Identifying objects seen from different viewpoints A PET investigation

Stephen M. Kosslyn,[1,3] Nathaniel M. Alpert,[2] William L. Thompson,[1] Christopher F. Chabris,[1] Scott L. Rauch[2,4] and Adam K. Anderson[1]

[1]*Department of Psychology, Harvard University, Cambridge and the Departments of* [2]*Radiology,* [3]*Neurology and* [4]*Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA*

Correspondence to: S. M. Kosslyn, Department of Psychology, Harvard University, Cambridge, MA 02138, USA

Summary

Positron emission tomography scans were acquired when subjects performed three tasks, each in a separate block of trials. They decided whether words named pictures of objects viewed from a canonical perspective, decided whether words named pictures of objects viewed from a non-canonical (unusual) perspective or saw random patterns of lines and pressed a pedal when they heard the word (this was a baseline condition). The dorsolateral prefrontal region was activated when subjects identified objects seen from non-canonical perspectives, as expected if the frontal lobes are involved in top-down perceptual processing. In addition, several areas in the occipital, temporal and parietal lobes were selectively activated when subjects identified objects seen from non-canonical perspectives, as specifically predicted by a recent theory. Overall, the pattern of results supported the view that the human brain identifies objects by using a system of areas similar to that suggested by studies of other primates.

Key words: object identification; PET; vision

Introduction

The power of the human perceptual system is revealed by our ability to identify objects under a wide range of circumstances. For example, we can identify objects when they are seen in different parts of the visual field, so that their images fall in different places on the retina, and when they are contorted or seen from unusual perspectives, so that their planar projections are very different from any we have seen before. Facts like these proved to be the demise of simple 'template theories' of visual object identification, which posited that the input is directly matched to stored templates, and the template that overlapped the most with the input image was used to establish its identity (*see,* for example, Neisser, 1967). Subsequently, many years of research have revealed that complex functions such as object identification are not accomplished by a single process, implemented in a single part of the brain. Rather, object identification is accomplished by a *system* of processes that work together; the brain apparently carves a very complex problem into a set of simpler problems, each of which is addressed by a distinct process (*see,* for example, Lowe, 1985, 1987*a,b*; Ullman, 1989; Kosslyn *et al.,* 1990; Hummel and Biederman, 1992; Kosslyn, 1994). The nature of the component processes that underlie object identification and their interactions is just now coming into focus, and in this

article we report the results of a PET investigation that expand and sharpen this emerging picture.

Recent theories of visual object identification have been illuminated by discoveries about how the brain encodes visual information. However, much of this work has focused on animal models, particularly the macaque monkey (*see,* for example, Maunsell and Newsome, 1987; Desimone and Ungerleider, 1989). In this article we use findings from animal models to formulate hypotheses about the component processes that allow humans to identify objects. We begin with the strong assumption that human visual processing closely mirrors visual processing in the macaque; numerous researchers have reported very similar visual psychophysical findings in the two species, which lends plausibility to our working hypotheses (e.g. *see,* for example, DeValois *et al.,* 1974*a,b*).

Because we assume that a system of processes underlies object identification, our hypotheses are necessarily complex. Hence, it is particularly important that our predictions be as specific as possible. Results from animal studies lead us to hypothesize six major component processing subsystems, which are illustrated in Fig. 1. We will characterize each component functionally, and propose that each is implemented in a separate, relatively small local region of the brain. First,
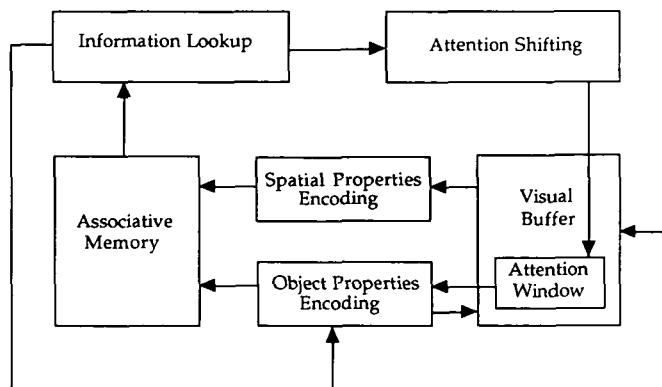
Information Lookup → Attention Shifting

Associative Memory ← Spatial Properties Encoding ← Visual Buffer

Object Properties Encoding ← Attention Window

Fig. 1 Six processing components that are used in visual object identification.

the input is represented in a structure we call the 'visual buffer' (see also Kosslyn et al., 1990). This functional structure corresponds to a set of retinotopically mapped areas in the occipital lobe; it segregates figure from ground and otherwise delineates the spatial organization of a stimulus. About half of the 32 cortical areas known to underlie vision in the macaque monkey are topographically organized (see Daniel and Whitteridge, 1961; Tootell et al., 1982; Van Essen, 1985; Felleman and Van Essen, 1991). There is ample evidence that the human occipital lobe contains such areas, from effects of brain damage (e.g. Holmes, 1918) and PET studies (Fox et al., 1986). Moreover, Kosslyn et al. (1993) used PET to show that topographically organized regions of visual cortex could be activated by visual mental imagery.

More information is present in the visual buffer than can be processed in detail, and hence only some of it is selected for additional processing. An *attention window* apparently selects a region of the visual buffer, and the activation within the region is allowed to be processed deeper in the system. The existence of such a mechanism was demonstrated, for example, by Sperling, (1960), who showed that subjects can covertly shift attention over an after-image. In addition, Treisman and Gelade (1980) found that when subjects search for a target in a field of distractors, they must look at each item separately if the target is defined by a conjunction of features (as is true for letters); in such situations, subjects apparently covertly shift attention to each item in an array, one at a time. Moran and Desimone (1985) provided some insight into the neural mechanisms that underlie this sort of attention in the macaque. They charted the receptive fields of neurons that responded selectively to specific stimuli, and then trained monkeys by reinforcing them only if the stimulus appeared in a certain quadrant of the receptive field. After training, the neuron would still begin to respond when the stimulus was placed in the non-reinforced part of the field, but would quickly cease activity. The pulvinar and anterior cingulate appear to be involved in this process of fixating visual attention (see LaBerge and Buchsbaum, 1990; Posner and Petersen, 1990).

The contents of the attention window are sent along two

major cortical pathways from the occipital lobe. In monkeys, one runs ventrally to the inferior temporal lobe whereas the other runs dorsally to the posterior parietal lobe. Numerous ablation and single-cell recording studies have shown that these pathways have different functions. For example, Ungerleider and Mishkin (1982) (see also Pohl, 1973) trained monkeys to discriminate between objects or spatial locations to find food. If an animal's inferior temporal lobes were removed, it had great difficulty making the object discrimination, but little difficulty making the spatial discrimination, and vice versa if an animal's posterior parietal lobes were removed. Such findings suggest that the ventral pathway encodes *object properties*, such as shape, colour and texture, whereas the dorsal pathway encodes *spatial properties*, such as location, size and orientation (see also Maunsell and Newsome, 1987). Neurophysiological studies have produced converging evidence for these inferences: neurons sensitive to shape and colour are often found in the temporal lobe (e.g. Desimone et al., 1984; Gross et al., 1984; Maunsell and Newsome, 1987), whereas neurons sensitive to location and motion are often found in the inferior parietal lobule (Hyvarinen, 1982; Andersen et al., 1985). Some of these results suggest that visual memories may actually be stored in the inferior temporal lobes (see, for example, Miyashita and Chang, 1988; for a review, see Kosslyn, 1994). Haxby et al. (1991) and Sergent et al. (1992a) report PET data that suggest that the occipital-temporal junction or middle temporal gyrus structures may be the human analogue of the ventral system. These structures were active during tasks involving object recognition (Sergent et al., 1992a) and face matching (Haxby et al., 1991; Sergent et al., 1992a).

The two processing streams must converge on an *associative memory* structure, where the input is matched to stored properties that are associated with objects. This system is a network that stores associations between both modality-specific and amodal representations, including the name of the object, its category and so on (see Chapter 8 of Kosslyn and Koenig, 1992). Associative memory is a cortical long-term storage structure, which should be distinguished from the processes that actually cause new information to be stored (which rely on hippocampal and related medial temporal structures; see Squire, 1987). However, we do not assume that there must be a distinction between the process that makes comparisons and the structure that is operated upon; rather, it is possible that the structures and processes are intermingled, as in neural network models. In such models, the comparison process corresponds to a pattern of activation within the structure itself. Moreover, it is possible that associative memory should be divided into more specialized memory stores, such as distinct 'semantic' and 'episodic' memories (Tulving, 1972) or 'category' and 'exemplar' memories; such finer divisions have proven difficult to defend in the cognitive psychology literature (see, for example, McKoon et al., 1986), and we need not make a commitment to them here (although as more findings are reported, such distinctions probably will be warranted).

We have at least three reasons to infer that such an associative memory exists. (i) In many circumstances, both object properties and spatial properties are used to identify a stimulus. (Various types of tomatoes, for example, are characterized in large part by differences in their size.) Hence, the two must be associated in memory. (ii) The mere fact that people can report from memory where objects are located (e.g. the locations of furniture in their living-rooms) is evidence that the two sorts of information have been conjoined in the brain. (iii) There is evidence that pathways project from the dorsal and ventral systems to dorsolateral prefrontal cortex (Goldman-Rakic, 1987). Indeed, cortex in the vicinity of the principal sulcus of the monkey (area 46) appears to function as a spatial 'working memory' structure (*see* Goldman-Rakic, 1988; Wilson *et al.*, 1993). However, the literature on amnesia (e.g. Squire, 1987) suggests that the frontal lobes are not the site of long-term memory storage. Based on the imagery results of Kosslyn *et al.* (1993), we speculate that the angular gyrus and/or parts of area 19 may play critical roles in implementing associative memory.

The sort of purely bottom-up processing just discussed, which is driven by properties of the stimuli, is apparently sufficient to identify an object in ideal circumstances. But such processing may fail if the object projects an unusual shape or in other conditions (e.g. occlusion, low luminance, etc.), in which case the initial input will not match representations stored in the ventral system very well. In such situations, we conjecture that bottom-up processing serves to formulate an hypothesis about the object's identity, and subsequent top-down processing is used to evaluate (i.e. confirm or refute) this hypothesis by collecting additional information about the stimulus (*see*, for example, Gregory, 1966, 1970; Lowe, 1987*a,b*; Ullman, 1989; Kosslyn *et al.*, 1990). People apparently do not search aimlessly over stimuli, but rather look for distinctive parts or properties of expected objects; this strategy relies on using knowledge to direct search for additional visual properties, which then are encoded and compared with those expected to be part of the hypothesized object (*see* Neisser, 1967; Yarbus, 1967; Loftus, 1972; Luria, 1980).

The frontal lobe clearly plays a role in such knowledge-guided search. Not only do the frontal eye fields (area 8) play a role in directing attention (Robinson and Fuchs, 1969), but many researchers have found that damage to the frontal lobe disrupts systematic visual search (e.g. Luria, 1980). In addition, Petersen *et al.* (1988) report a PET study in which a region in dorsolateral prefrontal cortex was active both when subjects looked up from memory ways in which objects can be used, and when they looked up properties (whether animals are dangerous) of named objects. Furthermore, the frontal lobes project to the superior parietal lobule (area 7, in particular) and superior colliculus, both of which appear to play critical roles in shifting attention (*see* Posner and Petersen, 1990; Haxby *et al.*, 1991; Corbetta *et al.*, 1993). Finally, the search process may involve holding information about the locations (actual and expected) of objects and parts

in a temporary 'working memory' and the frontal lobes clearly play a role in such processing (*see* Goldman-Rakic, 1987; Jonides *et al.*, 1993; Wilson *et al.*, 1993).

In summary, the emerging picture of object identification is as follows. Input from the eyes is first organized in a visual buffer, and key properties are selected by an attention window. The information passed through the attention window is sent to two systems; the object-properties encoding system (implemented in the inferior temporal lobe in monkeys) recognizes the shape, colour and texture of the object or part, and the spatial-properties encoding system (implemented in the posterior parietal lobe in monkeys) registers the location, size and orientation of the object or part of it. All of this information is then sent to associative memory, and the representation of the object whose properties are most consistent with those in the input is most strongly activated. However, if the match is not very good, a tentative object identification is treated as an hypothesis, and the frontal lobes access key properties associated with the object. The frontal lobes, superior parietal lobes and subcortical structures shift attention to the location where a distinctive property should be, and new information is encoded. If the property belongs to the object, and has the correct spatial properties, the hypothesis may be confirmed (for more details, *see* Chapter 3 of Kosslyn and Koenig, 1992; Kosslyn, 1994).

To some, the predictions that follow from this theory may not seem convincing because the theory has so many components. However, there is broad agreement in cognitive neuroscience that any complex activity, such as visual object identification, relies on a *system* of interacting components (*see*, for example, Posner and Petersen, 1990; Churchland and Sejnowski, 1992; Kosslyn and Koenig, 1992). Thus, the fact that we predict that numerous areas should be activated follows from the nature of the phenomenon we chose to study. Moreover, note that the predictions are not independent; we are predicting activation of a *set* of areas, not activation of individual, isolated areas; predictions about a pattern of active areas are a necessary concomitant of any attempt to study a system of processes. The specific components and the principles of interaction we hypothesize are grounded in large part on research with animal models, as noted above. Hence, the predictions are motivated well enough to warrant empirical testing—and to the extent that the results confirm the predictions, this is evidence that the approach (as well as the theory) is worth taking seriously.

In addition, the fact that many areas are predicted to be activated during picture identification does not imply that the theory will be difficult to disprove. Indeed, at first blush, results of Warrington *et al.* (*see* Warrington and Taylor, 1973, 1978; Warrington and James, 1991) appear to disprove one part of the theory. These researchers report that patients with damage to the frontal lobes did not have difficulty identifying objects that were depicted from unconventional points of view or that were presented as silhouettes; these results appear to be inconsistent with our claim that the frontal lobes play a major role in top-down processing during object

identification. However, patients with damage to the posterior right hemisphere were impaired in these tasks, which is consistent with our claim that the locations of additional properties must be encoded during the process of hypothesis testing. Kosslyn (1987) argues that the right parietal lobe plays a critical role in encoding metric spatial information, and Kosslyn (1994) argues that this sort of information is used to reconstruct depth information from pictures and to integrate encodings made during consecutive, but separate, eye fixations.

This apparent disconfirmation of our theory may not be fatal, however, for at least two reasons. First, more than one strategy (i.e. combination of processes) typically can be used to perform a task. Thus, patients with frontal lesions may identify objects seen from non-canonical viewpoints using a different strategy from that used by intact subjects. The failure to find a deficit does not imply that the frontal lobes are not used in normal processing in this situation. Secondly, it is possible that these patients did, in fact, have a deficit, but the deficit was apparent in response times, not error rates [Warrington and James (1991) did not report response times]. According to our theory, knowledge-guided search should be disrupted when the frontal lobes are damaged. Such damage would increase the time necessary to name unfamiliar shapes, but would not necessarily eliminate one's ability to complete the task accurately; even inefficient searching will often eventually allow one to encode additional information needed to evaluate an hypothesis. If so, then patients with frontal lobe damage should require more time to identify non-canonical shapes than canonical shapes, and this difference should be much larger than that found for normal subjects; whereas normal people would use knowledge to guide search, these patients would rely on inefficient and haphazard search strategies. This prediction has not been tested, but the mere fact that it can be formulated implies that it is premature to reject our theory.

We designed a PET experiment to evaluate the emerging theory as a whole, and also to test the specific hypothesis that the frontal lobes are used when one identifies objects seen from unfamiliar perspectives. Subjects participated in three conditions. In the first, they saw a series of objects depicted from a canonical perspective and heard a word when they saw each picture; on each trial, they decided whether the word accurately named the picture. In the second condition they saw a series of objects depicted from an unusual point of view and heard a word when they saw each picture; on each trial, they decided whether the word accurately named the picture. In the third condition, subjects saw random patterns of line segments, and heard a word when they saw each pattern; they now simply pressed a pedal when they heard the word (this was a baseline condition).

We chose a name-verification task, instead of a naming task, for a number of reasons. First, in a name-production task it is difficult to ensure that subjects produce the correct responses (particularly in the non-canonical condition); if subjects make many errors, it would not be clear what

sorts of processing the PET results reflected. Secondly, the cognitive literature on picture identification (reviewed by Kosslyn and Chabris, 1990; Kosslyn, 1994) is based primarily on name-verification paradigms. A major reason for this choice is that a name-verification paradigm allows researchers to control the relation between the type of name (e.g. at a superordinate, subordinate or 'entry' level) and the picture. We wanted to ensure that all names were at the 'entry' level, even for the non-canonical pictures. In order to do this in a naming paradigm, subjects must memorize the list of acceptable names in advance, which introduces a host of other factors that could affect the results (e.g. how well the names are memorized, how easily they are recalled, subjects' strategies in preparing to guess certain names). Thus, a name-verification paradigm was better suited to our purposes.

Methods
Subjects
Twelve males volunteered to participate as paid subjects. The mean age of the subjects was 22 years 3 months, with a range of 18 years 7 months to 28 years 4 months. The subjects all reported having good vision and being in good health. Eleven subjects were right-handed and one was left-handed. All subjects were unaware of the specific purposes or predictions of the experiment at the time of testing. This experiment was approved by the Harvard University and Massachusetts General Hospital Institutional Review Boards, and all subjects gave informed consent.

Procedures and equipment
We created four versions of each of 27 pictures of common objects. As illustrated in Fig. 2, two versions depicted the object from a canonical viewpoint, and two depicted it from a non-canonical viewpoint. The pictures were first drawn by hand, then digitized in black and white (i.e. 1 bit per pixel) at 75 dots per inch (~30 dots per centimetre) using a Microtek Scanmaker 600ZS scanner to create bit-mapped files for presentation on a Macintosh computer. The bit maps were then resized to make them all ~6.35 cm along their longest axis, or ~7° of visual angle from the subjects' viewpoint (which was ~52 cm from the computer screen).

In addition, we created 27 'patterns' by arbitrarily rearranging parts of each of the objects. The resulting drawings were meaningless configurations of line segments, and were equated with the drawings of objects for total number of pixels and size. In addition, we rotated these patterns three times, in 90° increments, thereby creating four versions of each.

The words were recorded on the Macintosh computer, using a Farallon Computing MacRecorder sound digitizer, sampling at 11K Hz, controlled by the SoundEdit program. We recorded the 'entry-level name' of each picture (Jolicoeur *et al.*, 1984; *see also* Kosslyn and Chabris, 1990), and two names of similarly shaped objects (as judged by three of the
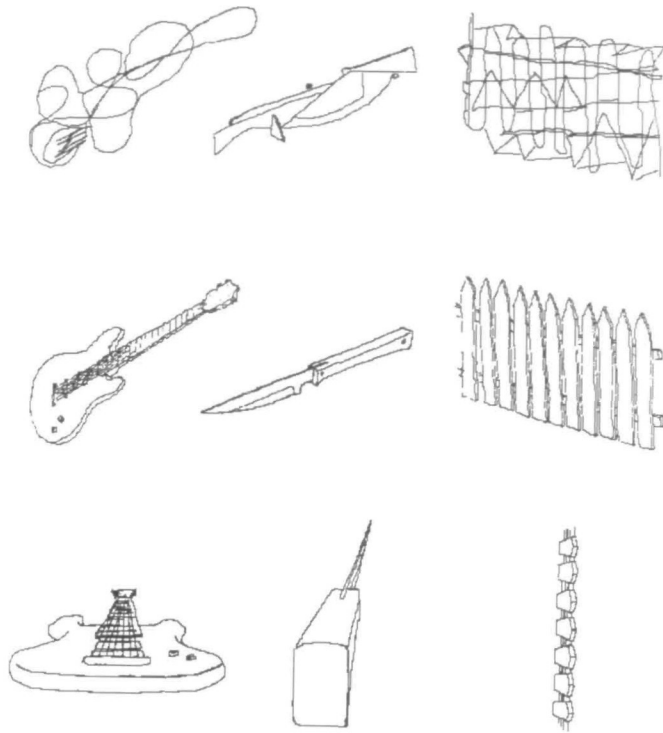
Fig. 2 Illustrations of stimuli. The top row are baseline patterns created from the pictures; the second row are pictures of objects seen from a canonical point of view; the bottom row are pictures of objects seen from a non-canonical point of view.

authors) to be used as distractors. We determined the entry-level name of a picture by testing an additional 20 Harvard University undergraduates. We asked these subjects simply to name objects that were depicted from a canonical viewpoint, and took the most frequent name as the entry-level name, provided that it was produced by at least 80% of the subjects. Table 1 presents a list of the objects and the corresponding words for 'yes' and 'no' trials.

An additional set of stimuli was created for practice trials. The practice set included eight trials; each of four objects was shown twice, once as a 'yes' trial and once as a 'no' trial. The practice trials were created the same way that the test trials were created, and neither the pictures nor words used in practice trials appeared in test trials.

Six versions of the experiment were prepared, each including three conditions: baseline, canonical and non-canonical pictures. All subjects received the baseline condition first; half then received the canonical followed by the non-canonical condition, and half received the non-canonical followed by the canonical condition. The 27 objects were divided into three groups of nine, which were assigned to the three conditions; thus, for a single subject, no object occurred in more than one condition. Counterbalancing ensured that within our group of 12 subjects, each object appeared equally often in each condition, and the canonical and non-canonical conditions were presented equally often in the two possible orders discussed above.

Table 1 *The names of the 27 objects that appeared in the experiment. The first column in each group presents the correct name for the object ('yes' trials). The next two columns list the two distractor words used for each object ('no' trials)*

Object (correct name)	Distractor 1	Distractor 2
Group 1 objects		
Airplane	Dragonfly	Helicopter
Bottle	Club	Rolling-pin
Flower	Lamp	Lollipop
Guitar	Tennis racket	Shovel
Hat	Muffin	Egg
Ring	Pacifier	Belt
Shirt	Bathmat	Pillow
Snake	Hose	Rope
Table	Awning	Bookshelf
Group 2 objects		
Apple	Heart	Bean bag
Cake	Bottlecap	Stadium
Car	Eraser	Footstool
Doll	Teddy bear	Cushion
Fence	Railroad tracks	Hedge
Golf club	Rake	Broom
Gun	Hairdryer	Telescope
Knife	Nail-file	Icepick
Rug	Bedspread	Shawl
Group 3 objects		
Dog	Bear	Cotton candy
Glasses	Wheelbarrow	Scissors
Lettuce	Wig	Cloud
Pen	Cigar	Paintbrush
Sandwich	Kite	Bowtie
Saw	Doorstop	Spatula
Shoe	Boat	Trough
Stove	Washing machine	Pinball machine
Watch	Tire	Collar

Within each condition, each object occurred four times, twice as a 'yes' trial (in which the picture was paired with the correct name) and twice as a 'no' trial (in which the picture was paired with the name of a similar-looking distractor). Naturally, in the baseline condition, the subjects were neither able nor requested to determine whether or not the word was a correct name. Furthermore, each of the two variants of each drawing appeared once in a 'yes' trial and once in a 'no' trial.

PET procedure

Subjects were tested individually. Each subject was told that the experiment was being conducted to study how people identify objects in different circumstances, and the procedure was described to him. After he filled out an informed consent form, he was fitted with a thermoplastic custom molded face mask (TRUE SCAN, Annapolis, Maryland, USA). The subject then entered the scanner, where his head was aligned relative to the cantho-meatal line. After mounting the mask

so that the subject's head was stabilized, we attached nasal cannulae to a radiolabelled gas inflow and hooked an overlying face mask to a vacuum. We took several transmission measurements with an orbiting-rod source prior to scanning. Following this, the experiment began. We took 20 measurements on each PET run; the first three measurements each were 10 s in duration, and the following 17 each were 5 s in duration. We began the scan by starting the camera acquisition program (which measured residual background from previous studies); 15 s later, presentation of the stimuli began, and the subject started to perform the task. Administration of $[^{15}O]CO_2$ gas began 15 s after this, and scanning ended after an additional 60 s, at which point the gas was stopped. The concentration of the delivered $[^{15}O]CO_2$ was 2800 MBq/l at a flow rate of 2 litres per minute and diluted by mixture with room air so that the measured peak countrate from the brain was 100 000 to 200 000 events per second.

The PET machine was a GE Scanditronix PC4096 15-slice whole-body tomograph, which we used in its stationary mode (*see* Rota-Kops *et al.*, 1990). The camera produced contiguous slices 6.5 mm apart (centre-to-centre; the axial field was equal to 97.5 mm); the axial resolution was 6 mm full width at half maximum (FWHM). The PET machine was in a suite built specifically for this purpose, and the same conditions were used for all testing—the lights were dimmed and there was no conversation or other distracting noise.

Task procedure

After a subject was placed in the scanner, he read instructions for the baseline task, which were presented on a sheet in front of the computer screen. He began the task only after he reported understanding the instructions. The stimuli were presented on a Macintosh Plus computer, using a version of the MacLab program (Costin, 1988) that was modified to present sounds as well as pictures. The computer recorded responses and response times via a foot-pedal device to which the keyboard was mounted.

Baseline task.

Baseline task. Each trial began with a blank screen, which appeared for 200 ms, followed by the auditory presentation of a word. Once the sound of the word ended, a pseudo-random pattern of lines appeared. The subjects were instructed simply to press a pedal as quickly as they could when the pattern appeared on the screen. We asked the subjects to alternate foot pedals during the baseline task because the pedals were used to make yes/no judgements in the other tasks, and the two judgements appeared equally often; we wanted to ensure that subjects pressed each pedal in approximately the same proportion as in the other tasks. Each pattern appeared exactly once in each block of nine trials, and the pattern was never paired with the word that named the object from which it was formed, nor was it

paired with either of the two distractors for that object. For example, the shape that was based on the drawing of the guitar was never paired with the word 'guitar' nor with 'tennis racket' or 'shovel' (which were the distractors for 'guitar'). We always administered the baseline task first so that the subjects would not be aware of the object identification task; we feared that if the other tasks were performed before the baseline, the subjects would try to find the named objects in the baseline patterns (even though they were not actually there). Debriefing after the experiment revealed that none of the subjects could recall making such efforts.

Picture identification tasks. After completing the baseline task, each subject read the instructions for either the canonical or non-canonical task, depending on his counterbalancing group. As soon as the subject reported understanding the instructions, he completed eight practice trials in which four objects appeared twice, once paired with a word that correctly named the object (a 'yes' trial) and once paired with a word that did not name the object (a 'no' trial). Following this, the subjects began the actual task (15 s before scanning began). As in the practice trials, a blank screen appeared for 200 ms, followed by the auditory presentation of a word. Immediately after the word, a picture of a common object appeared. The subject was told to press the pedal under his right foot if the word named the object correctly, and the pedal under his left foot if it did not. The subject was told to respond as quickly and accurately as possible. After each response, the blank screen returned and a new trial began. Each subject completed at least the full 36 trials of each condition; if there was still time remaining in the scan, the trials of the same condition were repeated (in the same order). All subjects reached a second trial cycle in the canonical condition, and all but two reached it in the non-canonical condition. The subjects completed more canonical trials than non-canonical ones (57 versus 53, respectively) $[F(1, 11) = 9.20, P < 0.01]$.

This procedure was repeated for the third condition of the experiment (either the canonical or the non-canonical task, depending on the subject's counterbalancing group). The third condition was identical to the second except that different objects were illustrated, and the objects were portrayed from a different viewpoint (canonical or non-canonical, as the case may be). Each condition began 10 min after the conclusion of the previous condition, enough time for most of the radioactivity from the previous condition to be washed out (^{15}O has a half-life of ~2 min).

PET image reconstruction

The images of relative blood flow were computed on the basis of scans 4–16, which were summed after reconstruction. The terminal count rates were between 100 000 and 200 000 events per second. Using radial artery cannulation, we have found that integrated counts over periods up to 90 s are a

linear function over the flow range of 0–130 ml/min/100 g. Hence, an arterial line was not necessary to ensure that data can be characterized in units of flow relative to the whole brain (*see* Kosslyn *et al.*, 1993).

We reconstructed the images using a measured attenuation correction and a Hanning-weighted reconstruction filter; the filter was set so that there was an 8 mm in-plane spatial resolution (FWHM). In reconstructing the images, we also corrected for effects of random coincidences, scattered radiation and counting losses that result from dead time in the camera electronics.

We pooled each slice of the scan data across all behavioural conditions, and then identified the coordinates of midline structures across all slices. We estimated the parameters of the midsagittal plane by applying a least squares procedure to these coordinates. We then re-sliced the images parasagittally at 5.1 mm intervals. The brain surface of a 10.2 mm parasagittal slice was outlined by hand at the 50% threshold level (nominal). If necessary, missing data from the surfaces of the parasagittal emission slices were filled in from more complete sagittal transmission images.

We then transformed the PET data to Talairach coordinates by deforming the 10 mm sagittal planes specified in the Talairach and Tournoux (1988) brain atlas until we obtained the best match to a standard template (with 'best' being defined in a least-squares sense; *see* Alpert *et al.*, 1993). Using this procedure, we estimated the locations of the frontal pole, occipital pole, vertex, anterior commissure (AC), posterior commissure (PC) and the tilt angle. The locations of these structures and the midsagittal plane allowed us to compute the piecewise linear transformation to Talairach coordinates. We evaluated the quality of this transformation not only by examining the standard errors of the parameters, but also simply by comparing visually the manually drawn brain surface to the atlas contour. In addition, we projected a computerized version of the Talairach *et al.* (1967) atlas onto the transformed data; this procedure allowed us to confirm that the transformed image conformed to the outlines of structures and key features in the atlas.

Results and discussion
Behavioural analysis
We began by analysing the response times and error rates from the canonical and non-canonical conditions. If the subjects had to collect additional information to evaluate the objects seen from non-canonical viewpoints, they should have required more time. And in fact, the subjects required a mean of 846 ms for the non-canonical trials compared with a mean of 657 ms for the canonical trials [$F(1, 11) = 5.74$, $P < 0.03$]. In addition, the subjects committed a mean of 8.9% errors for the non-canonical trials compared with a mean of 4.6% for the canonical trials [$F(1, 22) = 4.87$, $P < 0.04$]. These results confirm that we successfully manipulated the ease of encoding the pictures.

One could argue that any differences between the non-

canonical and canonical conditions resulted because subjects spent more time processing in the non-canonical condition. However, when we examined the total viewing time, we found that such a difference did not occur. Although the subjects did require more time in the non-canonical condition, as we predicted if additional processing was necessary, they performed fewer of these trials. We summed the total time each subject spent processing in each condition (i.e. we summed the time from stimulus presentation to response for each trial) and found that the subjects were processing a total of 37.09 s, on average, in the canonical condition compared with a mean of 44.55 s in the non-canonical condition, which was not a significant difference ($t < 1.15$). Thus, any differences between the two tasks are not due to the sheer amount of integrated brain computing time.

PET statistical analysis
The mean concentration in each slice for each run was specified as an area-weighted sum, which we adjusted to a nominal value of 50 ml/min/100 g. We then scaled and smoothed the images with a two-dimensional Gaussian filter (20 mm wide, FWHM). We next pooled the images over subjects for each condition, and a baseline image was subtracted from a test image; these images were subtracted within subjects. The results were images of the mean differences, standard deviations and a t-value for each pixel. Each t-statistic image was then submitted to a statistical parametric mapping (SPM) analysis (Friston *et al.*, 1991). This procedure produces 'omnibus subtraction images', which are images of standardized normal deviates. This method allows us to be confident that chance activations are not reported as significant. Two major elements of the analysis address the inherent problem of multiple comparisons: (i) the smoothness of the image, and (ii) the number of pixels implied by a hypothesis. The latter element deals specifically with multiple comparisons by Bonferonni-like correction, requiring a much higher threshold to achieve significance than would be indicated when considering the smoothness of the images alone. We measured image smoothness using the method of Friston *et al.* (1991) and found it to be 14.2 mm. We next used Friston *et al.*'s (1991) formula to adjust the threshold for statistical significance to account for multiple comparisons and the smoothness of subtraction images. We adjusted the significance value by multiplying it by the number of pixels in the region of activation. When we had predictions, we were justified in using one-tailed tests; when we did not have predictions, we used two-tailed tests.

For example, in a midbrain slice ~20 mm above the AC–PC line, there are ~2000 pixels, and a Z-score of ~3.5 is required to reach the 5% significance level if the investigator has no a priori hypothesis about the localization of the activation within the slice. On the other hand, if the investigator's hypothesis is so fine-grained that activation at a particular pixel is hypothesized, this method reduces to the equivalent of a single t-test. In our work, which features

theory-driven, a priori hypotheses, the 5% significance level is often reached with Z-scores of 2.5–3.0, depending on whether the activation is lateralized and whether the hypothesis refers to a large region, like the inferior parietal lobe, or a small region, like the pulvinar. In these analyses, Z-scores of 2.5–3.0 correspond to *t*-scores that without correction for multiple comparisons approach or achieve the *P* < 0.001 level. For additional details on the PET methods and analysis procedures, *see* Kosslyn *et al.* (1993).

Finally, we must note that the technique we used has only a limited capacity for distinguishing among a large contiguous territory of activation, multiple discrete areas of activation or a confluence of activation emanating from a single or multiple sources. This is a potential problem because the theory being tested specifies that a number of different areas should be activated, some of which happen to be adjacent to one another. We attempted to deal with this problem as follows: (i) we selected a threshold for the SPM image at a level corresponding to the Z-score needed for statistical significance in the largest area within a region of contiguous areas that appeared to be activated (this was conservative, because in SPM larger regions generally require higher Z-scores for significance); (ii) this produced an image that delineated the entire territory exceeding the criterion for significance; (iii) in cases where such a territory of significant activation extended into multiple adjacent structures, all are reported in the text (although only the locus of the single most-activated pixel is reported in the tables or figures); (iv) the Z-score we report for each such region is the value of the maximally activated pixel falling within the boundaries of that structure. We also provide, in the tables, descriptions of these territories based on *x*, *y* extent, and area within each slice.

Non-canonical–canonical

We began by subtracting the patterns of blood flow in the canonical picture condition from the patterns of blood flow in the non-canonical picture condition. This subtraction allowed us to discover whether the frontal lobes and other structures that are putatively recruited during top-down hypothesis testing were in fact activated (*see above*). The results of this SPM analysis are presented in Table 2 and the location of the single pixel with the greatest activation difference for each region of activation is illustrated in Fig. 3. Although the SPM analysis sometimes specified pixels that were very close together, in this and subsequent analyses we only report foci that were at least 14 mm apart (in three-dimensional space), which was the width of the smoothing function we used. As noted below, some of the regions of activation extended over several areas.

Table 2 *Coordinates (in millimetres, relative to the anterior commissure) and* P-values for *regions in which there was more activation in the non-canonical picture condition than in the canonical picture condition. Regions are presented from posterior to anterior. Seen from the rear of the head, the* x *coordinate is horizontal (with positive values to the right), the* y *coordinate is in depth (with positive values anterior to the anterior commissure) and the* z *coordinate is vertical (with positive values superior to the anterior commissure). Areas labelled with* * *or* † *are part of a territory of significant activation, described below*

	x	*y*	*z*	Z-score	*P*
Left hemisphere regions					
Area 18	−35	−73	4	2.84	0.01
Superior parietal	−25	−68	44	2.46	0.05
Middle temporal*	−44	−60	−4	4.64	0.00005
Dorsolateral prefrontal (area 47)	−22	40	−8	3.02	0.01
Right hemisphere regions					
Area 17	15	−93	−4	3.20	0.004
Area 18	22	−86	8	3.65	0.0009
Angular gyrus	33	−67	28	2.64	0.05
Inferior temporal†	52	−58	−8	4.16	0.0002
Inferior parietal†	38	−56	44	3.79	0.002
Dorsolateral prefrontal (area 9/46)	35	15	28	3.25	0.02

*The territory of significant activation included part of area 19 and the inferior temporal gyrus, and extended over two slices, $z = -8$ and $z = -4$. On slice −8, the territory covered 280 mm² and had an average Z-score of 3.24; on slice −4, the territory covered 722 mm² and had an average Z-score of 3.70. †The territory of significant activation included over portions of area 19, the middle temporal gyrus, and the fusiform gyrus, and extended over three slices, $z = -12$, $z = -8$ and $z = -4$. On slice −12, the territory extended over 449 mm² and had an average Z-score of 3.16; on slice −8, the territory extended over 455 mm² and had an average Z-score of 3.40; on slice −4, the territory extended over 195 mm² and had an average Z-score of 2.93.

Looking up expected properties and locations. As expected, the dorsolateral prefrontal cortex was activated, in both the left and right hemispheres. Specifically, a region extending over areas 9 and 46 was activated in the right hemisphere, and area 47 was activated in the left hemisphere. At least some of these areas presumably are involved in looking up stored information, as opposed to merely storing input from the ventral and dorsal systems for a brief period of time. According to our theory, the dorsolateral prefrontal cortex plays a key role in looking up information about a part or a property that the system is seeking in the visual input.

The following other areas were also activated more when subjects evaluated non-canonical pictures than when they evaluated canonical pictures (in all but two cases, activation was in areas we predicted):

Shifting attention. We found activation in the left superior parietal lobe, which apparently plays a role in shifting attention (Corbetta *et al.*, 1993). However, we did not find more activation in the frontal eye fields (area 8) for the non-canonical pictures than the canonical pictures.

Visual buffer. If hypotheses are tested by encoding additional information, then we should have found more activation in cortex used to organize visual information (which implements the visual buffer). And in fact, areas 17 and 18 were activated 18 bilaterally, 17 in the right hemisphere only.

Encoding new parts or properties. If additional parts or properties are encoded to test a hypothesis, then we should have found activation in the inferior temporal and/or middle temporal lobes (*see* Haxby *et al.*, 1991; Sergent *et al.*, 1992*a*). A large activated region that included both of these areas was activated in both hemispheres. This region also included activation in the right fusiform gyrus; all three of these areas were activated in tasks that required subjects to form visual mental images (Kosslyn *et al.*, 1993). These areas may be the human analogues to the monkey ventral system (*see* Levine, 1982).

Encoding new spatial relations. At the same time that additional parts and properties are encoded, we also expect spatial relations among them to be encoded. As predicted, we found massive activation in the inferior parietal lobe in the right hemisphere. As discussed previously, Warrington *et al.* have found that this area is the principal region where damage impairs performance when non-canonical pictures are identified. We found no selective activation in the left inferior parietal lobe, however, which is also consistent with the findings of Warrington and James (1991).

Associative memory. Finally, we expected greater activation in associative memory during the non-canonical picture condition. We predicted that associative memory should be activated not only because it is the locus where a name and visual inputs are compared, but also because
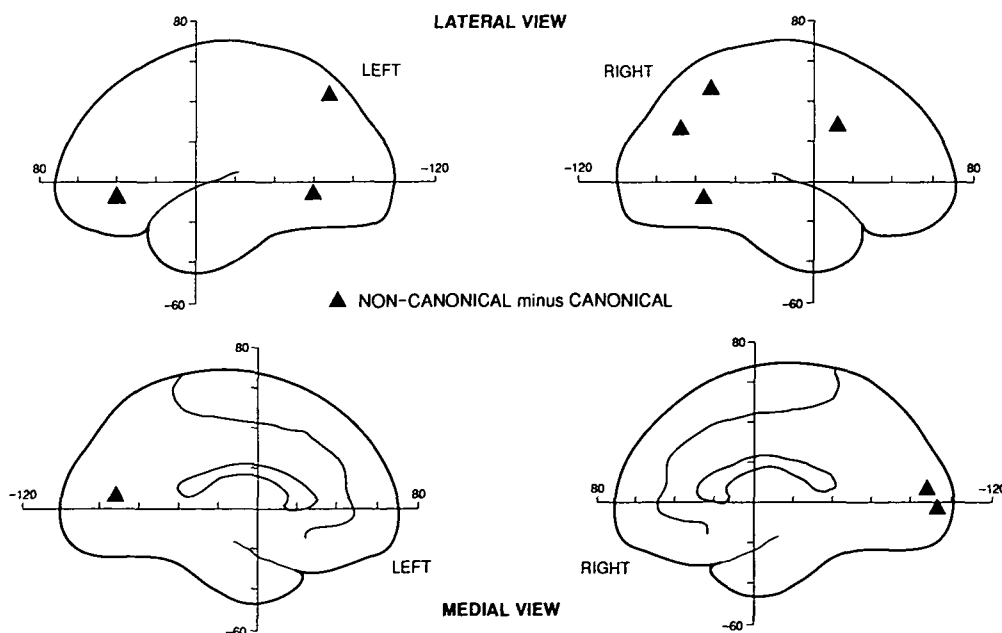

Fig. 3 The results when blood flow in the canonical condition was subtracted from blood flow in the non-canonical condition. The left and right cerebral hemispheres, seen from lateral and medial views. The tick marks on the axes specify 20 mm increments relative to the AC. Points illustrate the location of the single most activated pixel in a region; *see* text for description of regions.

information in associative memory must be activated to direct top-down search. Based on our earlier imagery results (Kosslyn et al., 1993), we hypothesized that parts of area 19 and the angular gyrus might be involved in this processing. In fact, we found more activation in the right angular gyrus and bilateral activation in area 19, however the coordinates of the activation in area 19 were very inferior; indeed, this activation was a portion of a continuous region spanning across portions of the fusiform, inferior and middle temporal gyri (see the bottom of Table 2). Sergent et al. (1992a) also found increased activation in area 19 in an object categorization task, when blood flow in a line gratings baseline task was subtracted. This activation was not evident when blood flow in a face gender discrimination task was subtracted from blood flow in a face identification task, which may indicate that both tasks involve associative memory. In fact, when blood flow in the gratings task was subtracted from blood flow in the gender discrimination task, Sergent et al. (1992a) again found activation in area 19.

Canonical–baseline

The design of this experiment also allowed us to distinguish between two variants of our overall theory. The *directed processing theory* posits that top-down processing mechanisms only come into play when necessary. In contrast, the

reflexive processing theory posits that all systems run virtually all the time, and that top-down processing will be engaged even if it is not essential. According to this theory, although such processing is 'automatic' (in the sense that decoding the meaning of a word in a Stroop task is automatic), it nevertheless is effortful (cf. Shiffrin and Schneider, 1977). Both variants of the theory predict the same pattern of activation when pictures seen from non-canonical points of view are identified. However, the theories make different predictions about activation when objects seen from canonical points of view are identified.

Within the framework illustrated in Fig. 1, the directed processing theory would lead us to expect only the bottom-up processes to be activated when one views canonical pictures; we would not expect activation of the dorsolateral prefrontal areas used to look up stored information or of parietal and subcortical structures used to shift attention. In contrast, the reflexive processing theory would lead us to expect the same patterns of activation for objects seen from canonical and non-canonical perspectives, including the areas used in top-down processing.

Hence, we next subtracted the blood flow in the baseline condition from the blood flow in the canonical condition. The results of this SPM analysis are presented in Table 3 and illustrated in Fig. 4. We first consider the results that were predicted by both the directed processing and reflexive processing variants of the theory.

Table 3 *Coordinates (in millimetres, relative to the anterior commissure) and P values for regions in which there was more activation in the canonical condition than in the baseline condition. Regions are presented from posterior to anterior. Seen from the rear of the head, the x coordinate is horizontal (with positive values to the right), the y coordinate is in depth (with positive values anterior to the anterior commissure) and the z coordinate is vertical (with positive values superior to the anterior commissure). Areas labelled with * are part of a territory of significant activation, described below*

	x	y	z	Z-score	P
Left hemisphere regions					
Area 17	−6	−79	0	3.36	0.003
Area 19	−26	−78	20	3.44	0.006
Superior parietal*	−24	−71	40	3.08	0.003
Fusiform	−33	−22	−12	2.33	0.07 (NS)
Frontal eye fields	−13	21	40	3.13	0.008
Right hemisphere regions					
Area 18	24	−87	4	2.92	0.01
Area 19	22	−81	40	2.72	0.02
Superior parietal	8	−73	44	3.20	0.007
Frontal eye fields†	4	29	32	2.87	0.006
Anterior cingulate	7	34	24	3.42	0.05
Dorsolateral prefrontal (area 10)	24	58	0	2.83	0.02

*The territory of significant activation included the inferior and superior parietal lobe, and extended over slice z = 40, extending over 208 mm² and had an average Z-score of 3.00. †The frontal eye fields and anterior cingulate were separated by a slightly smaller distance than our criterion; however, visual inspection suggested that these were in fact distinct areas of activation. NS = not significant.

Visual buffer. We found activation in area 17 in the left hemisphere, and area 18 in the right. Such activation presumably reflects processing to organize the figure during encoding.

Encoding parts or properties. We found only non-significant trends for the left middle temporal gyrus and left fusiform gyrus to be activated in this task. It is possible that the patterns in the baseline task activated these regions to some extent (people may have sometimes seen objects in these patterns, the way that they see faces in clouds), and hence the difference in blood flow was not as large as expected.

Encoding new spatial relations. At the same time that shape is encoded, relevant spatial properties should be encoded. We found activation in a large portion of the left parietal lobe, which included part of the inferior parietal lobe. We did not, however, find activation in the right inferior parietal lobe; this finding is consistent with the idea that the right parietal lobe encodes information used to reconstruct three-dimensional structure, which might not have been necessary to process the canonical pictures.

Associative memory. Area 19 was activated in both hemispheres, but the angular gyrus was not. However, different portions of area 19 appear to have been activated in this comparison than in the previous one.

The reflexive processing theory also predicts activation in the following additional systems:

Looking up expected properties and locations. Although we found activation in dorsolateral prefrontal cortex, it was in right hemisphere area 10. The result was predicted by neither variant of the theory: although dorsolateral prefrontal cortex was activated, it was not activated in the same way as when non-canonical pictures were evaluated.

Shifting attention. We found activation in the frontal eye fields (area 8) and in the right superior parietal lobe; we also found activation in a large left-parietal region that included the superior parietal lobes. We would expect these areas to be active in this task if they are involved in shifting attention to the location of an expected part or property. In addition, we found activation in the anterior cingulate in the right hemisphere; this area apparently is involved in attention (*see* Posner and Petersen, 1990).

Non-canonical–baseline

Finally, we also analysed the blood flow in the non-canonical condition when the blood flow in the baseline condition was subtracted. This was of interest in part because we failed to find activation in the frontal eye fields or anterior cingulate in our previous analysis of the data from the non-canonical
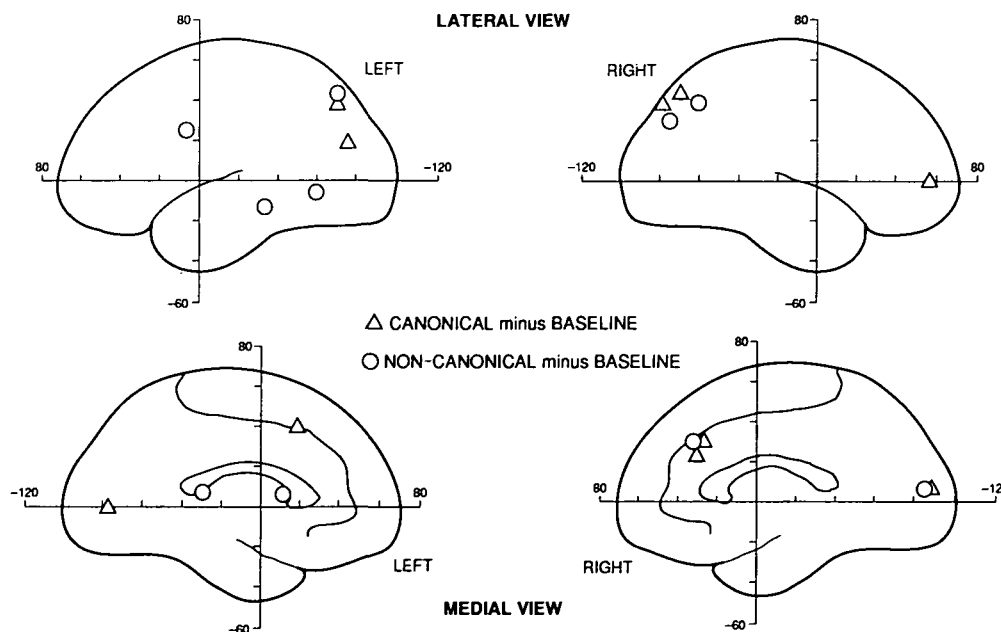

Fig. 4 The results when blood flow in the baseline condition was subtracted from blood flow in the canonical condition (triangles) or from the non-canonical condition (circles). The left and right cerebral hemispheres, seen from lateral and medial views. The tick marks on the axes specify 20 mm increments relative to the AC. Points illustrate the location of the single most-activated pixel in a region; *see* text for description of regions.

condition, but predicted both areas to be involved in top-down processing. We found that both areas were activated when subjects evaluated canonical pictures, as expected by the reflexive processing theory, and thus subtracting the blood flow in the canonical condition from the blood flow in the non-canonical condition may have concealed activation of these areas when subjects identified non-canonical pictures.

The results of this analysis are presented in Table 4 and illustrated in Fig. 4. As is evident, the right frontal eye field was activated, as was the left thalamus, both of which are involved in shifting attention (but there was only a trend for activation in the right anterior cingulate). We also found activation in Broca's area, for reasons that are not clear. The remaining results are consistent with those found by subtracting activation from the canonical pictures instead of the baseline condition. Note, however, that the left hemisphere caudate activation seems to be in white matter according to the Talairach and Tournoux (1988) atlas, but is clearly within the caudate in the Talairach et al. (1967) atlas.

Overall pattern of results

One question that arises in evaluating our results stems from the large number of theoretically motivated, a priori predictions. One might wonder about the likelihood of similar results occurring by chance. To address such concerns we calculated the probability of chance results under simplifying, but very conservative assumptions. The basic idea of the

calculation was to compute a likelihood product of independent probabilities, one factor per hypothesis, while accounting for important spatial correlations and multiple comparisons. Neglecting the correlations and multiple comparisons, an upper limit of the likelihood is given by the product of the P-values in Tables 2, 3 or 4 for the a priori hypotheses; this produces a vanishingly small number. We refined this calculation by assuming that spatial correlations within hypothesized regions were important but correlations were negligible between hypothesized regions. These spatial correlations are three-dimensional, arising from the finite spatial resolution of the PET scanner, interpolations in the stereotactic image reslicing steps and additional image smoothing. The correlations and multiple comparisons are already accounted for by the SPM method, in a two-dimensional way (by slice). But, in order to make a more conservative three-dimensional correction, we weighted each probability by the number of resolution elements in the hypothesized volume (accounting for in-slice smoothing and axial resolution, an effective resolution volume was taken as $20 \times 20 \times 6$ mm = 2400 mm^3). On the basis of these conservative criteria we estimate the probability of the ensemble of results in the non-canonical–canonical analysis occurring due to chance as $P = 0.000058$, approximately. Similarly, we estimate the corresponding probability for the canonical–baseline results as approximately one in one million, and for the non-canonical–baseline results as approximately one in 100 million.

Table 4 *Coordinates (in millimetres, relative to the anterior commissure) and P-values for regions in which there was more activation in the non-canonical picture condition than in the baseline condition. Regions are presented from posterior to anterior. Seen from the rear of the head, the x coordinate is horizontal (with positive values to the right), the y coordinate is in depth (with positive values anterior to the anterior commissure) and the z coordinate is vertical (with positive values superior to the anterior commissure). Areas labelled with * or † are part of a territory of significant activation, described below*

	x	y	z	Z-score	P
Left hemisphere regions					
Superior parietal*	−33	−73	44	3.97	0.0005
Middle temporal†	−40	−59	−4	4.31	0.0005
Fusiform†	−35	−35	−12	3.71	0.02
Thalamus	−20	−28	12	3.00	0.01
Inferior frontal (Broca's area)	−33	7	24	2.32	0.05
Caudate	−13	11	8	3.71	0.02
Right hemisphere regions					
Area 18	25	−85	4	4.55	0.00005
Angular gyrus	31	−78	28	2.51	0.07 (NS)
Inferior parietal	44	−60	40	3.09	0.01
Frontal eye fields	7	34	32	2.62	0.009

*This territory of activation included the superior and inferior parietal lobe, and extended over slice 44. It covered an area of 377 mm^2 and had an average Z-score of 3.42. †This territory included portions of area 19, the middle temporal gyrus, inferior temporal gyrus and fusiform gyrus, and extended over three slices (with z-values of −12, −8 and −4). On slice −12, the territory covered 267 mm^2 and had an average Z-score of 3.11; on slice −8, the territory covered 682 mm^2 and had an average Z-score of 3.05; on slice −4, the territory covered 793 mm^2 and had an average Z-score of 3.50. NS = not significant.

The overall pattern of results also allows us to rule out the possibility that the non-canonical condition was simply more difficult, and hence there was more blood flow in general during it. If top-down processing is actually needed in a specific task, and hence such processing runs to completion prior to a response, then there should be greater activation in areas that encode additional visual information when non-canonical pictures are seen instead of canonical ones. Specifically, we would expect particularly great activation in the fusiform, inferior and middle temporal gyri, which putatively store visual information and match input to these representations. If the reflexive version of the theory is correct, as suggested in the analysis of the canonical–baseline results, then we would not expect differences in the other areas.

To test this hypothesis, for each subject we measured the level of activation for each region found to be significantly more activated in the non-canonical condition than in the canonical condition (*see* Table 2). Regions of interest having a 5 pixel radius were centred on the pixel having the highest value within each region. The percentage of change from canonical to non-canonical conditions was then calculated for each region, for each subject, and an analysis of variance was performed on these values. The interaction between condition and regions of interest was $F(14,11) = 1.61$, $P < 0.08$. We then specified a contrast in which activation in the areas noted above plus area 19 (all of these sites were part of a single territory of activation) was compared with activation in all other areas. This contrast revealed that the change in activation in the non-canonical minus canonical subtraction was significantly greater in these areas (mean change = 7.4%) than in the other areas (mean change = 4.5%), $F(1,11) = 3.86$, $P = 0.05$. Thus, the difference in activation cannot be ascribed to a general effect of task difficulty—which would have predicted only a main effect between conditions.

One unexpected aspect of the overall results is that different regions of the dorsolateral prefrontal cortex were prominent in different comparisons. Reviewing the findings, mapped to the Talairach and Tournoux (1988) atlas, it is clear that we have evidence for activation of a number of different areas; these areas of activation are not contiguous, and are in many instances quite far apart from one another. It is possible that different parts of dorsolateral prefrontal cortex work together as a network; alternatively, these regions may be specialized for specific functions, such as accessing information in associative memory prior to directing attention versus holding shape and spatial information in working memories (*see* Goldman-Rakic, 1987). Although communicating the findings in terms of coordinates has the benefits of universality and nomenclature neutrality, it does not indicate whether different activations were truly in distinct areas. Given the size and multiplicity of functional subunits within this region, we have designated the specific Brodmann area corresponding to each site of activation.

The fact that different regions of dorsolateral prefrontal

cortex were activated in the non-canonical and canonical conditions allows us to rule out the possibility that activation of dorsolateral prefrontal cortex occurred merely because we provided a name along with each picture. If providing the name alone was responsible for the activation of the dorsolateral prefrontal region, the same region should have been activated in both conditions. Because names were presented in both conditions in exactly the same way, any differences between the canonical and non-canonical conditions cannot be ascribed to this aspect of the paradigm.

In some cases, if activation was observed when we performed the non-canonical–canonical subtraction but was not present when we performed the canonical–baseline subtraction, it was also present when we performed the non-canonical–baseline subtraction (i.e. for left middle temporal, left inferior temporal, right angular gyrus and the right inferior parietal cortex). However, this was not always observed; the results of the comparison of non-canonical and canonical conditions sometimes were not evident when we compared the non-canonical and baseline conditions, even though there was no difference between the canonical and baseline conditions.

There are a number of reasons why this 'non-transitivity' may have occurred. First, as noted above with respect to the dorsolateral prefrontal cortex, region labels can be misleading. In some cases, coordinates that are far apart lie in the same region (as specified by the taxonomy we used), and are thus labelled identically. Nevertheless, it is likely that each of these large regions in fact corresponds to a set of distinct areas. For example, activation of different portions of the dorsolateral prefrontal cortex was evident in the different comparisons, which explains why non-canonical–canonical revealed activation in areas 9 and 46 (right) and 47 (left), canonical–baseline revealed activation in area 10 (right), but non-canonical–baseline revealed no significant differences. Similarly, the fact that activation is reported in area 19 in the non-canonical–canonical comparison as well as in the canonical–baseline comparison, but not in the non-canonical–baseline comparison would seem to suggest a violation of transitivity. However, on examination of the coordinates reported for 'area 19' in these two subtractions, one realizes that they in fact correspond to two different regions that are separated by 48 mm on the z-axis alone. The standard taxonomy of functional areas obviously leaves much to be desired.

Secondly, as will be discussed in greater detail below, the baseline condition should not be interpreted as a task that preserves every aspect of the canonical condition except for the identification of a picture. Thus, we should not expect it to show perfect transitivity with respect to the other two conditions. For example, consider the results from the right hemisphere regions of inferior temporal, middle temporal and fusiform, which were significantly activated in the non-canonical condition when compared with canonical condition, but not by the non-canonical condition when compared with the baseline. A more detailed examination of the data revealed

that for each of these three regions, there were positive, but non-significant, tendencies for 'inhibition' (negative activation), which suggests that the baseline task may have taxed these areas slightly more than did the canonical task.

Thirdly, the tables list only differences that were significant according to the SPM analyses, which are inherently conservative. Thus, the patterns of activation by region may preserve the expected transitive pattern across conditions, but some of the specific comparisons may have just missed significance—and were thus omitted. Visual inspection of the blood flow maps suggests that this was the case for some of the apparent violations of transitivity.

Finally, it is worth noting that the SPM technique selects the coordinates of the pixels with maximal differences between images, and these pixels may vary depending on the images being compared. Thus, the location of the maximal difference in a given region between images 1 and 2 may not be the same as the location of the maximal difference when the same region is compared in images 2 and 3, which would lead to the appearance of a lack of transitivity.

These observations raise an important methodological point about the design of activation studies (for similar observations, see Sergent et al., 1992b). In fact, the logic of inference differs for the different comparisons we made. First, we used a simple subtraction logic when the blood flow in the baseline condition was subtracted from the other conditions. Following the convention established by the St Louis group (see, for example, Petersen et al., 1988), we have compared the two test conditions to a baseline condition that not only involved different stimuli, but also a different task. We have assumed, as is common in this field, that the elementary encoding and response processes used in the baseline task were also used in the picture evaluation tasks, and hence by subtracting patterns of blood flow in the baseline condition we removed the contributions of those processes to the results of the picture evaluation conditions. However, as was evident at the turn of this century, when the 'fallacy of pure insertion' was first noted, this need not be true (see, for example, Külpe, 1895, pp. 406–22; Woodworth, 1938, pp. 309–10; Boring, 1950, pp. 148–9; Luce, 1986, pp. 212–17). Rather, it is possible that subjects scan the baseline patterns differently than actual pictures, organize them differently, match them (often without success, presumably) to stored representations differently and respond differently than they do when meaningful pictures are evaluated.

The other logic of inference we employed is rooted in the 'additive factors' method (as developed by Sternberg, 1969). The additive factors method relies on preserving the nature of the task, and manipulating a variable that selectively taxes a specific type of processing used in it. For example, to study the process of scanning a list of items in short-term memory, Sternberg (1969) varied the length of the list—which engendered more or less such scanning. This logic eliminates the potential problems of comparing two different tasks, which may be performed using different strategies. It is a

truism that a good experiment should vary only one thing at a time, and it would seem that an easily interpretable PET experiment should include comparison conditions that preserve the nature of the task itself. Our comparison between the canonical and non-canonical conditions has this property: only the projected viewpoint of the picture was varied. To the extent that there are discrepancies between the various comparisons reported above, we urge caution in interpreting those in which the baseline results were subtracted, since we have no proof that 'pure insertion' did, in fact, occur.

Conclusions

This study produced two notable findings. First, we found clear evidence that the predicted pattern of activation was in fact present. Indeed, when examining the difference between the non-canonical and canonical picture conditions, we found activation in virtually all of the additional areas we predicted, and we found activation in only two unexpected regions. These findings are of interest in part because our predictions were, for the most part, based on results from animal models. Secondly, we found that portions of dorsolateral prefrontal cortex are activated when one identifies an object seen from a non-canonical perspective. This result was predicted by our claim that these regions are used to look up stored information in the course of top-down search.

Our findings appear to converge well with those of Sergent et al. (1992a) in a study of face and object processing. Sergent et al. (1992a) presented subjects with black-and-white photographs of common objects, faces of well-known people, and line gratings of various orientations. Activation engendered by an object classification task (in which subjects were to determine whether stimuli were natural or man-made) was compared with activation in a baseline task with the line gratings stimuli. Furthermore, the face recognition and object classification conditions were compared in order to identify areas of activation specific to each.

The regions that proved significantly more active in the object-classification task than in the gratings task were the left inferior temporal gyrus, left fusiform gyrus, left middle temporal gyrus, left middle occipital gyrus (area 19), left superior parietal lobe, bilateral supramarginal gyrus and gyrus rectus. We also found evidence for the involvement of the left fusiform and left middle temporal gyri in the recognition of objects. We attribute the middle temporal activation [found by Sergent et al. (1992a) and by us] to the encoding of parts and properties and to visual memory activation. We also found activation of the left area 19, which we propose as an associative memory structure. We did not, however, find activation in the left inferior temporal lobe. In addition to the areas reported by Sergent et al. (1992a) for object classification, we also found activation in right extrastriate and right superior parietal lobe, as well as in areas thought to be involved in attention-shifting and top-down search: the frontal eye fields, anterior cingulate and dorsolateral prefrontal cortex, all bilaterally.

Sergent *et al.*'s (1992*a*) face and object condition subtractions revealed shared activation of the left fusiform, left middle temporal area and gyrus rectus, while the left inferior temporal and left middle occipital lobes were activated solely by the object classification task. They propose that the left fusiform and left middle temporal area are involved in visual analysis of all shapes, and they suggest that activation in the gyrus rectus may reflect accessing of visual memory. Consistent with their inferences, in neither of our picture naming tasks did we obtain evidence of increased blood flow in most of the areas Sergent *et al.* (1992*a*) posit to be involved exclusively in face identification. The one exception is the right fusiform gyrus, which was active in the non-canonical condition relative to the canonical condition. The function of this area may be to extract the perceptual invariants of a particular face or the non-accidental properties of an object (*see* Lowe, 1987*a,b*). Warrington and Taylor (1978), as noted by Sergent *et al.* (1992*a*), report that lesions to the right fusiform gyrus cause patients to have difficulty recognizing pictures presented from a non-standard viewpoint.

In short, the emerging picture of high-level vision is remarkably consistent with what one would expect based on findings in non-human primates and analyses of information-processing requirements. With the coarse outline in place, we can now begin to specify in detail the different individual processes that are used in specific types of tasks.

Acknowledgements

References

Alpert NM, Berdichevsky D, Weise S, Tang J, Rauch SL. Stereotactic transformation of PET scans by nonlinear least squares. In: Uemura K, Lassen NA, Jones T, Kanno I, editors. Quantification of brain function: tracer kinetics and image analysis in brain PET. Amsterdam: Elsevier, 1993: 459–63.

Andersen RA, Essick GK, Siegel RM. Encoding of spatial location by posterior parietal neurons. Science 1985; 230: 456–8.

Boring EG. A history of experimental psychology. 2nd ed. New York: Appleton-Century-Crofts, 1950.

Churchland PS, Sejnowski TJ. The computational brain. Cambridge (MA): MIT Press, 1992.

Corbetta M, Miezin FM, Shulman GL, Petersen SE. A PET study of visuospatial attention. J Neurosci 1993; 13: 1202–26.

Costin D. MacLab: a Macintosh system for psychology labs. Behav Res Meth Instrum Comput 1988; 20: 197–200.

Daniel PM, Whitteridge D. The representation of the visual field on the cerebral cortex in monkeys. J Physiol (Lond) 1961; 159: 203–21.

Desimone R, Ungerleider LG. Neural mechanisms of visual processing in monkeys. In: Goodglass H, Damasio AR, editors. Handbook of neuropsychology, Vol. 2. Amsterdam: Elsevier, 1989: 267–99.

Desimone R, Albright TD, Gross CG, Bruce CJ. Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 1984; 4: 2051–62.

De Valois RL, Morgan H, Snodderly DM. Psychophysical studies of monkey vision. 3. Spatial luminance contrast sensitivity tests of macaque and human observers. Vision Res 1974a; 14: 75–81.

De Valois RL, Morgan HC, Polson MC, Mead WR, Hull EM. Psychophysical studies of monkey vision. 1. Macaque luminosity and color vision tests. Vision Res 1974b; 14: 53–67.

Felleman DJ, Van Essen DC. Distributed hierarchical processing in primate cerebral cortex. [Review]. Cereb Cortex 1991; 1: 1–47.

Fox PT, Mintun MA, Raichle ME, Miezin FM, Allman JM, Van Essen DC. Mapping human visual cortex with positron emission tomography. Nature 1986; 323: 806–9.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. Comparing functional (PET) images: the assessment of significant change. J Cereb Blood Flow Metab 1991; 11: 690–9.

Goldman-Rakic PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Mountcastle VB, Plum F, editors. Handbook of physiology, Sect. 1, Vol. V, Pt. 1. Bethesda (MD): American Physiological Society, 1987: 373–417.

Goldman-Rakic PS. Topography of cognition: parallel distributed networks in primate association cortex. [Review]. Annu Rev Neurosci 1988; 11: 137–56.

Gregory RL. Eye and brain: the psychology of seeing. New York: McGraw-Hill, 1966.

Gregory RL. The intelligent eye. London: Weidenfeld and Nicholson, 1970.

Gross CG, Desimone R, Albright TD, Schwartz EL. Inferior temporal cortex as a visual integration area. In: Reinoso-Suárez F, Ajmone-Marsan C, editors. Cortical integration. New York: Raven Press, 1984: 291–315.

Haxby JV, Grady CL, Horwitz B, Ungerleider LG, Mishkin M, Carson RE, et al. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc Natl Acad Sci USA 1991; 88: 1621–5.

Holmes G. Disturbances of vision by cerbral lesions. B J Ophthalmol 1918; 2: 353–84.

Hummel JE, Biederman I. Dynamic binding in a neural network for shape recognition. Psychol Rev 1992; 3: 480–517.

Hyvarinen J. Posterior parietal lobe of the primate brain. [Review]. Physiol Rev 1982; 62: 1060–129.

Jolicoeur P, Gluck MA, Kosslyn SM. Pictures and names: making the connection. Cognit Psychol 1984; 16: 243–75.

Jonides J, Smith EE, Koeppe RA, Awh E, Minoshima S, Mintun MA. Spatial working memory in humans as revealed by PET [see

comments]. Nature 1993; 363: 623–5. Comment in: Nature 1993; 363: 583–4.

Kosslyn SM. Imagery, mental. In: Adelman G, editor. Encyclopedia of neuroscience, Vol. 1. Boston: Birkhäuser, 1987; 12: 521–2.

Kosslyn SM. Image and brain: the resolution of the imagery debate. Cambridge (MA): MIT Press, 1994.

Kosslyn SM, Chabris CF. Naming pictures. J Visual Lang Comput 1990; 1: 77–95.

Kosslyn SM, Koenig O. Wet mind: the new cognitive neuroscience. New York: Free Press, 1992.

Kosslyn SM, Flynn RA, Amsterdam JB, Wang G. Components of high-level vision: a cognitive neuroscience analysis and accounts of neurological syndromes. Cognition 1990; 34: 203–77.

Kosslyn SM, Alpert NM, Thompson WL, Maljkovic V, Weise SB, Chabris CF, et al. Visual mental imagery activates topographically organized visual cortex: PET investigations. J Cogn Neurosci 1993; 5: 263–87.

Külpe, O. Outlines of psychology. New York: MacMillan, 1895.

LaBerge D, Buchsbaum MS. Positron emission tomographic measurements of pulvinar activity during an attention task. J Neurosci 1990; 10: 613–19.

Levine DN. Visual agnosia in monkey and in man. In: Ingle DJ, Goodale MA, Mansfield RJW, editors. Analysis of visual behavior. Cambridge (MA): MIT Press, 1982: 629–70.

Loftus GR. Eye fixations and recognition memory for pictures. Cogn Psychol 1972; 3: 525–51.

Lowe DG. Perceptual organization and visual recognition. Boston: Kluwer, 1985.

Lowe DG. Three-dimensional object recognition from single two-dimensional images. Artif Intell 1987a; 31: 355–95.

Lowe DG. The viewpoint consistency constraint. Int J Comput Vision 1987b; 1: 57–72.

Luce RD. Response times: their role in inferring elementary mental organization. New York: Oxford University Press, 1986.

Luria AR. Higher cortical functions in man. 2nd rev. ed. New York: Basic Books, 1980.

McKoon G, Ratcliff R, Dell GS. A critical evaluation of the semantic-episodic distinction. J Exp Psychol: Learn, Mem Cogn 1986; 12: 295–306.

Maunsell JHR, Newsome WT. Visual processing in monkey extrastriate cortex. [Review]. Annu Rev Neurosci 1987; 10: 363–401.

Miyashita Y, Chang HS. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 1988; 331: 68–70.

Moran J, Desimone R. Selective attention gates visual processing in the extrastriate cortex. Science 1985; 229: 782–4.

Neisser U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME. Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 1988; 331: 585–9.

Pohl W. Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. J Comp Physiol Psychol 1973; 82: 227–39.

Posner MI, Petersen SE. The attention system of the human brain. [Review]. Annu Rev Neurosci 1990; 13: 25–42.

Robinson DA, Fuchs AF. Eye movements evoked by stimulation of frontal eye fields. J Neurophysiol 1969; 32: 637–48.

Rota-Kops E, Herzog HH, Schmid A, Holte S, Feinendegen LE. Performance characteristics of an eight-ring whole body PET scanner. J Comput Assist Tomogr 1990; 14: 437–5.

Sergent J, Ohta S, MacDonald B. Functional neuroanatomy of face and object processing: a positron emission tomography study. Brain 1992a; 115: 15–36.

Sergent J, Zuck E, Lévesque M, MacDonald B. Positron emission tomography study of letter and object processing: empirical findings and methodological considerations. Cereb Cortex 1992b; 2: 68–80.

Shiffrin RM, Schneider W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychol Rev 1977; 84: 127–90.

Sperling G. The information available in brief visual presentations. Psychol Monogr 1960; 74 No. 498: 1–29.

Squire LR. Memory and brain. New York: Oxford University Press, 1987.

Sternberg S. The discovery of processing stages: extensions of Donders' method. Acta Psychol (Amst) 1969; 30: 276–315.

Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme, 1988.

Talairach J, Szikla G, Tournoux P, Prossalentis A, Bordas-Ferrer M, Covello L, Iacob M, Mempel E. Atlas d'anatomie stéréotaxique du télencéphale; études anatomo-radiologiques. Paris: Masson, 1967.

Tootell RBH, Silverman MS, Switkes E, De Valois RL. Deoxyglucose analysis of retinotopic organization in primate striate cortex. Science 1982; 218: 902–4.

Treisman AM, Gelade G. A feature-integration theory of attention. Cognit Psychol 1980; 12: 97–136.

Tulving E. Episodic and semantic memory. In: Tulving E, Donaldson W, editors. Organization of memory. New York: Academic Press, 1972: 381–403.

Ullman S. Aligning pictorial descriptions: an approach to object recognition. Cognition 1989; 32: 193–254.

Ungerleider LG, Mishkin M. Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJW, editors. Analysis of visual behavior. Cambridge (MA): MIT Press, 1982: 549–86.

Van Essen DC. Functional organization of primate visual cortex. In: Peters A, Jones EG, editors. Cerebral cortex, Vol. 3: visual cortex. New York: Plenum Press, 1985: 259–329.

Warrington EK, James M. A new test of object decision: 2D silouettes featuring a minimal view. Cortex 1991; 27: 370–83.

Warrington EK, Taylor AM. The contribution of the right parietal lobe to object recognition. Cortex 1973; 9: 152–64.

Warrington EK, Taylor AM. Two categorical stages of object recognition. Perception 1978; 7: 695–705.

Wilson FAW, Scalaidhe SP, Goldman-Rakic PS. Dissociation of object and spatial processing domains in primate prefrontal cortex [see comments]. Science 1993; 260: 1955–8. Comment in: Science 1993; 260: 1876.

Woodworth RS. Experimental psychology. New York: Henry Holt, 1938.

Yarbus AL. Eye movements and vision. New York: Plenum Press, 1967.