased prefrontal activation for masked (relative to unmasked) object stimuli that were presented for very brief durations.

The present study, by contrast, was designed to allow us to observe evidence that top-down control processes are normally recruited during the categorization of impoverished objects, and we obtained the opposite pattern of effects in ventral regions from what a bottom-up account would predict.

Network operation in object model verification

We have suggested that the VLPFC is responsible for the top-down control processes that evaluate stored models in posterior object-sensitive regions, consistent with model verification theories (Kosslyn, 1994; Lowe, 1985, 2000). Moreover, our finding of greater activation across entire VLPFC-posterior networks not only motivates us to focus on VLPFC as having a role in facilitating neural processes in other brain regions, but also leads us to propose that the VLPFC is engaged in four particular types of cognitive control processes. Specifically, (a) *Retrieving Distinctive Attributes*: VLPFC activates perceptual information in long-term memory that is used to guide subsequent top-down processing. (b) *Attribute WM*: VLPFC maintains retrieved knowledge about distinctive attributes of candidate object models. (c) *Covert Attention Shifting*: VLPFC initiates attentional shifts so that additional attributes, especially those suggested by the hypotheses activated by a candidate object, are examined further and tested against the candidate match. These VLPFC processes result in facilitation and inhibition of activity in object-processing regions, driven also by IPS areas (Beauchamp et al., 2001; Corbetta and Shulman, 2002). (d) *Attribute Biasing*: VLPFC biases neural representations of salient visual features in the object-sensitive regions, thereby facilitating detection and selection of these relevant features. These subprocesses may occur simultaneously and continue until sufficient information has been obtained to select a response, including deciding that the available evidence is not sufficient to categorize the visual input. The role of the VLPFC is critical

when the ventral pathway and the initial attentional set produce only a weak match for the impoverished visual input.

Although strong inferences about the time courses of processing are not possible because the hemodynamic responses are a temporally smoothed reflections of the underlying neural responses, we find it intriguing that our proposal provides a possible (although far from conclusive) interpretation of the time course of the hemodynamic response in the ventral pathway and VLPFC. We proposed that categorization of MI stimuli recruits top-down control processes after an initial bottom-up pass through posterior object-sensitive regions, and that top-down control processes enhance later activation of object-processing areas until categorization succeeds. Consistent with this idea, in the ventral pathway, activation is initially (at 4 s) above baseline for both MI and LI objects, but reliable differences between the two conditions begin only later (at 6 s). On the other hand, activation in the VLPFC and other prefrontal regions for cognitive control differs between MI and LI objects both at the earlier and later times (4 and 6 s). Further work will be required to evaluate the actual timing of the neural response.

Alternative explanations?

One might argue that the posterior effects we observed reflect visual differences among the stimuli at different fragmentation levels: By construction, MI images have fewer pixels than LI images, on average. However, this difference would predict that activation would be stronger in visual areas for the LI stimuli with more pixels than the MI stimuli with fewer pixels. To the contrary, the opposite results were obtained: We found stronger activation in all extrastriate areas when subjects categorized MI images than when they categorized LI images.

Generic task difficulty is often discussed in the literature as an alternative explanation for condition differences. By this account, our results implicate an extended network of brain regions, all comparably modulated by task difficulty, although one cannot infer whether the modulation in visual cortex is an inherent local response or a result of prefrontal projections. To our knowledge, such a widespread network has not been previously reported to differ between two conditions within the same task. However, such a generic task difficulty explanation does not seem compelling on empirical and theoretical grounds. First, our results are not consistent with such an account. RT is often taken as a metric of task difficulty. Given this, we observed that RTs for the pseudo-objects were much longer (by about 500 ms) than those for the objects, which would predict more activation for pseudo-objects than objects. To the contrary, no region shows even a hint of stronger activation to pseudo-objects than to objects, and, in some areas, the opposite occurred: Activation to objects was reliably stronger than activation to pseudo-objects. Furthermore, although the difference in RTs between MI and LI object stimuli was similar in magnitude to those between MI and LI pseudo-object stimuli, the difference in brain activation (i.e., impoverishment effect) was reliably larger for objects than for pseudo-objects.

Second, from a theoretical perspective, generic task difficulty is vaguely defined and does not make clear predictions for the direction of effects: Although task difficulty is often taken to predict greater activation, greater prefrontal activity has been found not only for harder than easier tasks but also for the opposite, for easier than harder tasks (Bor et al., 2003). More important, our object model verification theory is an attempt to provide a well-

articulated explanation of how specific neural processes differ when the stimuli are more or less impoverished. It is important to realize that being harder to categorize is an inherent characteristic of impoverished stimuli, compared to intact ones. Hence, one could try to match generic task difficulty between the two conditions by equating the RTs, say, through reducing the visual contrast of intact pictures of objects, thereby making them harder to categorize. However, that would merely create another type of impoverished stimulus. Image impoverishment thus is a more precise way to think about factors affecting categorization processes than generic task difficulty.

In addition, these considerations stand against an explanation that the observed results are simply due, in a generic way, to the subjects' paying more attention to any image that is more impoverished, as opposed to the specific engagement of a prefrontal control network subserving the goal of visual object categorization. A generic attention explanation would predict greatest activation for pseudo-objects because the search for a match can potentially continue indefinitely with these stimuli. As just mentioned, no region shows this pattern.

Alternatively, one might argue that our findings reflect differential eye movements in the MI and LI conditions, but the results of our standard eye movement localizer make this interpretation unlikely. Only six scattered voxels that were more strongly activated in the MI condition than in the LI condition overlapped with saccade regions. Moreover, most saccade regions defined by the eye movement localizer task showed no differences between MI and LI conditions. However, because our eye movement localizer task did not use pictures of objects, we cannot rule out the possibility that some parietal effects were influenced by differential eye movements specifically to objects.

Finally, one could also hypothesize that some of the observed effects might be due to post-categorization processes, given that the objects were on the screen for 2200 ms, longer than the RTs. According to this hypothesis, stimuli that were classified more quickly might undergo more post-categorization processes and eye movements/attention shifts than those that subjects classified more slowly. If this were the case, the pattern of results should be exactly the opposite of what was found in Experiment 1.

Top-down control processes are necessary for categorization of MI objects

The logic of Experiment 2 rested on a straightforward idea: If top-down control processes are required when people categorize impoverished objects, then a concurrent task (i.e., Shape WM) that engages processes recruited for object model verification should interfere with the categorization of MI objects more than a concurrent task (i.e., Digit WM) that engages processes not recruited during object model verification (although some interference may be present, due to overall effort expended under dual task conditions). Furthermore, this effect should be smaller or absent during the categorization of LI objects, because, in this case, top-down control processes should be less important for correct categorization. The results were as predicted.

By our top-down account of categorization of the MI objects, visuospatial WM is a necessary component of model verification processes. This explains why a WM task that competes for the same neural resources used in such top-down processing slows the categorization of impoverished images of objects. Our Shape WM task required subjects to keep in WM not only the identity and

structure of the shapes but also their spatial order. In contrast, the relative positions of the items in the Digit WM task could be kept in WM by means of phonological rehearsal. This result provides evidence that top-down WM control processes, which appear to be supported by the VLPFC-parietal network revealed in Experiment 1, are necessary for the categorization of impoverished pictures of objects, and not an epiphenomenon.

Summary and conclusions

In this study, neuroimaging and behavioral results were used to test three key predictions of object model verification theory, which implicates top-down control processes during the categorization of impoverished objects. The findings confirm our predictions. First, the categorization of impoverished objects engages a VLPFC-IPS network involved in top-down control. Second, coincident with the operation of this frontoparietal network, the categorization of impoverished objects engages object-sensitive regions in the ventral stream more strongly than does the categorization of less impoverished pictures of objects. Third, top-down control processes, which convergent evidence suggests are supported by this VLPFC-IPS network, play a causal role in categorizing impoverished pictures of objects. This evidence supports the view that a comprehensive account of object categorization under all viewing conditions requires characterizing the top-down processing contributions that have been little studied with neuroimaging.

Acknowledgments

G.G. and H.E.S. contributed equally to this research endeavor. We thank two anonymous reviewers for insightful comments. We are grateful to Dr. Chantal E. Stern and the cognitive neuroimaging lab for insightful discussions of prefrontal anatomy and function. This research was supported in part by Tufts University faculty research funds and grants NIMA: NMA201-01-C-0032; DARPA: FA8750-05-2-0270; NIH: 1 R21 MH068610-01A1, and 2 R01 MH060734-05A1; NIA: NRSA 5 F32 AG005914. The MRI infrastructure at the A. Martinos Center for Biomedical Imaging was supported in part by the National Center for Research Resources (P41RR14075) and the Mental Illness and Neuroscience Discovery (MIND) Institute.

References

- Awh, E., Jonides, J., 2001. Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci.* 5, 119–126.
- Bar, M., Tootell, R.B., Schacter, D.L., Greve, D.N., Fischl, B., Mendola, J.D., Rosen, B.R., Dale, A.M., 2001. Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29, 529–535.
- Beauchamp, M.S., Petit, L., Ellmore, T.M., Ingeholm, J., Haxby, J.V., 2001. A parametric fMRI study of overt and covert shifts of visuospatial attention. *NeuroImage* 14, 310–321.
- Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147.
- Bor, D., Duncan, J., Wiseman, R.J., Owen, A.M., 2003. Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* 37, 361–367.
- Brewer, J.B., Zhao, Z., Desmond, J.E., Glover, G.H., Gabrieli, J.D.E., 1998. Making memories: brain activity that predicts how well visual experience will be remembered. *Science* 281, 1185–1187.
- Bunge, S.A., 2004. How we use rules to select actions: a review of evidence from cognitive neuroscience. *Cogn. Affect. Behav. Neurosci.* 4, 564–579.
- Bussey, T.J., Wise, S.P., Murray, E.A., 2001. The role of ventral and orbital prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (*Macaca mulatta*). *Behav. Neurosci.* 115, 971–982.
- Chao, L.L., Martin, A., 2000. Representation of manipulable man-made objects in the dorsal stream. *NeuroImage* 12, 478–484.
- Collins, D., Neelin, P., Peters, T., Evans, A., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev., Neurosci.* 3, 201–215.
- Corbetta, M., Miezin, F.M., Shulman, G.L., Petersen, S.E., 1993. A PET study of visuospatial attention. *J. Neurosci.* 13, 1202–1226.
- Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.
- Dale, A.M., 1999. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* 8, 109–114.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage* 9, 179–194.
- de Fockert, J.W., Rees, G., Frith, C.D., Lavie, N., 2001. The role of working memory in visual selective attention. *Science* 291, 1803–1806.
- Downing, P., Liu, J., Kanwisher, N., 2001. Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia* 39, 1329–1342.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9, 195–207.
- Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246.
- Friedman-Hill, S.R., Robertson, L.C., Treisman, A., 1995. Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science* 269, 853–855.
- Fuster, J.M., Bauer, R.H., Jervey, J.P., 1985. Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Res.* 330, 299–307.
- Ganis, G., Kosslyn, S.M., in press. Multiple mechanisms of top-down processing in vision. In: Funahashi, S. (Ed.). *Representation and brain*. Springer Verlag.
- Grezes, J., Armony, J.L., Rowe, J., Passingham, R.E., 2003. Activations related to “mirror” and “canonical” neurons in the human brain: an fMRI study. *NeuroImage* 18, 928–937.
- Grill-Spector, K., Malach, R., 2004. The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677.
- Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* 3, 837–843.
- Hasson, U., Harel, M., Levy, I., Malach, R., 2003. Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37, 1027–1041.
- Heekeren, H.R., Marrett, S., Bandettini, P.A., Ungerleider, L.G., 2004. A general mechanism for perceptual decision-making in the human brain. *Nature* 431, 859–862.
- Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C., 1998. Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* 22, 324–333.
- Hopfinger, J.B., Buonocore, M.H., Mangun, G.R., 2000. The neural mechanisms of top-down attentional control. *Nat. Neurosci.* 3, 284–291.
- James, K.H., Humphrey, G.K., Vilis, T., Corrie, B., Baddour, R., Goodale, M.A., 2002. “Active” and “passive” learning of three-dimensional object structure within an immersive virtual reality environment. *Behav. Res. Methods Instrum. Comput.* 34, 383–390.
- Kastner, S., Ungerleider, L.G., 2000. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341.

- Kirchhoff, B.A., Wagner, A.D., Maril, A., Stern, C.E., 2000. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *J. Neurosci.* 20, 6173–6180.
- Konishi, S., Kawazu, M., Uchida, I., Kikyo, H., Asakura, I., Miyashita, Y., 1999. Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the Wisconsin Card Sorting Test. *Cereb. Cortex* 9, 745–753.
- Kosslyn, S.M., 1994. *Image and Brain*. MIT Press, Cambridge, MA.
- Kosslyn, S.M., Alpert, N.M., Thompson, W.L., Chabris, C.F., Rauch, S.L., Anderson, A.K., 1994. Identifying objects seen from different viewpoints. A PET investigation. *Brain* 117 (Pt 5), 1055–1071.
- Kosslyn, S.M., Thompson, W.L., Alpert, N.M., 1995. Identifying objects at different levels of hierarchy: a positron emission tomography study. *Hum. Brain Mapp.* 3, 107–132.
- Kosslyn, S.M., Thompson, W.L., Alpert, N.M., 1997. Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *NeuroImage* 6, 320–334.
- Kosslyn, S.M., DiGirolamo, G.J., Thompson, W.L., Alpert, N.M., 1998. Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography. *Psychophysiology* 35, 151–161.
- Lavie, N., Hirst, A., de Fockert, J.W., Viding, E., 2004. Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen.* 133, 339–354.
- Lerner, Y., Harel, M., Malach, R., 2004. Rapid completion effects in human high-order visual areas. *NeuroImage* 21, 516–526.
- Lowe, D., 1985. *Perceptual Organization and Visual Recognition*. Kluwer Academic, Boston.
- Lowe, D., 2000. Towards a computational model for object recognition in IT cortex. Paper presented at: First IEEE International Workshop on Biologically Motivated Computer Vision (Seoul, Korea).
- Macwhinney, B., Cohen, J., Provost, J., 1997. The PsyScope experiment-building system. *Spat. Vis.* 11, 99–101.
- Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. U. S. A.* 92, 8135–8139.
- Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb. Cortex* 14, 1256–1265.
- Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Miller, B.T., D'Esposito, M., 2005. Searching for the “top” in top-down control. *Neuron* 48, 535–538.
- Monchi, O., Petrides, M., Petre, V., Worsley, K., Dagher, A., 2001. Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J. Neurosci.* 21, 7733–7741.
- Moore, T., Armstrong, K.M., 2003. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421, 370–373.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Oliver, R.T., Thompson-Schill, S.L., 2003. Dorsal stream activation during retrieval of object size and shape. *Cogn. Affect. Behav. Neurosci.* 3, 309–322.
- Passingham, R.E., Toni, I., Rushworth, M.F., 2000. Specialisation within the prefrontal cortex: the ventral prefrontal cortex and associative learning. *Exp. Brain Res.* 133, 103–113.
- Perrett, D., Oram, M., 1993. Neurophysiology of shape processing. *Image Vis. Comput.* 11, 317–333.
- Petrides, M., 2005. Lateral prefrontal cortex: architectonic and functional organization. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* 360, 781–795.
- Poggio, T., Edelman, S., 1990. A network that learns to recognize three-dimensional objects. *Nature* 343, 263–266.
- Rainer, G., Miller, E.K., 2000. Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron* 27, 179–189.
- Ranganath, C., D'Esposito, M., 2005. Directing the mind's eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Curr. Opin. Neurobiol.* 15, 175–182.
- Riesenhuber, M., Poggio, T., 2002. Neural mechanisms of object recognition. *Curr. Opin. Neurobiol.* 12, 162–168.
- Rolls, E.T., Tovee, M.J., Panzeri, S., 1999. The neurophysiology of backward visual masking: information analysis. *J. Cogn. Neurosci.* 11, 300–311.
- Romanski, L.M., 2004. Domain specificity in the primate prefrontal cortex. *Cogn. Affect. Behav. Neurosci.* 4, 421–429.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P., 1976. Basic objects in natural categories. *Cogn. Psychol.* 8, 382–439.
- Rousselet, G.A., Thorpe, S.J., Fabre-Thorpe, M., 2003. Taking the MAX from neuronal responses. *Trends Cogn. Sci.* 7, 99–102.
- Schendan, H.E., Kutas, M., 2002. Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia* 40, 931–945.
- Schendan, H.E., Ganis, G., Kutas, M., 1998. Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology* 35, 240–251.
- Smith, E.E., Jonides, J., 1999. Storage and executive processes in the frontal lobes. *Science* 283, 1657–1661.
- Smith, E.E., Medin, D.L., 1981. *Categories and Concepts*. Harvard University Press, Cambridge, MA.
- Snodgrass, J.G., Corwin, J., 1988. Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture set. *Percept. Mot. Skills* 67, 3–36.
- Snodgrass, J.G., Feenan, K., 1990. Priming effects in picture fragment completion: support for the perceptual closure hypothesis. *J. Exp. Psychol. Gen.* 119, 276–296.
- Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215.
- Snodgrass, J.G., Smith, B., Feenan, K., Corwin, J., 1987. Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications. *Behav. Res. Methods Instrum. Comput.* 19, 270–274.
- Sugio, T., Inui, T., Matsuo, K., Matsuzawa, M., Glover, G.H., Nakai, T., 1999. The role of the posterior parietal cortex in human object recognition: a functional magnetic resonance imaging study. *Neurosci. Lett.* 276, 45–48.
- Tanaka, K., 1996. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139.
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., Miyashita, Y., 1999. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401, 699–703.
- Turnbull, O.H., Beschin, N., DellaSala, S., 1997. Agnosia for object orientation: implications for theories of object recognition. *Neuropsychologia* 35, 153–163.
- Wager, T.D., Jonides, J., Reading, S., 2004. Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage* 22, 1679–1693.
- Wager, T.D., Smith, E.E., 2003. Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274.
- Wagner, A.D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, A.M., Rosen, B.R., Buckner, R.L., 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281, 1188–1191.
- Wallis, G., Rolls, E.T., 1997. Invariant face and object recognition in the visual system. *Prog. Neurobiol.* 51, 167–194.
- Wojciulik, E., Kanwisher, N., 1999. The generality of parietal involvement in visual attention. *Neuron* 23, 747–764.
- Xiong, J., Gao, J.-H., Lancaster, J.L., Fox, P.T., 1995. Clustered analysis for functional MRI activation studies of the human brain. *Human Brain Mapp.* 3, 287–301.