ELSEVIER

Cognitive Brain Research 5 (1996) 55–67

COGNITIVE
BRAIN
RESEARCH

Functional imaging of human visual recognition

Nancy Kanwisher [a,b,*], Marvin M. Chun [a,b], Josh McDermott [a,b], Patrick J. Ledden [b]

[a] *Department of Psychology, Harvard University, Cambridge, MA 02138, USA*
[b] *Massachusetts General Hospital NMR Center, Charlestown, MA 02129, USA*

Keywords: Visual recognition; Human

1. Introduction

While human visual recognition has been extensively studied with behavioral, computational, and neuropsychological techniques for decades, it is only in the last 3–5 years that functional brain imaging methods have been exploited to investigate the brain structures involved in this process. In this paper we review this work, discuss the unique methodological and interpretive challenges it raises, and present some of our own preliminary studies on the brain basis of visual recognition.

We take visual recognition proper to include two main components: (i) the high-level perceptual analysis of a visual stimulus (for example, the construction of a structural description of an object's shape), and (ii) the matching of that perceptual description to a stored visual representation in memory (e.g., determining that the shape looks more like a chair than a bicycle). Note that this definition excludes both very 'early' visual processes such as edge extraction, and very 'late' processes such as accessing the name or meaning of a visually-presented stimulus.

Functional imaging studies can advance our understanding of three different aspects of human visual recognition: its neuroanatomy, neurophysiology, and modularity. First, cortical areas are defined not only by aspects of brain hardware like cytoarchitectonics and anatomical connectivity, but also by function [11]. Thus to the extent that functional imaging studies reveal regions of cortex selectively involved in specific functions, these regions become candidates for distinct anatomical areas in human cortex. Second, functional mapping of human cortex can help establish homologies between visual areas in humans and macaques [32]. This will allow us to bring together an understanding of the specific computations carried out in a given visual area with the vast body of knowledge from single-unit studies of the homologous area in macaques, making possible a new and powerful convergence between the cognitive psychology and the single-unit physiology of higher-level vision. Third, functional brain imaging provides another technique for cognitive psychologists to investigate the modularity of the mind: to the extent that different regions of human cortex are activated when different component processes within visual recognition are engaged, this will provide evidence that these components are indeed functionally dissociable processes.

Before the advent of functional brain imaging, the main technique for approaching these research goals was to study the behavioral deficits of patients with focal brain damage. While this is a venerable and powerful technique (see [10]), functional brain imaging can complement it in several ways. Most importantly, the interpretation of data from patients with brain damage can be clouded by the possibility that the brain has become reorganized as a result of the damage. Second, brain damage tends to affect large and diffuse regions, making isolation and anatomical localization of any one function difficult. Third, brain damage may disrupt performance at a specific task not because the damaged region is critical to the computations underlying that task, but instead because the damage has disrupted neural pathways that would normally carry critical information to another cortical region which carries out that computation. In contrast, functional imaging studies in normal subjects allow generalization to other normal subjects, can provide precise anatomical localization (see for example Fig. 4C), and are unlikely to indicate regions involved merely in the passage of information between cortical areas (because white matter rarely if ever shows activation in functional imaging studies). On the other hand, it should be kept in mind that studies of patients with focal brain damage have other significant advantages over imaging studies, such as the power to support inferences that a given area of the brain is *necessary* to, rather than simply involved in, a given task. Thus the two approaches

[*] Corresponding author.

PII S0926-6410(96)00041-9

are both important in different ways and can serve as converging sources of evidence bearing on the same set of research questions.

However, the design of a functional brain imaging study can be tricky, and unless a few key rules are followed, the results may be uninterpretable. Most importantly, when trying to localize a particular mental process X in the brain by comparing the brain activation that results from two different conditions, A (which engages X) and B (which does not), it is critical that conditions A and B not also differ in the other mental processes they engage [1]. How can conditions be designed such that the difference between them includes a single kind of mental process? The two main options are to vary (i) the stimuli presented, or (ii) the task the subject is requested to carry out on those stimuli. A good rule of thumb in functional imaging is to vary *either* the stimulus *or* the task between conditions, not both.

A PET study illustrating the kind of difficulties that arise when both stimuli and task are varied at the same time was reported by Sergent et al. [33]. They asked whether there are distinct brain regions specialized for face and object recognition, and found that face recognition primarily activated a ventro-medial region in the right hemisphere, whereas object recognition primarily activated an occipitotemporal region of the left hemisphere. To find object recognition areas, they subtracted the activation resulting when subjects judged the orientation of sine-wave gratings from the activation that resulted when subjects categorized photographs of objects as living or nonliving. Because both the stimulus and the task changed between conditions, it is not clear what processes go on in the areas activated by the object task but not the grating task, with possibilities including not only visual recognition processes, but also (i) extraction of any visual features (other than those included in the gratings), (ii) covert object naming, and (iii) extraction of the meanings of the objects recognized. To find areas involved in face recognition, Sergent et al. subtracted the activation which resulted when subjects discriminated the gender of photographs of unfamiliar faces from the activation which resulted when subjects categorized photographs of familiar faces as actors or nonactors. This subtraction is an improvement over the object subtraction in that the stimuli did not differ greatly while the task changed. Nonetheless, this comparison cannot distinguish between activations which result from (i) matching perceptual descriptions to stored visual representations of faces in memory, (ii) covert naming of the individuals depicted, and (iii) accessing semantic information about the individuals depicted. Thus Sergent et al.'s findings do not demonstrate that different brain regions are involved in the high-level visual analysis of faces and objects, but could instead simply reflect differences in either lower-level feature extraction processes or in the postrecognition semantic/linguistic processing of these stimuli.

The Sergent et al. [33] study is not alone in the confounding of stimulus and task manipulations or in the existence of multiple possible interpretations of the data collected. Indeed it is probably true that all imaging studies leave open more than one possible interpretation of the mental processes which underlie each of the activations reported. What is important is to design experiments that keep the number of possible interpretations to a minimum, and to be clear about each of them in the discussion of data. Ultimately we suspect that the only way to deal with this problem will be to use converging operations involving several different orthogonal tests of the same hypothesis; we describe preliminary data from one such approach at the end of this paper (see Fig. 4C).

Given that in a well-designed experiment either the stimulus or the task will be manipulated, but not both, how should one decide which kind of manipulation to use to answer a given question? Suppose one wants to distinguish between (i) early visual feature extraction processes and (ii) higher-level components of visual object recognition. Imagine as the subject that you are presented with a sequence of photographs of familiar objects, each displayed one at a time near the center of gaze for 2 s, and you are instructed to either (i) recognize the pictures (in one condition) or (ii) only analyze the features of the objects but not recognize them. While a subtraction of the second condition from the first should in principle reveal just the higher-level stages of object recognition, the obvious problem here is that it will be impossible to follow the instructions in the second condition. A number of studies indicate that (i) visual object and word recognition is automatic in the sense that it occurs even when subjects are engaged in another simultaneous task while trying to avoid recognizing the objects or words [8,12,13,34,36], (ii) some indirect evidence suggests that face recognition is also automatic [9], and (iii) words, objects, and faces are apparently recognized even when subjects are not attending to them [7,19,37] and in some cases even when they do not enter awareness [3]. It is of course possible that visual recognition might be modulated under some conditions – for example, if the stimuli are presented very briefly and/or subjects are given a sufficiently demanding simultaneous task [22]. But under the conditions of most imaging experiments, stimuli are typically presented for a second or more near fixation and visual recognition is likely to be automatic. In this kind of situation, when subjects cannot control their own mental processes, task manipulations are not likely to be effective. On the other hand imagine presenting subjects with a display that says '64 × 7 = ?'. Most subjects can solve this problem if they try to

[1] Note that although widely used, this standard logic of subtraction assumes that single component processes can be added or deleted without affecting other component processes – an assumption that is testable and can often be incorrect [4].

but also have the option of just looking at the numbers and not bothering to figure out what the answer is. The key difference here is that some tasks (like visual feature extraction and object recognition) are highly automatic, whereas others (like mental arithmetic and visual attention) are controlled (but see also [5]). Task manipulations are most sensible for controlled mental processes and stimulus manipulations are most appropriate for automatic mental processes.

Although we have so far only discussed experimental designs which attempt to localize single mental processes in the brain, another approach is to ask what brain regions are involved in the execution of an entire complex task. An example of a study using this approach was reported by Kosslyn et al. [21], who used PET to localize the many different component processes entailed in Kosslyn's model of object recognition. Subtracting the brain activity that results when subjects name line drawings of canonical views of objects from the activations when subjects name line drawings of noncanonical views of objects, these researchers found significant activation in six areas within the right hemisphere and four in the left. Kosslyn et al. offer explanations for the particular processes underlying each of these activations in the context of their multi-component model of object recognition. While such studies can be powerful in their potential to localize many different processes at once, because of the large number of activations observed, the only way to determine which activations are due to which component processes is to rely heavily on prior knowledge and/or theories.

2. Literature review

In just the last few years a large number of studies have used PET and fMRI to explore the brain loci involved in human visual recognition. This work is briefly reviewed here.

2.1. Task manipulations

A series of studies by Haxby and his colleagues [14,15] asked whether human visual cortex is organized into the same 'what' and 'where' pathways that have been extensively studied in the macaque [39]. Although their first study [14] confounded task and stimulus manipulations, Haxby et al. [15] used an improved design in which the same stimuli were used for both the face-matching and location-matching tasks (an array of three faces, each in its own box but slightly off center). They found occipitotemporal activations largely in the fusiform gyrus bilaterally for the face-matching task but occipitoparietal activations in the location-matching task, consistent with the organization of the visual cortex in the macaque. Kohler and colleagues [20] used a similar design to ask whether the ventral pathway in humans is also involved in the visual recognition of objects (i.e. as well as faces). They showed subjects pairs of two sequentially-presented displays, each containing three objects. Subjects were asked to judge in one condition whether the three locations were the same, and in another condition whether the three objects were the same. Areas that were significantly more active in the identity task than the location task included the inferior temporal cortex in the region of the fusiform gyrus (Brodmann areas 19 and 37) in the left hemisphere, extending posteriorly into the lingual gyrus (Brodmann areas 18 and 17), and in the ventral occipital cortex of the right hemisphere in the region of the fusiform gyrus–suggesting that these areas are involved not only in face recognition (as Haxby et al. had shown) but also in visual object recognition.

Note that these results contrast with those reported earlier by Sergent et al. [33] who placed object recognition processes in the left hemisphere and face recognition processes in the right. For the reasons outlined above, we find the Haxby et al. [15] and Kohler et al. [20] studies more convincing. Nonetheless, within-subject testing will be necessary before any solid conclusions can be reached about the overlap (or lack thereof) in brain areas involved in visual face and object recognition. In the third section of this paper we present some of our own results which suggest that at a finer grain different patches of ventral occipitotemporal cortex may be involved in the visual recognition of faces and objects (see also [1,2]).

While the Haxby et al. [15] and Kohler et al. [20] studies are well designed, it is perhaps somewhat surprising that task manipulations were so effective in manipulating visual recognition processes, which we have suggested are usually highly automatic. One possibility is that the subjects made different patterns of eye movements in the two tasks, foveating the faces or objects in the identity tasks but different aspects of the array in the location tasks. This possibility is strengthened by the fact that the stimuli were very large and were presented for several seconds, conditions which would encourage eye movements. If so, then the retinal stimulation in the two conditions might be different and the reported activations would reflect low-level feature-analysis processes as well as higher-level visual recognition processes, consistent with the fairly large swaths of cortex activated in these studies. (This possibility could be tested by presenting the three faces in each trial sequentially while subjects maintain fixation.) However, even if retinal stimulation was not a confound in the Haxby et al. and Kohler et al. studies, the activations reported in these experiments still may reflect not visual recognition per se–which would have occurred automatically in both tasks and hence be invisible in the subtraction–but instead the effects of attending to such information [6,25], encoding it in short-term memory, and/or using it as the basis for a decision. Of course it is also possible that we have exaggerated the automaticity of visual recognition, and that it may be modulated by some tasks even when the items are presented clearly for several seconds and retinal stimulation is identical in the two tasks.

Schacter et al. [30] reported another PET study which focused more specifically on visual shape extraction. Subjects viewed line drawings of 3-D novel objects which were either physically possible or impossible. Compared to passive viewing of the same stimuli, an 'object decision' task (deciding whether the objects were possible or not) activated areas in the inferior temporal and inferior fusiform gyri – but only for the physically possible objects. These and other data were taken as evidence that the inferior temporal and fusiform regions are "selectively involved in computing global representations of structurally coherent three-dimensional objects" (p. 590). One might wonder why a task manipulation was so effective in this experiment given the arguments above. One possibility is that shape analysis is not automatic when the stimuli are presented extremely briefly, as they were in this experiment (50 ms). Another possibility is that the kind of shape processing that was required in Schacter et al.'s object decision task is much more difficult, effortful, and controlled than that involved in 'normal' object recognition. Whether there is a single shape-analysis system that can be activated to different degrees depending on task difficulty, or whether the areas activated in Schacter's study are different from those involved in normal object recognition remains to be determined.

In sum, the three studies described above used task manipulations and found activations in occipitotemporal regions during visual shape and object recognition tasks. However if we are correct that visual shape analysis and

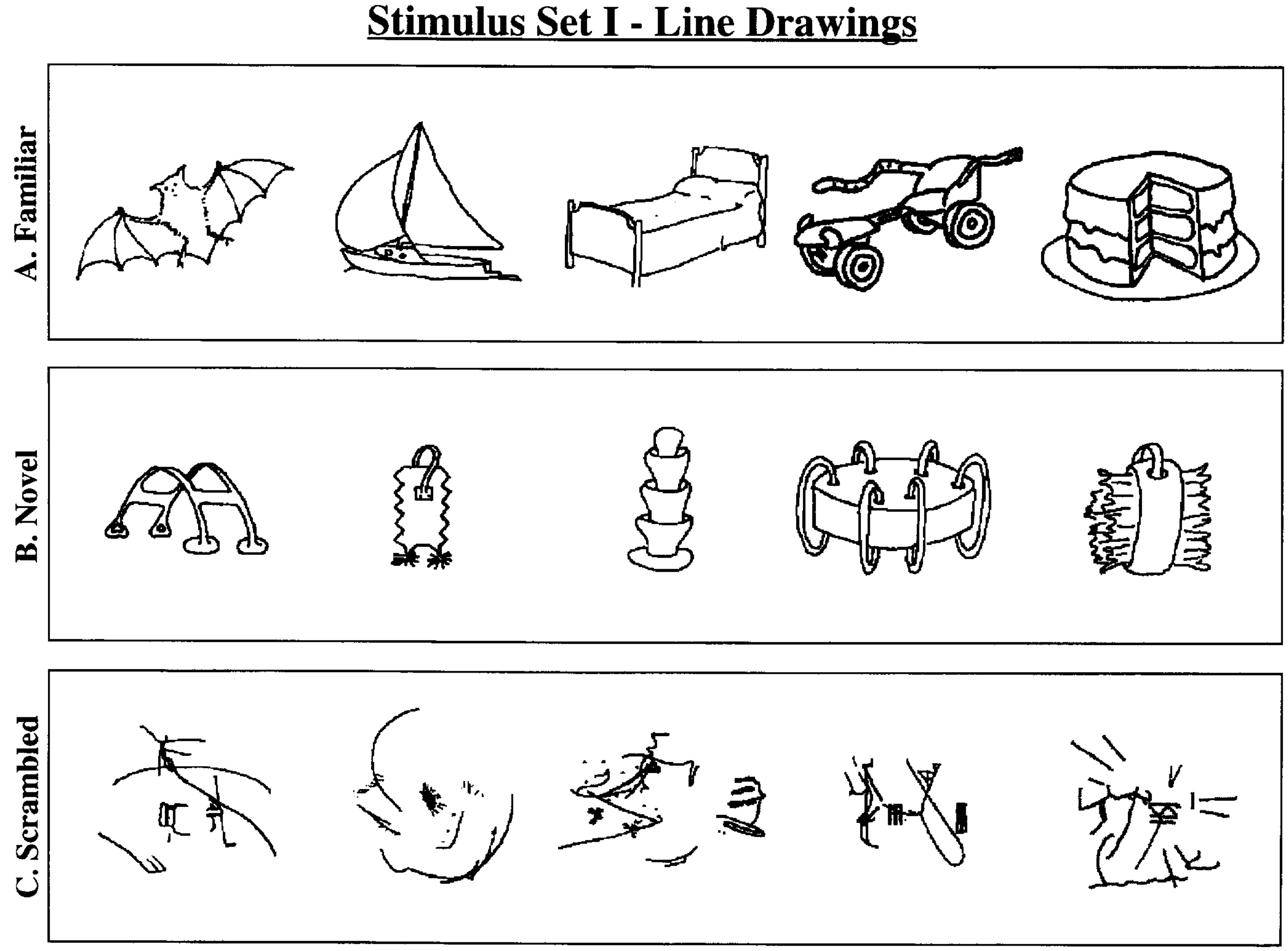


Fig. 1. Examples of the three different stimulus conditions used in Kanwisher et al.'s [18] PET study and in the present Expt. 1. The pictures in the novel-object set were drawn by hand to resemble the familiar objects in every possible way except for familiarity. The familiar objects, which were adapted from the Snodgrass and Vanderwart [35] set, were then redrawn by the same person to match the line width and handwriting used in constructing the novel set. The scrambled objects were created by first digitally dividing each of the 140 familiar object stimuli into five line-component subsets (that recreated the entire familiar object without overlap when superimposed), and then superimposing 5 subset images from 5 different objects (sampling without replacement) to make each scrambled image. Thus, the scrambled and familiar stimulus sets were perfectly matched for average luminance at each pixel, total line contour length, and local orientation (although object structure inevitably covaries to some extent with spatial frequency and with certain visual features like T-junctions).

object recognition is automatic for clearly-presented stimuli under normal viewing conditions, then the activations revealed under such task manipulations may reflect other controlled processes (e.g. attention, working memory, or decision processes) rather than visual recognition per se. If so, then varying the stimulus might be a more effective way to study visual recognition. We turn next to several studies which did just that.

2.2. Stimulus manipulations

In a now-classic PET study, Petersen et al. [27] found a left medial extrastriate area which was more active when subjects passively viewed words or orthographically regular nonwords than when they fixated on a point, but which was not more active during viewing of orthographically irregular nonwords or words written with 'false fonts' than during fixation. They argued that this area is involved in the extraction of the orthographic structure of a visually-presented word, which is likely to be a key component process in word reading. Although this study has been criticized on other grounds [4,16,28], the logic of its elegantly simple design has been recruited in a number of more recent studies.

In one such study on face recognition Puce et al. [29] used fMRI to find brain loci that responded more strongly to intact than scrambled faces. These researchers found significant activations in the fusiform and inferior temporal gyri (in addition to some other areas) in 9 out of 12 subjects, consistent with results of Haxby et al. [15] as well as with earlier electrophysiological recordings by Allison et al. [1]. As the authors note, their data support only the claim that these areas are face-sensitive, not that they are face-selective. Further, the scrambling procedure used to generate the face stimuli in this experiment produced changes in the low-level feature components of the two sets of images. For example, numerous tiny edges were created in the scrambled-face images, which may account for the significant activations also found in the inverse comparison of scrambled versus intact faces. This is a common problem in brain imaging studies in which 'scrambled' control images are created by cutting and pasting parts of test images – and one which we grapple with later in this paper.

Malach et al. [24] used fMRI data to argue that a new extrastriate area ('LO', for lateral occipital complex) at the lateral-posterior aspect of the occipital lobe just posterior to area MT, is involved in an intermediate stage of visual object recognition. This claim is based on the fact that area LO responded more strongly to photographs of familiar objects, famous faces, and unfamiliar 3-dimensional objects, compared to texture fields and a variety of other control stimuli. While these results are important and provocative, and there is some safety in the sheer number of kinds of stimuli tested, most of the control stimuli used by Malach et al (e.g. Aldus Superpaint textures) were not matched in any particular way to the object pictures (e.g., a teddy bear, a pond with ducks, etc). Such images are bound to differ in a host of low-level visual features, so it may be premature to argue that areas responding more strongly to the objects represent medium- or higher-level stages of visual object analysis. Malach et al. also generated control stimuli which were matched to the object images in Fourier power spectra, by scrambling the phase of the original object images. Such controls are helpful in ruling out power-spectrum accounts of the observed activations, but are open to other alternative accounts because they differ from the object stimuli in other significant ways such as the complete lack of edges [23].

A recent PET study by Kanwisher et al. [17,18] asked subjects to passively view line drawings of either (a) familiar objects, (b) novel objects that were similar to the familiar objects in complexity, three-dimensionality, and part structure, or (c) scrambled versions of the familiar objects (see Fig. 1) which preserved the exact retinotopy, average luminance, total contour length, and other features of the familiar-object set. A lateral and inferior extrastriate area straddling the anterior occipital sulcus was more active bilaterally when subjects passively viewed familiar or novel compared to scrambled stimuli. Because this area was at least as strongly activated by novel as familiar objects, the activation is unlikely to reflect processes associated with memory-matching, naming, or accessing semantic information [2]. Kanwisher et al. therefore proposed that it is involved in the bottom-up construction of shape descriptions from simple visual features.

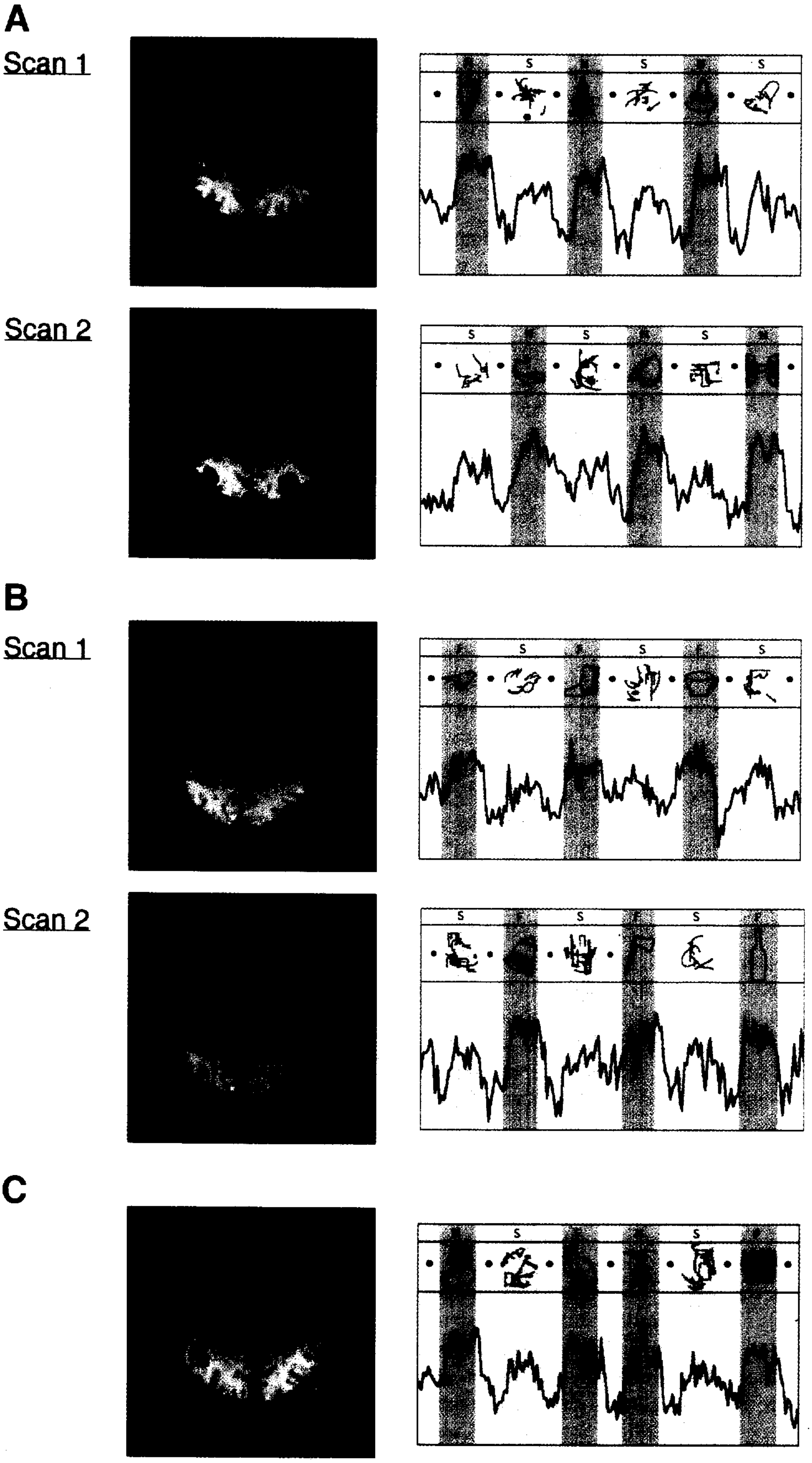
Although the exact areas activated in the above studies differ, this brief review serves to demonstrate that results from many different labs, paradigms, and techniques are beginning to demonstrate that different components of visual recognition activate different areas in the human ventral pathway. However, because of the significant anatomical variability across subjects, the only way to carry out a real comparison of different components of visual recognition is to do extensive within-subject studies. Because fMRI allows essentially unlimited repeated testing of the same individual subjects it is ideally suited to this kind of work. Next we report some preliminary results from single subjects on some of the experiments we have carried out in the last year addressed to this goal.

3. Preliminary results from fMRI

3.1. Experiment 1A

In our first study, we replicated with fMRI the main result from the Kanwisher et al. [18] PET study – namely,

[2] Note that this interpretation is based on the assumption that more neural activity occurs when a given computation runs successfully and produces an output than when it either does not run on a given input or runs but does not produce an output.


A
Scan 1
S
S
S
Scan 2
S
S
S
B
Scan 1
S
S
S
Scan 2
S
S
S
C
S
S

activation of a bilateral inferolateral extrastriate area by line drawings depicting clear 3-dimensional shapes compared to line drawings not depicting shapes. The subject was tested in a 1.5-Tesla GE scanner using a gradient echo pulse sequence. We also tested whether the activation would be found in the same area in two independent runs within a single subject. We used the Novel and Familiar line-drawing stimuli from the Kanwisher et al. set (see Fig. 1), in a standard fMRI experimental design. While lying in the scanner the subject viewed a 5.5-min long sequence of visual stimuli, back-projected onto a ground-glass screen and reflected to a convenient viewing position by a mirror just over the subject's head. Seven 7-mm thick slices perpendicular to the calcarine cortex were scanned. Each scan lasted 5.5 min, during which time an image was collected from each slice once every 2 s (TR = 2), for a total of 165 images per slice.

A diagram of the stimulus presentation sequence is shown in the top-right panel in Fig. 2A: 30-s epochs of visual stimulation (during which stimuli were presented at a rate of 1.5 pictures/second with no blank interval between successive pictures) alternated with 20-s periods of fixation. Activation epochs (indicated by grey vertical bars) alternated between those in which Novel stimuli were presented (dark grey) and those in which Scrambled stimuli were presented (light grey), with 20-s fixation intervals after each. Order was counterbalanced across the scans. The subject was instructed before each scan to simply lie still, view the stimuli attentively, and fixate on the fixation point when it was present.

To analyze the data, Kolmogorov-Smirnov statistics were applied to each voxel in each of the 7 slices, testing for the significance of any differences between the 45 images collected while the subject was viewing Novel drawings to the 45 images collected while the subject was viewing Scrambled drawings.

The results from two scans in one subject are shown in Fig. 2A. Because in this and subsequent experiments we used a surface coil placed over the back of the head, our signal was only strong from posterior regions and we report only occipital and posterior temporal activations. On the left is a high-resolution anatomical (T1-weighted) image of a single slice perpendicular to the calcarine sulcus, overlaid with the color-coded statistical results: a bilateral region of activation at the inferolateral surface of the brain showed significantly greater signal intensity during the Novel than Scrambled epochs. A number of adjacent voxels in this slice reached a $P < 10^{-8}$ level of significance. As in the PET study, this area of significant activation is on the inferolateral surface of the brain near the occipitotemporal junction.

This result, while highly statistically significant, could nonetheless arise from a number of different artifacts, such as motion, physiological noise, or machine noise. These concerns can, however, be allayed by visual inspection of the raw data. Fig. 2A (top right) shows the time course of raw signal intensity over the 5.5 min of the scan in a rectangular region of interest (ROI) centered in one of the activated regions (indicated by the yellow square). Raw time course data such as this makes it clear that the significance represented in the colorized brain image is not artifactual but instead reflects the dependence of signal strength in this region (and hence presumably neural activity in it) on the type of stimuli viewed by the subject. A similar pattern is visible in individual voxels in the activated area.

A skeptic might nonetheless worry that with the large number of voxels available to sample from one might be able to find ROIs with this characteristic pattern arising due to chance alone. This concern is easily allayed by conducting a second independent run in the same subject with the same stimuli. If the pattern observed in the scan described above were due to chance then we would not expect a similar activation pattern to occur in the same brain region in an independent run. Yet Fig. 2A (bottom) shows just such a replication in the identical ROI in the same subject during the scan that immediately followed the one shown above it. This replication provides strong evidence that MR signal intensity in this region is reliably affected by whether the stimuli viewed by the subject depict coherent 3-D shapes.

3.2. Experiment 1B

Experiment 1A compared only the Novel and Scrambled stimuli from the Kanwisher et al. [18] set, but did not include the Familiar stimuli. To demonstrate that this area is also active when familiar objects are recognized, the same result must be shown for a comparison of Familiar and Scrambled objects. Fig. 2B shows the data from a

Fig. 2. Results from Expt. 1. A: the top of the top right figure, Scan 1, diagrams the sequence of stimulus events during a single 5.5-min scan in Expt. 1A. The brain image at the top left shows a slice perpendicular to the calcarine cortex, cutting diagonally through the cerebellum which appears here above the occipital cortex. The areas in which the signal strength was significantly greater during the novel than scrambled epochs are shown in color, with region of interest (ROI) indicated by the yellow rectangle in the activated area in the left hemisphere. (All brain images in this paper are shown in conventional radiological coordinates with the left hemisphere on the right.) The the top right figure shows the raw average of signal strength in that ROI, sampled once every 2 s, over the 5.5 min of the scan. The bottom brain slice and time-course figure, Scan 2, shows the analogous information from a later independent scan in the identical ROI in the same subject, with the order of stimulus epochs reversed. B: analogous results for two different scans in a single subject in Expt. 1B. Activation images and time courses of signal intensity in the identical ROI for each scan reveal an area that is significantly more active during viewing of Familiar than Novel stimuli. C: results for Expt. 1C. A near-axial brain slice showing regions with significantly higher signal intensity during periods when either familiar or novel images were presented, compared to periods when scrambled images were presented.

different subject on an experiment which was the same as Expt. 1A except that (i) epochs containing Familiar (rather than Novel) items were alternated with epochs containing Scrambled items, and (ii) while a gradient echo (GE) pulse sequence was used in Expt. 1A, less vein-sensitive asymmetric spin echo (ASE) pulse sequence was used in Expt. 1B (and in the rest of the experiments in this paper).

The results from one subject in Expt. 1B are shown in Fig. 2B. Here a region in the inferior surface of the brain can be seen which produced significantly higher MR signal ($P < 10^{-8}$ in many different voxels) when the subject viewed Familiar compared to Scrambled items. The brain slices on the left show areas of significant activation and ROIs selected in those areas; the time courses on the right show the average of raw signal intensity in that ROI over the 5.5 min of the scan. Again, visual inspection of the raw time course data leaves little room for any interpretation other than that the MR signal in this area increased when the subject viewed Familiar compared to Scrambled objects. The bottom row of Fig. 2B shows a replication of this result in a second scan in the same subject and the same ROI about a half hour after the first scan. In a third scan (not shown here) the MR signal also increased in this same region when the subject viewed Novel compared to Familiar stimuli (as in Expt. 1A). Hence this experiment represents another replication of the Kanwisher et al. [18] result.

3.3. Experiment 1C

One difficulty with Expts. 1A and 1B is that because Novel and Familiar stimuli were not presented in the same scan, we cannot directly compare the response of these two kinds of stimuli to each other. In this experiment we included two epochs for each of the three kinds of stimuli (Familiar, Novel, and Scrambled; see Fig. 2C, top right), so that we could directly compare the responses of the Familiar and Novel stimuli to each other, while keeping the Scrambled condition as a baseline.

The results from one run in one subject are shown in Fig. 2C. We looked for any voxels in which the MR signal was significantly greater for Familiar and Novel compared to Scrambled stimuli; these areas are shown on the left. Inspection of the time course of an ROI selected from this region shows that Novel and Familiar epochs did indeed

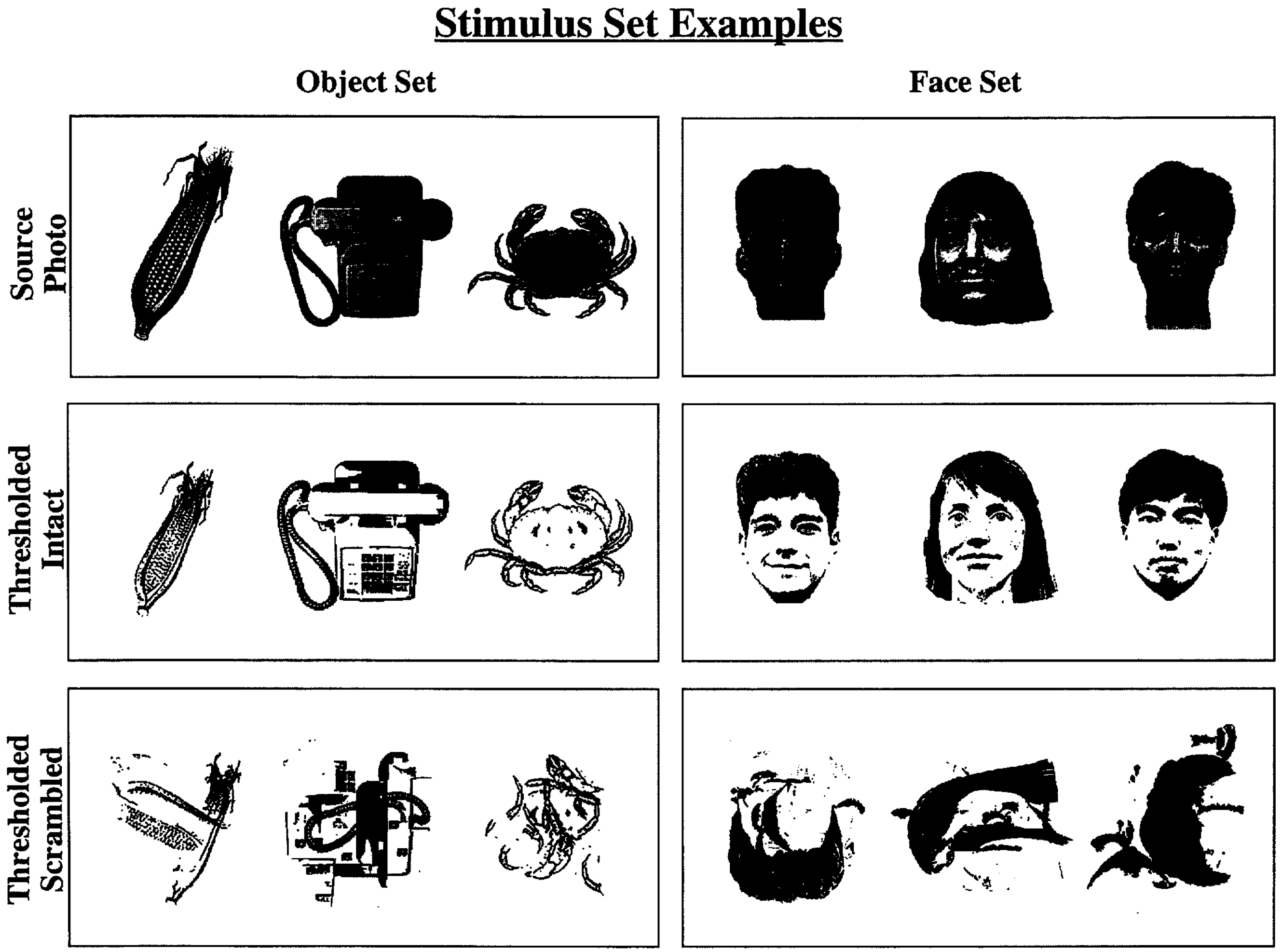


Fig. 3. Example stimuli for Expts. 2 and 3. See text.

produce higher signal intensity in this region (at right) than did Scrambled epochs (as expected), but – more importantly – no difference was visible in the heights of the peaks for Novel and Familiar stimuli. (Note however that because we used a surface coil placed over the back of the head, we could only observe occipital and posterior temporal activations in these experiments.) This finding is consistent with the PET results of [18], and with the interpretation that this area is involved in the analysis of visual shape.

Similar patterns of data to those described above in Expts. 1A–1C were observed in at least one run each in six different subjects. These results replicate the Kanwisher et al. PET study and strengthen the argument that a bilateral extrastriate area at the occipitotemporal junction is involved in the extraction of visual shape information. On the other hand, a number of other subjects run on these experiments did not show any discernible difference between the conditions. Although it is possible that such variation results from stable differences in brain organization or hemodynamic response between individuals, we also thought it possible that line drawings might be relatively weak stimuli [24,26] and that more robust results might obtain if more realistic stimuli were used.

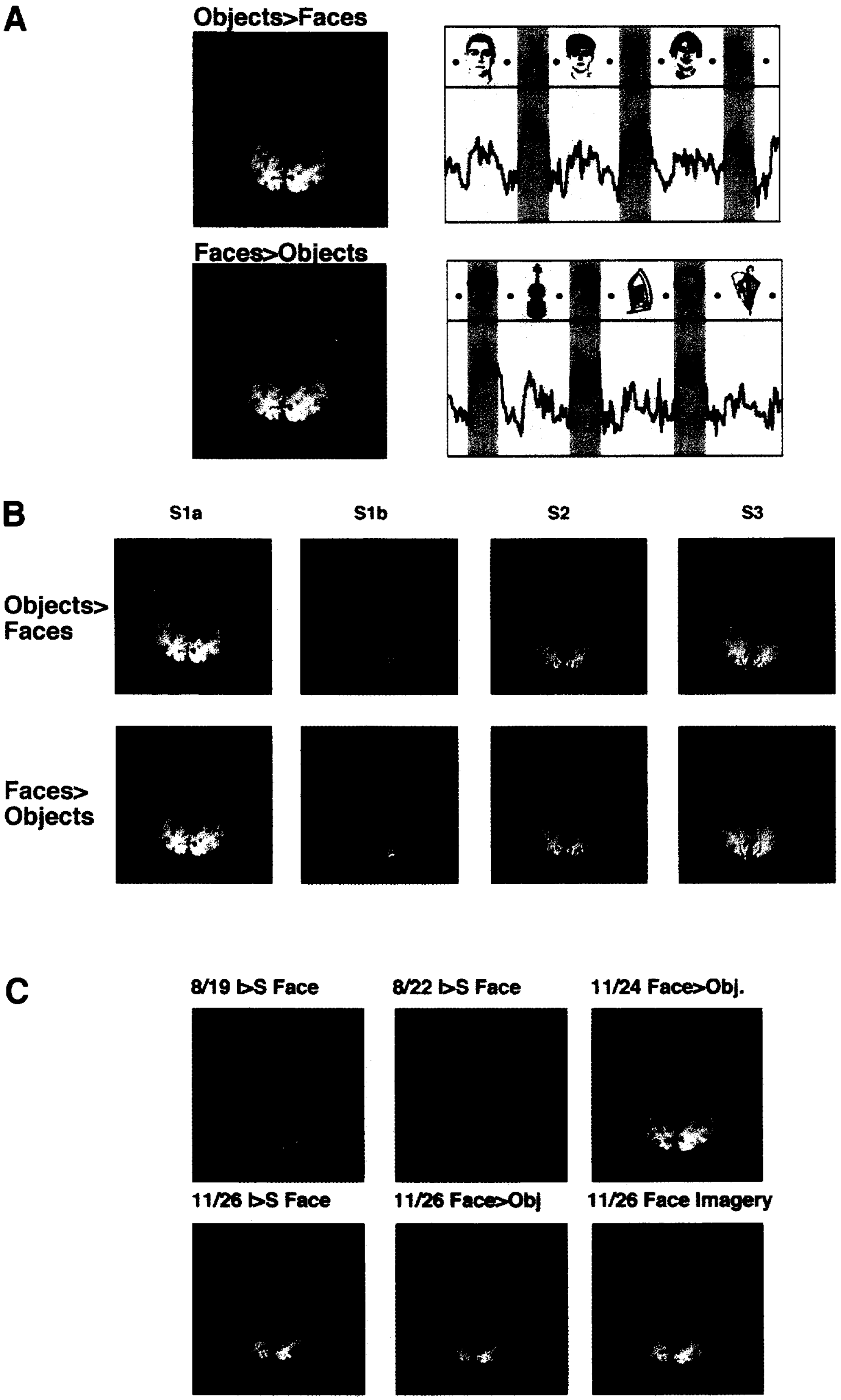
3.4. Experiment 2

The reason we began our investigations with line drawings in the first place, rather than photographs, was that we felt the scrambled control drawings could be better matched to their intact drawings than in any analogous manipulation with photographs. However, given the possibility that line drawings simply do not drive the visual system as strongly as photographs [26], and given the further problem that line drawings are just not adequate in capturing faces, we devised a new strategy. A set of 92 object/scrambled object pairs and 92 face/scrambled face pairs were created as follows. Grey-level photographs (Fig. 3, top row) of objects and (unfamiliar) faces were 'thresholded'; that is, all grey levels below a certain brightness threshold were assigned to black and all above were assigned to white (Fig. 3, middle row). For most objects and faces (and for all those included in our study), the thresholded images remained highly recognizable. However because these images contained relatively large patches of black in a sea of white, we were able to move the black regions relative to each other (without leaving behind cut-and-paste marks) to generate our scrambled control stimuli (Fig. 3, bottom row). The scrambled control stimuli were thus exactly matched to the experimental stimuli in average luminance, and approximately matched in complexity and in many low-level visual features. Note that this procedure does not create spurious edges in the scrambled condition which are not present in the intact condition, a problem that has plagued numerous previous studies of visual recognition.

Thus by comparing the brain regions which are more active when subjects view intact than scrambled objects with those which are more active when subjects view intact than scrambled faces, we hoped to isolate the intermediate and higher-level stages of visual recognition of objects and faces, while minimizing confounds from the different low-level features present in the face and object stimulus set. Any areas which are active in the intact versus scrambled object subtraction but not the intact versus scrambled face subtraction would be candidate loci specialized for object but not face recognition and vice versa; areas active in both subtractions would be implicated in a domain-general aspect of visual processing.

These stimuli have now been run on ten subjects, using methods similar to those described in Expt. 1. The results have been rather less than we hoped for. Three of the ten subjects showed areas which had a significantly higher signal intensity while subjects viewed intact than scrambled faces (but no analogous object effect), and three subjects showed the intact versus scrambled object effect but no face effect, whereas only one subject showed both areas which responded to objects and areas which responded to faces. The loci of the activations were in general either near the inferolateral areas on the occipitotemporal junction which were activated by the line drawings in Expt. 1 or in nearby, usually more medial, locations. (See Fig. 4C for an example of an area activated by intact compared to scrambled faces.) For the subject who showed areas responsive to both the face and object subtractions, there was a great deal of overlap in these areas. Further, although there were some areas which responded significantly in the object comparison but not the face comparison, there were few if any pixels which showed the reverse effect. The region of the object activations was adjacent to (and indeed partially overlapping with) an area, presumably MT, which responded more strongly to expanding and contracting rings than to stationary rings [38]. The finding from this one subject fits nicely with the results from Expts. 1A–1C, and with earlier work [18,24], all of which found that activity in this neighborhood of cortex increased when coherent 3-D shapes were presented, compared to various control stimuli not depicting clear 3-D shapes.

Although this experiment raises important issues, it did not answer the question we set out to address: whether the high-level components of face and object recognition are carried out in the same or different brain areas. This question can only be answered when the same subject shows significant activations for both the face and object comparisons, but as noted above only one of ten subjects run on both stimulus sets showed both effects. Why are the results from these stimuli not more robust? One possibility is that the thresholded stimuli are (as we hypothesized for line drawings) just too impoverished to drive the visual system effectively. Alternatively, perhaps the scrambled stimuli are just too good! If higher-level processing of


A
Objects>Faces
Faces>Objects
B
S1a
S1b
S2
S3
Objects>
Faces
Faces>
Objects
C
8/19 I>S Face
8/22 I>S Face
11/24 Face>Obj.
11/26 I>S Face
11/26 Face>Obj
11/26 Face Imagery

faces and objects consists in large part in the extraction of complex features of the items presented (rather than the construction of shape representations), and if similar complex features are present in the intact and scrambled objects, then the key areas involved in face and object recognition might be activated to a similar degree by our scrambled and intact stimuli. Indeed, large regions of visual cortex were strongly activated by both scrambled and intact items, compared to fixation controls, which is consistent with this possibility (though far from proving it).

A third possibility is that in visual recognition, most of the neural activity tends to reflect computational *effort* rather than computational *success*. If so, then it is possible that just as much computational effort is expended (and hence just as strong an MR signal results) in processing the scrambled stimuli as the intact stimuli, even though a coherent perceptual representation is ultimately delivered only in the intact-stimulus case. This points to a general caveat against overinterpreting null results when stimulus manipulations are used (see also footnote 1). Recent progress has been made, however, in devising imaging paradigms which can distinguish between activations which result from computational effort from those which result from computational success [31].

The considerations above suggest that scrambled pictures may have significant shortcomings as control stimuli in efforts to localize specific components of visual recognition. There is yet another possible problem with scrambled stimuli, however. If visual attention is more strongly recruited by intact than scrambled stimuli, then some of the activation observed in intact-versus-scrambled comparisons in our own and numerous other studies may reflect the operation of general attentional mechanisms, rather than of computations specific to visual recognition. In the next experiment we tried a different tack, one which circumvents these concerns about scrambled stimuli by avoiding them altogether.

3.5. Experiment 3

In this experiment subjects viewed either photographs of (unfamiliar) faces or photographs of objects (Fig. 3, top row), and the activation in each of these two conditions was compared directly to the other. Several possible advantages over the previous experiments were envisioned with this simple design. First, photographs may drive the visual system more effectively than either line drawings or thresholded photographs. Second, intact photographs of faces and objects are of similar interest and presumably of similar attention-capturing power, so any differential activation observed in these two conditions is unlikely to reflect the operation of a general attentional mechanism. Finally, the other concerns about comparing intact to scrambled stimuli discussed above can be avoided because this comparison does not use scrambled stimuli. On the other hand, it must be kept in mind that this comparison could in principle reveal differences in face and object perception at any level of processing, from low-level feature extraction to high-level visual analysis, memory matching, and access to semantic information and/or names.

Fig. 4A shows the results of such a scan, with the areas which responded more strongly to objects than faces (and the signal intensity time course for an ROI in this region) shown on the top and the area that responded more strongly to faces than objects (and the time course of an ROI in this region) in the same subject and scan on the bottom. Inspection of the time courses from these areas reveals that the bilateral medial area on the top shows a clear selectivity for objects over faces, and the more anterior right-hemisphere area on the bottom shows a clear selectivity for faces compared to objects.

Fig. 4B shows that this result is highly consistent across three subjects (S1–S3) run on the same paradigm (with activations in the area of the parahippocampal and fusiform gyri bilaterally for objects and the right fusiform gyrus for faces), as well as across two different scanning sessions within the same subject (S1a and S1b). These results would be very difficult to explain in terms of differential recruitment of visual attention by the two stimulus types, and instead strongly imply that different extrastriate areas are involved in the visual processing of faces and objects. While they do not allow us to determine the stage of processing at which these differences occur, the fairly anterior loci and the virtual absence or more posterior activations (in either the face versus object subtraction or vice versa) argue against a low-level interpretation (i.e. a retinotopic or simple featural confound). It is also unclear from the present data whether the more robust results observed in Expt. 3 than Expts. 1 and 2 were due to the use of photographs rather than line drawings or thresholded photographs, or to the direct comparison of objects and faces. Future research will explore this question, for

Fig. 4. A: data from a single scan in a single subject in Expt. 3, showing regions which were more active during object-photo viewing than during face-photo viewing at the top, and areas showing the opposite pattern in the same slice below, with time course data for the ROIs marked by yellow rectangles. B: near-axial slices showing the loci of activation for three subjects (plus a replication of one of them) in Expt. 3. Each slice shows in yellow an ROI in which the time course data (not shown) clearly revealed either higher signal intensity during the object than face epochs (top) or vice versa (bottom). C: the results of six different scans in a single subject showing what appears to be the same area in the right hemisphere which is significantly more active when the subject (i) viewed intact than scrambled thresholded faces in three different scanning sessions (labelled I > S with the date of the scan), (ii) viewed photographs of faces than photographs of objects (labelled Face > Object with the date of the scan), and (iii) imagined faces of familiar people (heard spoken through earphones at a rate of one name every 2 s) compared to rest.

example by directly comparing thresholded faces to thresholded objects.

While we are now only beginning to chart the direction this work will take, it is clear that multiple independent tests will be necessary before we can specify the exact computations that go on in a given cortical area. A preliminary example of how this might work is illustrated in Fig. 4C, which shows the results of six different scans over four different scanning sessions in the same subject. Although the exact slice planes differed across each session the same small face-selective region in the right hemisphere is clearly visible in each scan. This region responded more strongly both to intact than scrambled faces, and to photographs of faces than photographs of objects. Replications of these effects are shown both within and across scanning sessions in this subject. Despite the ambiguity of each effect when considered alone, the fact that these two independent comparisons activate the same area suggests that this region is specialized for construction of high-level visual representations of faces. The lower right scan shows that this region even appears to become active when the subject mentally imaged familiar faces (compared to rest). Although this result will have to be replicated, it further suggests that this face-representing area is involved whether the faces are familiar or unfamiliar, and whether they are visually perceived or simply imagined. Thus, these three independent tests illustrate the power of converging operations in neuroimaging, and demonstrate that a high degree of specificity and replicability can be found in fMRI studies of visual recognition.

3.6. Conclusions

To sum up, we have argued in this chapter that functional imaging can be a powerful technique for exploring the modular structure of visual recognition. However, if this research program is to be successful, careful attention will have to be paid to the design of experimental conditions and the exact mental processes which vary between them. We have argued that for largely automatic processes such as visual recognition, task manipulations may not be appropriate: to the extent that recognition is automatic it will occur in all conditions independent of task and any differences in activations between task conditions must reflect postrecognition processes. On the other hand, stimulus manipulations are not without their own shortcomings, which include (i) the difficulty of designing stimulus sets which engage higher-level processes to different degrees but are well-balanced in terms of their low-level visual properties, (ii) the difficulty of obtaining significant activations in a large enough proportion of subjects when carefully-controlled stimuli are used, (iii) the possibility that attentional confounds arise because general-purpose attentional mechanisms may be engaged to different degrees by different stimuli, and (iv) questions of whether computations run even on inappropriate input (does the shape analysis system operate on the scrambled stimuli from Expt. 1? does the face recognition system operate even on object stimuli?), and whether neural activation is more likely to reflect the successful generation of a coherent output for a given computation, or the effort expended in carrying out the computation independent of its success.

Despite these challenges, we feel optimistic that the ambiguities in the interpretation of any single experiment can be reduced or eliminated by the use of multiple independent tests of the same hypothesis. For example past evidence [18,24] as well as the our own preliminary evidence from Expts. 1 and 2 above is beginning to converge on the conclusion that an inferolateral area at the occipitotemporal junction is involved in extraction of object shape. Second, the data presented in Expt. 3 suggests that more medial and anterior cortical areas are involved in higher-level components of visual face and object recognition, with different areas appearing to be specialized for face and object recognition. Although in several of the experiments described above, the reported activations were significant in less than half of the subjects tested, the object effect described in Expt. 3 has been observed in five out of five subjects tested. This result will have to be demonstrated on more subjects and bolstered by a number of further control conditions [3], but it points the way toward further explorations of modular processes in the human ventral pathway.

Further work in this area will explore a number of questions. Is the modularity of the human ventral pathway genetically hardwired in the brain, or the product of self-organizing neural networks? What is the role of expertise in the construction and/or maintenance of visual modules? How fine-grained are the functions which are carried out by visual modules? Will it be possible to discover with functional brain imaging new functional components of the mind that were not predicted from purely behavioral measures? While the data necessary to answer these questions does not yet exist, we feel optimistic that functional brain imaging involving multiple independent tests in individual subjects will allow us to approach these questions with anatomical precision and cognitive sophistication.

Acknowledgements

This work was supported by NIMH Grant MH45245 to N.K. and NIH Award EY06592 to M.M.C. We thank Roy Hamilton, Charlotte Chui, and Dave Wamback for help with stimulus preparation, Ken Kwong, Roger Tootell, Kathy O'Craven, Ewa Wojciulik, Jody Culham, Raynald

[3] For example, we are now comparing the activation that results when subjects view sequences of houses versus sequences of faces, in order to unconfound the object-face distinction from other factors such as the heterogeneity of the category presented.

Comtois, Mary Foley, Robert Savoy, Bruce Rosen, Ken Nakayama, and Terry Campbell for technical assistance and discussion of the research, and Molly Potter, Janine Mendola, and Alex Holcombe for comments on the manuscript.

References

[1] Allison, T., Ginter, H., McCarthy, G., Nobre, A., Puce, A., Luby, M. and Spencer, D., Face recognition in human extrastriate cortex, *J. Neurophysiol.*, 71 (1994) 821–825.

[2] Allison, T., McCarthy, G., Nobre, A., Puce, A. and Belger, A., Human extrastriate visual cortex and the perception of faces, words, numbers, and colors, *Cereb. Cortex*, 5 (1994) 544–554.

[3] Berti, A. and Rizzolatti, G., Visual processing without awareness: Evidence from unilateral neglect, *J. Cogn. Neurosci.*, 4 (1992) 345–351.

[4] Bookheimer, S.Y., Zeffiro, T.A., Blaxton, T., Gaillard, W. and Theodore, W., Regional cerebral blood flow during object naming and word reading, *Human Brain Mapping*, 3 (1995) 93–106.

[5] Cheng, P.W., Restructuring versus Automaticity: Alternative Accounts of Skill Acquisition, *Psychol. Rev.*, 92 (1985) 414–423.

[6] Corbetta, M. Miezin, F.M., Dobmeyer, S.M., Shulman, G.L., Petersen, S.E., Attentional modulation of neural processing of shape, color, and velocity in humans, *Science*, 248 (1990) 1556–1559.

[7] Dalrymple-Alford, E.C. and Budayr, B., Examination of some aspects of the Stroop colour-word test, *Percept. Motor Skills*, 23 (1966) 1211–1214.

[8] Dunbar, K. and MacCleod, A horse race of a different color: Stroop interference patterns with transformed words, *J. Exp. Psychol. Human Percept. Perform.*, 10 (1984) 622–639.

[9] Farah, M.J., *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision*, MIT Press, Cambridge, MA, 1990.

[10] Farah, M., Dissociable systems for visual recognition: A cognitive neuropsychology approach. In S.M. Kosslyn and D.N. Osherson (Eds.), *Visual Cognition*, MIT Press, Cambridge, MA, 1995, pp. 101–119.

[11] Felleman, D.J. and Van Essen, D.C., Distributed heirarchical processing in the primate cerebral cortex, *Cereb. Cortex*, 1 (1991) 1.

[12] Frith, C.D., Kapur, N., Friston, K.J., Liddle, P.F. and Frackowiak, R.S.J., Regional cerebral activity associated with the incidental processing of pseudo-words, *Human Brain Mapping*, 3 (1995) 153–160.

[13] Glaser, W.R. and Dungelhoff, F.-I., The time course of picture-word in interference, *J. Exp. Psychol. Human Percept. Perform.*, 10 (1984) 640–654.

[14] Haxby, J.V., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B. and Rapoport, S.I., Dissociation of spatial and object visual processing pathways in human extrastriate cortex, *Proc. Natl. Acad. Sci. USA*, 88 (1991) 1621–1625.

[15] Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P. and Grady, C.L., The functional organization of human extrastriate cortex: A PET-fCBF study of selective attention to faces and locations, *J. Neurosci.*, 14 (1994) 6336–6353.

[16] Howard, D., Patterson, K., Wise, R., Brown, W.D., Friston, K., Weiller, C., et al., The cortical localization of the lexicons: positron emission tomography evidence, *Brain*, 115 (1992) 1769–1782.

[17] Kanwisher, N., Woods, R., Iacoboni. M. and Mazziotta, J., PET studies of object recognition. Paper presented at the *Society for Neuroscience*, November, 1994.

[18] Kanwisher, N., Woods, R., Ioacoboni, M. and Mazziotta, J., A locus in human extrastriate cortex for visual shape analysis, *J. Cogn. Neurosci.*, (1996) in press.

[19] Khurana, B., Smith, W.C., Baker, M.T. and Huang, C., Face representation under conditions of inattention, *Invest. Ophthalmol. Vis. Sci.*, 35 (1994) 2147 (Abstract #4135).

[20] Kohler, S., Kapur, S., Moscovitch, M., Winocur, G. and Houle, S., Dissociation of pathways for object and spatial vision: A PET study in humans, *Neuroreport*, 6 (1995) 1865–1868.

[21] Kosslyn, S.M., Alpert, N.M., Thompson, W.L., Chabris, C.F., Rauch, S.L. and Anderson, A.K., Identifying objects seen from different viewpoints: A PET investigation, *Brain*, 117 (1994) 1055–1071.

[22] Lavie, N., Perceptual load as a necessary condition for selective attention, *J. Exp. Psychol. Human Percept. Perform.*, 21 (1995) 451–468.

[23] Li, Z. and Atick, J.J., Toward a theory of striate cortex, *Neural Computation*, 6 (1994) 127–146.

[24] Malach, R., Reppas, J.B., Benson, R.B., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R. and Tootell, R.B.H., Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex, *Proc. Natl. Acad. Sci. USA*, 92 (1995) 8135–8138.

[25] O'Craven, K., Savoy, R. and Rosen, B., Attention modulates fMRI activation in human MT/MST. Paper presented at the *25th Annual Meeting of the Society for Neuroscience*, San Diego, CA, 1995.

[26] Perrett, D.I., Rolls, E.T. and Caan, W., Visual neurons respsponsive to faces in the monkey temporal cortex, *Exp. Brain Res.*, 47 (1982) 329–342.

[27] Petersen, S.E., Fox, P.T., Snyder, A.Z. and Raichle, M.E., Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli, *Science*, 249 (1990) 1041–1044.

[28] Price, C.J., Wise, R.J.S., Watson, J.D.G., Patterson, K., Howard, D. and Frackowiak, R.S.J., Brain activity during reading: The effects of exposure duration and task, *Brain*, 117 (1994) 1255–1269.

[29] Puce, A., Allison, T., Gore, J.C. and McCarthy, G., Face-sensitive regions in human extrastriate cortex studies by functional MRI, *J. Neurophysiol.*, 74 (1995) 1192–1199.

[30] Schacter, D.L., Reiman, E., Uecker, A., Polster, M.R., Yun, L.S. and Cooper, L.A., Brain regions associated with retreival of structurally coherent visual information, *Nature*, 376 (1995) 587–590.

[31] Schacter, D.L., Alpert, N., Savage, C., Rauch, S. and Albert, M.S., Conscious recollection and the human hippocampal formation: Evidence from positron emission tomography, *Proc. Natl. Acad. Sci. USA*, 93 (1996) 321–325.

[32] Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J., Brady, T.J., Rosen, B.R. and Tootell, R.B.H., Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging, *Science*, 268 (1995) 889–893.

[33] Sergent, J., Ohta, S. and MacDonald, B., Functional neuroanatomy of face and object processing, *Brain*, 115 (1992) 15–36.

[34] Smith, M.C. and Magee, L.E., Tracing the time course of picture-word processing, *J. Exp. Psychol. Gen.*, 109 (1980) 373–392.

[35] Snodgrass and Vanderwart, A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity, *J. Exp. Psychol. Human Learn. Memory*, 6 (1980) 174–215.

[36] Stroop, J.R., Studies of interference in serial verbal reactions, *J. Exp. Psychol.*, 18 (1935) 643–662.

[37] Tipper, S.P. and Driver, J., Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli, *Memory Cogn.*, 16 (1988) 64–70.

[38] Tootell, R.B.H., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R. and Belliveau, J.W., Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging, *J. Neurosci.*, 15 (1995) 3215–3230.

[39] Ungerleider, L.G. and Mishkin, M., Two cortical visual systems. In D.J. Ingle, M.A. Goodale and R.J.W. Mansfield (Eds.), *Analysis of Visual Behaviour*, MIT Press, Cambridge, MA, 1982, pp. 549–586.