DOES VISUAL SUBORDINATE-LEVEL CATEGORISATION ENGAGE THE FUNCTIONALLY DEFINED FUSIFORM FACE AREA?

Isabel Gauthier
Yale University, New Haven, CT, USA

Michael J. Tarr
Brown University, Providence, RI, USA

Jill Moylan, Adam W. Anderson, Pawel Skudlarski, and John C. Gore
Yale University, New Haven, CT, USA

Functional magnetic resonance imaging was used to compare brain activation associated with basic-level (e.g. bird) and subordinate-level (e.g. eagle) processing for both visual and semantic judgements. We localised the putative face area for 11 subjects, who also performed visual matching judgements for pictures and aurally presented words. The middle fusiform and occipital gyri were recruited for subordinate minus basic visual judgements, reflecting additional perceptual processing. When the face area was localised individually for each subject, analyses in the middle fusiform gyri revealed that subordinate-level processing activated the individuals face area. We propose that what is unique about the way faces engage this region is the focal spatial distribution of the activation rather than the recruitment of the face per se. Eight subjects also performed semantic judgements on aurally presented basic- and subordinate-level words. The parahippocampal gyri were more activated for subordinate-level than basic-level semantic judgements. Finally, the left posterior inferior temporal gyrus was activated for subordinate-level judgements, both visual and semantic, as well as during passive viewing of faces.

INTRODUCTION

Recent neuroimaging studies have contributed to our understanding of part of the human visual system where objects appear to be coded as objects (rather than as simple features as in primary visual cortex). For instance, experiments comparing "objects" to "nonobjects" such as scrambled features or textures have identified an area in the lateral occipital gyrus that responds preferentially to any stimulus, novel or familiar, which has a clear three-dimensional interpretation (Kanwisher, Chun, McDermott, & Ledden, 1996; Kanwisher, Woods, Iacoboni, & Mazziota, 1997b; Malach et al., 1995; Schacter et al., 1995). Other neuroimaging studies have focused on comparisons between particular object categories, leading to a modular view of human inferotemporal cortex in which there are regions dedicated to the recognition of distinct object categories such as faces or letter strings. For example, inspired by the neuropsychological literature on prosopagnosia (face recognition deficits

---

143

following brain insults), several studies have compared faces to nonface objects. Results of these studies suggest that part of the middle fusiform gyrus is dedicated to face recognition (Haxby et al., 1994; Ishai, Ungerleider, Martin, Maisog, & Haxby, 1997; Kanwisher, McDermott, & Chun, 1997a; McCarthy, Puce, Gore, & Allison, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1996; Puce, Allison, Gore, & McCarthy, 1995; Puce, Allison, Spencer, Spencer, & McCarthy, 1997).

In a recent paper (Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997), we proposed that many studies have failed to control for a critical difference between faces and nonface objects: That of the level of categorisation at which such stimuli are typically recognised. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) presented evidence that familiar objects are first recognised at a level of abstraction referred to as the "basic" level (or "entry" level, when defined independently for each object as the level at which contact is made first in semantic memory, Jolicoeur, Gluck, & Kosslyn, 1984). Objects in different basic-level categories (e.g. chair or bicycle) generally differ in their parts and configuration and so are easily discriminable (Tversky & Hemenway, 1984). In contrast, recognising objects at the "subordinate" level (e.g. desk chair vs. arm chair) requires additional time and perceptual processing (Jolicoeur et al., 1984), relying more heavily on multiple perceptual dimensions such as shading, texture, colour, surface detail, pigmentation, and spatial arrangement of features (Bruce & Humphreys, 1994).

Within this framework, faces are typically recognised at a very subordinate level (the exemplar level, e.g. Jill's or Isabel's face, or "this particular individual" for an unfamiliar face) whereas most other objects are typically recognised at the basic level. Based on this observation, Gauthier et al. (1997) tested the possibility that this difference in categorisation level contributed to the activation obtained when faces and objects are compared. The hypothesis is that, especially when subjects *passively* view a stream of faces and nonface objects, they are more likely to process faces than objects at the subordinate level. This may be so for at least two reasons: The presentation of several objects exclusively from the same category may encourage such subordinate-level processing and a lifetime of experience recognising faces at the this level may lead subjects to do so automatically, regardless of the task.

In Gauthier et al. (1997), a picture was presented together with a basic-level or subordinate-level word and participants were asked to verify whether the picture matched the word. After removal of the contribution of semantic processing of the word using a double-subtraction (see the Semantic task described later in this paper), a region of the fusiform and inferotemporal cortex similar to that described in prior studies as face-sensitive was found to be selectively engaged by subordinate-level processing for nonface objects. However, a limitation of this finding is that each individual's putative face area occupies only a small portion of the group-defined face-sensitive region (Allison et al., 1994). It is therefore possible that face recognition and subordinate-level categorisation engage mutually exclusive areas within this region in each individual but that they average to the same area across a group of subjects. The present study was designed to address this limitation by measuring for the same individuals both the activation for passive viewing of faces minus objects and the activation for a manipulation of categorisation level for nonface objects. Again, we believe that the latter manipulation contributes to the faces minus objects effect.

Since our first study manipulating categorisation level, two neuroimaging studies have equated subordinate-level processing of nonface objects with face processing (Kanwisher et al., 1997a; McCarthy et al., 1997). Both studies led to the conclusion that the activation for faces could not be accounted for by a category-level manipulation. However, methodological issues may limit the strength of this interpretation. First, McCarthy et al. compared passive viewing of faces and flowers appearing on a continuously changing background of nonsense patterns or nonface objects. Both faces and flowers, when shown on a background of nonsense patterns, engaged the right fusiform gyrus. When compared to a baseline of nonface objects, faces, but not flowers, produced activation in this area. The authors concluded that faces are treated differently to other objects, in a specialised brain

area. However, flowers as a control stimulus class and passive viewing instructions may not be optimal for engaging subordinate-level processing. Unlike faces, which share common parts and configuration, most flowers have very distinctive and unique features, such as the number and shape of the petals. Face-like subordinate-level processing is likely to be engaged by those object categories that have been termed "homogeneous classes." These are defined as categories that share a configuration of features so that common points can be located on every member of the class; exemplars from homogeneous classes can be averaged together to produce another instance of the same category (Diamond & Carey, 1986; Rhodes & McLean, 1990). Thus, whereas a composite of faces progresses towards a prototype as the number of exemplars averaged together increases (Levin, 1996; Rhodes, Brennan, & Carey, 1987), the same effect cannot be obtained with flowers of various species in that they differ widely in appearance (e.g. tulip, rose, and daisy). As a second concern, passive viewing of exemplars from a single category in the context of other objects is less likely to automatically engage subordinate-level processing (unless perhaps with expert subjects). Finally, McCarthy et al. (1997) quantified the activation in the comparison with nonface objects by thresholding individual activation maps at a relatively high threshold, the same used for the comparison with nonsense patterns, and then counting activated voxels. Faces differ from common objects along more dimensions (conceptual category, homogeneity of the class, default level of categorisation, degree of expertise) than do flowers (conceptual category, homogeneity of the class). Therefore, the activation for faces compared to objects may be stronger than that for flowers. A lower threshold might reveal a peak of activation for flowers among objects in the putative face area. Thus, the selection of a high threshold on individual activation maps may produce an artificial floor effect that could compromise the interpretation of a task comparison. In contrast, the quantification method used in the present study, that of summing the percentage signal change or t-values for voxels over a very low threshold in a region of interest (ROI), does not have this limitation. This method

has been found to be less dependent on ROI size than voxel count and rather insensitive to the threshold value (Constable et al., 1998).

In a second study, Kanwisher, McDermott, and Chun (Kanwisher et al., 1997a) compared activation during "one-back" judgements for consecutive repetitions of identical stimuli for faces and hands and found significantly more activation for faces than hands as compared to a fixation baseline. In contrast to McCarthy et al. (1997), this study used a homogeneous class of stimuli and required subjects to process subordinate-level information. A face area ROI was also defined in each subject, and the signal change for other comparisons was evaluated in this ROI regardless of threshold. However, an important difference is that the one-back task includes only a small proportion of matching trials as compared to a task such as that used by Gauthier et al. (1997), which included 50% matching trials (see Dill & Fahle, 1998; Farell, 1985; for differences between same and different trials). In order to convince themselves that they are not seeing a repetition, subjects need only notice a single featural difference. On the other hand, in order to give a "same" response, subjects need to have searched more thoroughly for differences and found none: "Same" responses are thus more likely to recruit configural processing than are "different" responses. In addition, whereas Kanwisher et al. (1997a) found more activation for faces than hands in the face area, it may still be the case that the hand task activates the face area to a significant degree, because it requires subordinate-level processing. The fact that the hand task does not engage the face area as much as the face task could reflect a difference in difficulty based on within-class similarity (which cannot be equated across classes) or a difference in the subjects' level of expertise for each class, as well as possible preference of this area for faces.

To summarise, both McCarthy et al. (1997) and Kanwisher et al. (1997a) found activation for faces to be *stronger* than that obtained for nonface objects and rejected level of categorisation as an important factor in interpreting the specialisation of the putative face area. However, although categorisation level may not be sufficient to account for *all* of the specificity in the face area, it remains possible that

some of the activation in the face area, when faces are compared to objects at the basic level, may be due to subordinate-level processing (as it is engaged automatically for faces but not for objects). Thus, although categorisation level may be only one of the several organising principles for the inferior temporal cortex, it may be of somewhat greater importance in that it appears to play a role early in development (Rosch, 1976) and has universal influence (in contrast to the more idiosyncratic role of expertise with particular object categories). Subordinate-level processing could thus account for a coarse specialisation in the middle fusiform and inferior temporal gyri, while expertise (with faces or other categories) may further refine this specialisation.

The present study was designed to revisit the question of whether a significant part of the activation found in the face area can be attributed to subordinate-level processing. What is new, however, is that here the putative face area is defined functionally in *individual* subjects (Kanwisher et al., 1997a). Other manipulations from our earlier study remain relatively unchanged. As before, we isolate subordinate-level processing by manipulating the categorical level of a word (basic or subordinate) matched to identical objects. In Gauthier et al. (1997), a double-subtraction method was used to isolate this process, as it was important to control for the visually presented subordinate-level words being longer on average than the basic-level words. Here we used aurally presented words, so that a single-subtraction design was possible. Although we define the face area using passive viewing of faces minus objects, as in Kanwisher and colleagues' work, our approach varies in that we investigate not only the activation within this ROI, but also in the rest of the ventral temporal cortex and especially in ROIs bordering the face area. In contrast, many recent studies (Kanwisher et al., 1997a; Kanwisher, Tong, & Nakayama, 1998; Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 1998) investigating the response in the face area restrict their analyses to a small region of the fusiform gyrus generally defined using passive viewing of faces minus objects. Such an analysis would not distinguish, for instance, whether subordinate-level processing

leads to activation of a small magnitude but precisely centred on the face area or to a larger activation that is not as precisely focused.

A second goal was to explore further the neural basis of categorisation at different levels of abstraction. The preponderance of experiments investigating the inferotemporal cortex in terms of stimulus preferences may lead to a skewed picture of the organisation of this region. Crucially, implicit task manipulations may reside within stimulus manipulations (e.g. passive viewing of faces may imply a different level of categorisation than passive viewing of objects), whereas when the same stimulus is presented and the task is varied, any difference obtained must be attributed to the task manipulation. The importance of such endogenous mechanisms is well illustrated in experiments where attention is shifted between dimensions (e.g. colour vs. motion; Zeki, Watson, Friston, & Frackowiak, 1991) or between an identity and a location task (Haxby et al., 1994; Moscovitch, Kapur, Kohler, & Houle, 1995) with the same stimulus. It is our belief that experiments investigating the role of endogenous factors that recruit in different cortical areas are needed to understand the neural basis of visual object recognition. Here, we aim to test what extent an endogenous manipulation alone can activate the "face area". This will provide an estimate of the maximum possible contribution of the level of categorisation effect when comparing faces to objects (if an explicit endogenous manipulation cannot engage the face area, it is unlikely that an implicit manipulation can do so).

In addition to the primary goals stated, we hoped to pursue the hypothesis advanced in Gauthier et al. (1997), that early visual areas may be recruited by subordinate-level semantic judgements because subordinate words provide a more specific basis for detailed visual imagery. By using aurally presented words, the present study controls for the possibility that this effect in the original study was due to greater visual stimulation for subordinate-level words in that they were on average longer than the basic-level words.

As in Gauthier et al. (1997), we used an empirical definition of basic and subordinate levels, using object and name pairings that were selected by

Gauthier et al. so that name verification times were significantly slower at the subordinate level than at the basic level[1]. Moreover, semantic judgements on the same basic-level and subordinate-level words also showed a basic-level advantage in response times, although not as large as the advantage found for visual judgements (Gauthier, unpublished data). Indeed, some of the basic-level advantage in response time for picture naming and even name-verification may be due to nonperceptual factors such as name frequency, name length, and order of learning (Johnson, Paivio, & Clark, 1996). However, we believe that our object–name pairings require additional *perceptual processing* for subordinate judgements.

METHOD

Subjects

Fourteen neurologically normal subjects (all right-handed) took part in the study (approved by the Yale University Human Investigation Committee). Eleven subjects performed the Localiser and the Visual task (five males, six females) and eight performed the Semantic Task (five males, three females). Three subjects participated in all three tasks in the same session.

Materials and Procedure

The same 72 images of objects as in Gauthier et al. (1997) were used, except that words were presented aurally (see Appendix A). The words were recorded on a Macintosh computer by a female native English speaker using SoundEdit Pro Software (Macromedia, San Francisco). Here, the Semantic task was not designed as a control for the Visual task. Therefore, 38 basic-level words and 38 subordinate-level words, not directly matched with the words used in the Visual task, were used in the Semantic task (see Appendix B).

In the Visual task, participants performed name-verification judgements in which they decided whether a greyscale image of an object matched an aurally presented word (responding Yes or No by pressing one of two buttons). On each trial, a word was presented through headphones while an object appeared simultaneously and remained on the screen for 2750msec with an intertrial interval of 750msec. Nine 21.3sec epochs showed Basic and Subordinate trials in alternation (order counterbalanced between runs, 4 runs per subject) for a total of 192sec. Identical sets of images were shown during the Basic and Subordinate epochs, in a randomised order, and the only difference between the two conditions was the level of categorisation of the aurally presented word (either basic as in "bird" or subordinate as in "sparrow"). Seventy-two pictures were repeated an average of three times each during the experiment (an *average* of 1.5 times at each level, with either matching or nonmatching labels). In the Semantic task, the same design as in the Visual task was used, except that subjects were asked to close their eyes and judge whether each of the words presented aurally described an object that could move by its own power (stimulus onset asynchrony of 3500msec—identical to the Visual task). Words were randomised within each level and each word was repeated an average of 2.8 times during the experiment (there were four runs per subject). In order to define face-selective areas individually, a Faces minus Objects Localiser task was performed using 90 greyscale faces (all cropped in the same oval shape) as well as 90 greyscale images of common nonface objects, with no overlap with the subordinate-level lists used in Visual and Semantic tasks. In each run for this task, 26 faces or 26 objects were flashed, each for 750msec within 9 alternating epochs of 21.3sec. Stimuli were projected on a screen at the subjects' feet within a square region of approximately 4 × 4 degrees of visual angle. This resulted in each image being shown once in each run. Subjects were told to fixate on the centre of the screen during the entire run for both visual tasks, although there was no fixation cross on the screen

---

[1] Judgements were timed starting from onset of the presentation of the picture following the presentation of the word.

during the presentation of pictures. The experiments were conducted using RSVP software (Williams & Tarr, undated).

fMRI Scan Acquisition

Imaging was performed on a 1.5 Tesla Signa scanner (GE Medical Systems, Waukesha, WI), with Instascan echo planar imaging capabilities (Advanced NMR Systems, Wilmington, MA). A single shot, gradient echo, echo-planar pulse sequence was used to acquire 64 by 128 voxel images over a field of view of 20 by 40cm. Imaging parameters were T_E = 60msec, T_R = 1500msec, flip angle 60 degrees, slice thickness 7mm. Six contiguous axial-oblique slices were imaged during each repetition time interval. The image plane was aligned along the longitudinal extent of the temporal lobe, with the lowest slice capturing the inferior occipital gyrus and the lowest portion of the temporal poles. This slice orientation was chosen to maximise resolution in the fusiform gyrus (in-plane resolution = 3.125 × 3.125mm, through-plane resolution = 7mm). Each run produced 128 images per slice. Image data were corrected for motion using the SPM 96 software (Wellcome Department of Cognitive Neurology, London). Changes in image intensity were analysed on a voxel-by-voxel basis: Maps of t-values were created for each pair of conditions compared and corrected for a low-frequency drift in the signal with an estimated 2sec haemodynamic lag taken into account. In addition, the two images (per slice) that occurred at the beginning of each epoch were discarded because at this point, the rising signal for the starting task would be confounded with the declining signal for the task that just ended. The t-maps were superimposed on T1-weighted anatomical images of the corresponding slice. No statistical significance is attributed to these activation maps, which are then used as raw data for further analyses in ROIs.

For averaging purposes, anatomical landmarks were defined in T1-weighted axial-oblique and midline sagittal images of each subject (they consisted of, in the oblique plane, the outer edges of the brain, the optic chiasma, and the anterior edge of the cerebral aqueduct, and in the sagittal plane, the superior and inferior edges of the brain, the optic chiasma, and the most posterior point of the fourth ventricle). Functional and anatomical images were transformed by piece-wise linear warping in 12 brain subvolumes to register the results for each subject in a common coordinate system. The transformed functional maps were then combined across subjects using a median value for each voxel and thresholded to obtain composite functional maps.

Anatomically Defined Regions of Interest

For analyses on anatomically defined areas, regions of interest (ROIs) were drawn a priori on the standardised anatomical images (see Plate 2 of the colour section by clicking here), based on a comparison of the average and individual anatomy with several human brain atlases (H. Damasio, 1995; Duvernoy, 1991; Talairach & Tournoux, 1988). No ROIs were defined in the first and last standard slices because several subjects had no equivalent original slice and to avoid artefacts caused by the motion correction algorithm. Skew-corrected t-values were summed in each ROI (Skudlarski, Constable, & Gore, 1999), and normalised for ROI size and amount of activation in all ROIs in a given hemisphere: A value of 1 therefore indicates that an ROI has a density of activation that is identical to the mean activation in the hemisphere, and higher values signify higher-than-average densities. No probabilistic value is attributed to the t-maps or the density values in individual subjects. Rather, the statistical significance of effects is determined by their representation in the group sample.

RESULTS

Analyses on Anatomically Defined ROIs

Plate 2 (see colour section) shows the median activation maps for the subtractions of the Subordinate minus the Basic condition for both the Visual and

Table 1. *Mean Density of Activation for Each ROI in the Visual (N = 11) and Semantic (N = 8) Tasks*

	Visual		*Semantic*	
ROI (ROI No.)	*S–B*	*B–S*	*S–B*	*B–S*
amg (16)	1.22	.81	1.06	.98
antFG (11)	1.05	R(1.27) L (.79)	1.35	1.07
antITG (2)	.89	1.17	1.14	.82
antMTG (12)	1.19	1.60	1.27	1.04
cun (8)	.84	*2.07***	*1.05*	1.34*
hipp (14)	.83	1.38	1.02	1.41
infOG (5)	2.66**	.46	1.02	1.54*
latOG (9)	2.70**	.49	*1.15*	1.12
linG (7)	.85	*1.81***	*1.56*	1.13
midFG (10)	*1.56***	.86	1.26	1.24
parahippG (3)	1.28	.81	*1.29***	1.06
postFG (6)	2.20**	.67	1.32*	1.18
postITG (4)	R (.75) L (1.47**)	1.22	R(.81) L (1.41)	R(.77) L(1.69)
postMTG (13)	1.70**	1.19	1.44	1.12
supOG (15)	2.00**	.95	1.39	.97
tPol (1)	.96	1.05	1.05	1.10

Results are split by hemisphere only when the laterality effect was significant by paired *t*-tests, *P* < .05.
Significance only indicated for densities > 1. Italic with **: a priori tests, *P* < .05. Italic without
asterisk: a priori tests, n.s. **: post hoc, *P* < .01/number of post hoc tests. *: post hoc, *P* < .05. (df = 10
for the Visual task and 7 for the Semantic task).

S: subordinate; B: Basic; amg: amygdala; antFG: anterior fusiform gyrus; antITG: anterior inferior
temporal gyrus; antMTG: anterior middle temporal gyrus; cun: cuneus; hipp: hippocampus; infOG:
inferior occipital gyrus; latOG: lateral occipital gyrus; linG: lingual gyrus; midFG: middle fusiform
gyrus; parahippG: parahippocampal gyrus; postFG: posterior fusiform gyrus; postITG: posterior
inferior temporal gyrus; postMTG: posterior middle tempral gyrus; supOG: superior occipital gyrus;
tPol: temporal poles.

the Semantic tasks. Table 1 gives the mean density of activation for each ROI for each comparison.

Single-sample *t*-tests (Sheskin, 1997) were performed for each cell of the table to assess whether the mean density in each ROI was significantly higher than the average hemisphere activation density[2]. A priori hypotheses were formulated based on the results of Gauthier et al. (1997) and the corresponding tests were performed at the .05 level per comparison. These hypotheses concerned five ROIs for the Visual task, three expected to be associated with subordinate level (midFG, postITG, postMTG), and two with basic level (cun, linG, both visible in the composite images in

Gauthier et al.) and for ROIs expected to be associated with subordinate level for the Semantic task (cun, latOG, midFG, linG). For all other ROIs, the probability of committing a Type I error was minimised by dividing an α-level of .05 by the number of post hoc ROIs, independently for each task (for Visual Subordinate minus Basic, .05/13 = .0038, and Basic minus Subordinate, .05/14 = .0036; for Semantic Subordinate minus Basic, .05/12 = .0042 and Basic minus Subordinate, .05/16 = .0031). Effects that did not reach the corrected α-level but were significant at the .05 level are reported as such, so that they can form the basis of hypotheses for future studies.

___

[2] Because the two sides of each subtraction (Subordinate minus Basic and Basic minus Subordinate) are not independent for each task, densities significantly lower than 1 are not reported. In most cases where this happened, the density for the other side of the subtraction was significantly higher than 1.

Visual Subordinate Minus Basic

The Visual Subordinate minus Basic comparison revealed activation in the middle and posterior fusiform gyri (midFG and postFG), the left postITG, the entire occipital gyri (infOG, latOG, supOG) as well as the posterior middle temporal gyri (postMTG, see Plate 2 of the colour section by clicking here).

The activations in midFG and post ITG replicate the findings from the double subtraction in Gauthier et al. (1997), as can be seen in Fig. 1. This activation overlaps with that described in many studies as the group-averaged face sensitive area (Puce et al., 1995, 1996, 1997; Sergent et al., 1992). Below, we use the Faces minus Objects comparison to localise individual subject's face area. We then investigate whether the association between the face area and subordinate-level processing (Gauthier et al.) can be replicated in individually defined face areas (which vary in location from subject to subject). Activation in the occipital lobe (postFG and OG) was anticipated by the composite results for the single (Subordinate minus Basic) subtraction for the Visual task in Gauthier et al., although it did not reach significance in that study.

This difference in OG activation, stronger here than in the prior study, may be the major difference between the double- and single-subtraction designs. We address the extensive recruitment of the visual system by subordinate-level processing in the final discussion. These consistent results obtained by comparing subordinate-level to basic-level judgements with fMRI here and in our prior study may be contrasted to the results of a similar comparison with PET by Kosslyn, Alpert, and Thompson (1995), who found only a single area in the ventral pathway (left BA 19) that was more active for Subordinate minus Basic visual judgements.

The postMTG activation most likely reflects the recruitment of the secondary auditory cortex by the longer subordinate-level words (similar, albeit nonsignificant activation is found in the homologous comparison for the Semantic task).

Visual Basic Minus Subordinate

Whereas Gauthier et al. (1997) found no significant activation for basic *over* subordinate visual judgements (in contrast to the additional activation obtained for subordinate over basic), the subject-
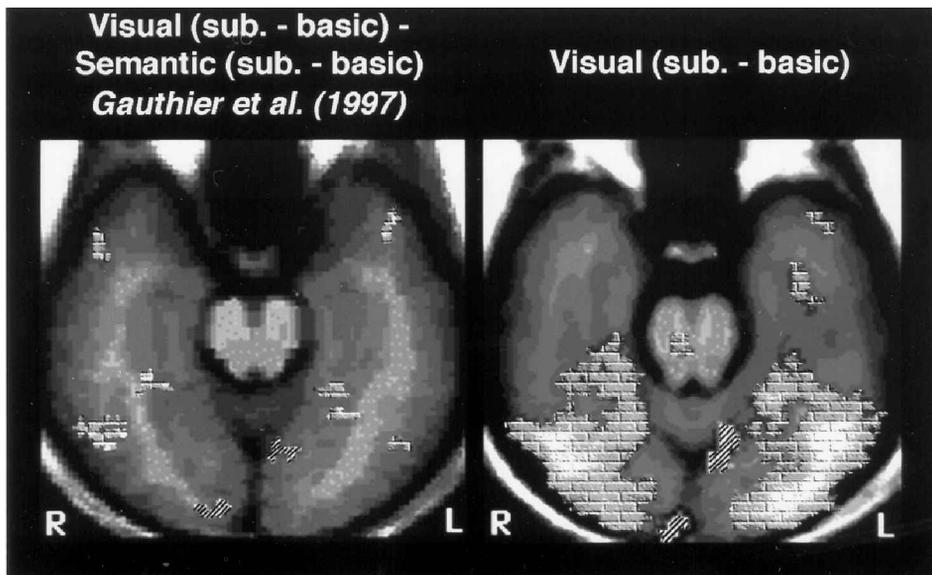
Fig. 1. *Comparison of slice 3 in the present study for the Visual task and the results from the double subtraction in Gauthier et al. (1997). The horizontal brick texture shows the result of the indicated subtractions.*

average composite map for the double subtraction suggested some activity in presumptive retinotopic visual cortex. A strikingly similar pattern is found here in the Visual single subtraction (Basic minus Subordinate) in both the cuneus and lingual gyri (see Fig. 1). These two areas are the only ones that survive our criterion for statistical significance. Thus, despite the differences in design between the two experiments (single vs. double subtraction, auditory vs. visual presentation of the words), there is a close correspondence in the activation patterns seen for the basic minus subordinate comparison. It is worth noting, however, that inspection of the individual activation maps suggests that at least some of this activation may be taking place in large vessels, although it is difficult to know without angiography, not performed in this study. The pattern is similar to that sometimes obtained when subtracting from a fixation baseline, suggestive of draining vessels.

Greater visual cortex activation for basic-level over subordinate-level recognition is inconsistent with almost all accounts of object categorisation. Models typically postulate that the basic (or entry) level is necessarily accessed before access to the subordinate level (Jolicoeur et al., 1984; Kosslyn et al., 1995) or that subordinate-level judgements require a higher criterion than basic-level judgements (without attributing a special status to basic-level categories). The assumption is that more features must be verified to distinguish among objects at the subordinate level as compared to objects at the basic level (G. Murphy & Smith, 1982; G.L. Murphy, 1991). Consequently, no additional processing would be predicted for making basic-level judgements relative to subordinate-level judgements. Because our finding would require a major revision of object recognition theories and given our uncertainty as to its source, we favour a cautious interpretation.

Semantic Subordinate Minus Basic

The Semantic Subordinate minus Basic comparison revealed significant activation in the parahippocampal gyri (parahippG). Early visual areas such as the linG and the postFG show higher than average densities of activation but are not sig-

nificant. Therefore, our results provide little support for Gauthier et al.'s (1997) hypothesis that subordinate-level words may recruit early visual cortex through visual imagery. However, postMTG, activated for visual subordinate judgements presumably because of longer subordinate-level words on average, also shows nonsignificant but higher than average density of activation. This suggests that the reduced statistical power in the Semantic task relative to the visual task (8 vs. 11 subjects) may have contributed to our null result and that the same issue should be addressed in a larger study. It should be noted that the lingual gyrus was more strongly activated during Semantic than Visual subordinate-level judgements $[F(1,17) = 5.05, P < .05]$.

Semantic Basic Minus Subordinate. No region was expected to be specifically recruited by basic-level semantic judgements and none showed activation surviving our corrected α-level. However the infOG came close ($P < .0004$) and also showed stronger activation for Semantic than Visual basic-level judgements $[F(1,17) = 28.0, P < .0001]$.

Visual vs. Semantic Judgements

Given the many differences between the Visual and Semantic conditions (different number of subjects, different words, basic and subordinate words conceptually related in the Visual task, as they described the same images, but not in the Semantic task), only a qualitative summary (rather than a quantitative analysis) of the regions engaged in the two experiments will be reported here.

Common Areas

The postMTG, significantly recruited for subordinate visual judgements and only showing numerically above-average density of activation for semantic subordinate judgements, is thought to reflect secondary auditory cortex activation caused by the longer subordinate-level words (Binder et al., 1997). The left posterior inferior temporal gyrus was significantly engaged by both visual and semantic subordinate-level judgements. Binder et al. investigated cerebral linguistic areas involved in making semantic judgements upon the presenta-

tion of aurally presented animal names (relative to pitch-based decisions about tones). Consistent with the present results, Binder et al. also found activation of the left inferior temporal gyrus and suggested a role in comprehension at a linguistic/ semantic level. Interestingly, the words and tones in this study were equated in duration (unlike our subordinate-level and basic-level words), suggesting that a linguistic component may be sufficient to engage this area.

Differentially-engaged Areas. Subordinate-level visual judgements engaged the midFG gyrus, consistent with the double-subtraction results reported by Gauthier et al. (1997). The entire OG (inferior, lateral and superior—predominantly BA 19) was also engaged for subordinate visual but not semantic judgements. This is consistent with the demonstration by Jolicoeur et al. (1984) that picture recognition at the subordinate level requires more perceptual processing than recognition of the same picture at the basic level. The parahippG was specifically engaged by subordinate-level semantic judgements. This area was also activated in the Semantic task using animal names in Binder et al. (1997), who suggested that processing at a semantic level may have enhanced storage of episodic memories (Craik & Lockhart, 1972; Zola-Morgan, Squire, Amaral, & Suzuki, 1989). The parahippocampus has been implicated in the encoding of novel material (Gabrieli, Brewer, Desmond, & Glover, 1997; Stern et al., 1996), but our Semantic task involved no novelty manipulation. However, the subordinate-level words clearly had lower word frequencies and it is possible that this factor could lead to a pseudo-novelty effect.

The absence of common activation for basic-level visual and semantic judgements is consistent with the hypothesis that there is no additional semantic/linguistic component involved in processing basic-level words as compared to subordinate-level words (that is, the comprehension of the word "pelican" includes the comprehension of the concept "bird"). This is because the only shared feature in the Visual and Semantic tasks is the auditory presentation of words. Therefore, any component that would be mediated solely by language process-

ing should be common to both tasks. Of course, differences may occur in brain areas not covered by our slices.

Results for Faces Minus Objects in Anatomically Defined ROIs

The normalised density of activation (sum of t-values over $t = 0.1$, normalised by the sum of t-values in all ROIs in a given hemisphere) for the Localiser task (Faces minus Objects and Objects minus Faces) in each anatomically defined ROI (Plate 2) was compared by t-test to a baseline of 1 (average density in the hemisphere). For Faces minus Objects, two ROIs (midFG and postITG) were predicted a priori to be activated and statistical significance was assessed at an α-level of .05. Alpha was corrected for all other ROIs (.0035 for 14 ROIs in Faces minus Objects, .0031 for 16 ROIs in Objects minus Faces). For Faces minus Objects, activation density in the midFG was not significantly higher than the average (mean = 1.065, n.s.). Figure 2 shows that the group-average "face area" is located within the midFG ROI but is only a small portion of this ROI: A measure of activation based on density is highly sensitive to the fit between the ROI size and activation size. Furthermore, Fig. 2 also illustrates that some of the midFG surrounding the putative face area is actually more activated for objects than faces (mean = 1.77, [$t(10) = 3.70$, $P < .004$]. There was a significant effect of hemisphere in post ITG [$F(1,10) = 5.15$, $P < .05$], with density of activation higher than average in the left hemisphere only [right = 1.19, n.s.; left = 1.66, $t(10) = 2.89$, $P < .02$]. The left postITG activation is of particular interest given that this same area appeared to be engaged for both Visual and Semantic subordinate judgements. The left-side activation in the postITG for Faces minus Objects can hardly be attributed to a linguistic difference (in fact, nonface objects might have been expected to lead to stronger naming responses than unfamiliar faces).

For the Faces minus Objects comparison, only one additional ROI (cuneus) showed higher than average activation but did not pass our corrected α criterion [mean = 2.23, $t(10) = 2.94$, $P < .02$]. The density of activation in the right parahippG (visible
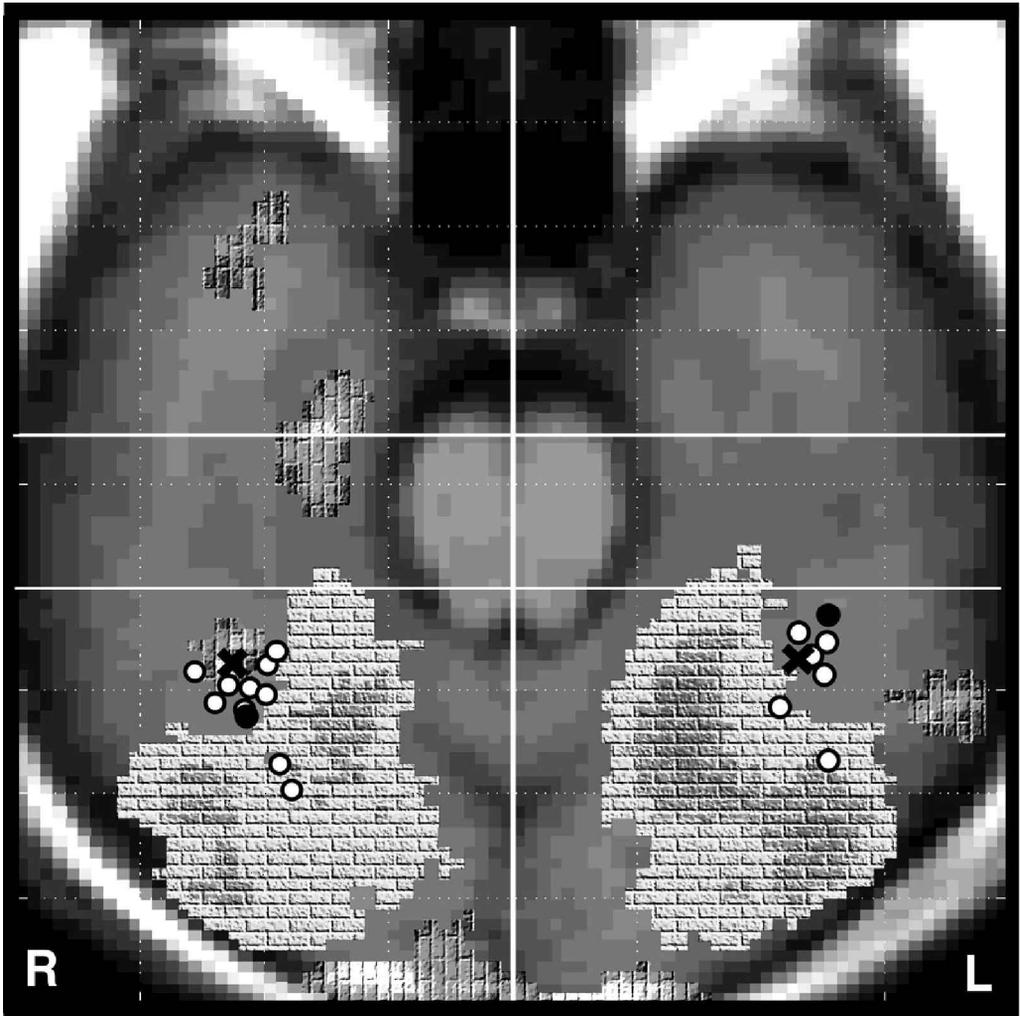
Fig. 2. *Median composite activation map for 11 subjects in the Localizer task (Faces minus Objects) in slice 3. The horizontal brick texture shows the voxels more activated for faces while the vertical brick texture indicates voxels more activated for non-face objects. The threshold is t = 0.1 for Faces minus Objects and t = 0.5 for Objects minus Faces, with a cluster size of 20 voxels. Black Xs indicate the centre of the average definition for the face area in each hemisphere. The criterion for lateralisation (see text) led to three subjects being right-lateralised, one subject being left-lateralised, six bilateral subjects, and one with no face area for the average definition. For the individual definition, this led to four subjects being right-lateralised and seven being bilateral. Centres of individual face areas in each hemisphere are shown with white-filled circles, the black-filled circles show the face area in slice 2 for a bilateral subject.*

in Fig. 2) was not significantly higher than 1 (mean = 1.12, n.s.) but there was a significant effect of Hemisphere in the hippocampus ROI, stronger on the right [right = .99; left = .56; $F(1,10) = 7.32$, $P < .02$]. Four additional ROIs were more activated bilaterally for objects than faces: the left infOG

[mean = 3.22, $t(10) = 4.31$, $P < .002$], the latOG [mean = 4.74, [$t(10) = 4.93$, $P < .0006$], the postFG [mean = 3.93, $t(10) = 9.67$, $P < .0001$] and the supOG [mean = 3.75, $t(10) = 5.14$, $P < .0004$].

The fact that the density of activation was significant in the midGF for the Visual Subordinate

minus Basic comparison, whereas it was not in the Faces minus Objects comparison, can be explained by the much wider focus of activation in the Visual task than the Localiser. To address this issue, in the next section we introduce a technique in which we positioned a small ROI on the putative face area in the average as well as the individual results for the Localiser task, the purpose being to compare the activation in the Visual task for these two definitions; in doing so, we also verify that there is a local peak of activation in the midFG for the Faces minus Objects comparison.

COMPARISON OF AVERAGE AND INDIVIDUAL DEFINITIONS OF THE FACE AREA

Definitions of the Face Area ROI

Because of the known variability in the location of the face area, Gauthier et al. (1997) used a large ROI, which should have included most subjects' face areas based on results from intracranial recordings and fMRI (Allison et al., 1994; Puce et al., 1995, 1996). The caveat of this technique, as shown in the previous section, is that such a large ROI is also likely to include cortical areas that are actually more responsive to the presentation of nonface objects than faces.

Here we compare two different definitions of the face area. In the *average* definition, a small ROI is centred on the face area of a composite map for the 11 subjects. If such a group-average definition is found to be an efficient way of localising the face area within a restricted region, it has the advantage that only a limited number of Localiser images are required of any subject in an experiment. This is important if a group-average definition is to be used in further imaging studies—the number of images that can be obtained from one subject in one session is limited. In contrast, when sufficient data for a Localiser task is gathered in each subject (here for instance, we took as many images for the Localiser task for the Visual task), an *individual* definition of the functional face area can be used. This method has proven useful in prior studies but it has never

been compared with an average definition (Kanwisher et al., 1997a; Gauthier, Tarr, Anderson, & Gore, 1999).

The centre of the *average* face area was positioned in the Localiser task, for the midFG foci in each hemisphere (Slice 3) that showed the strongest activation for Faces minus Objects in the median composite for all subjects (see Fig. 2). *Individual* face area ROIs were centred on the strongest midFG focus (or postFG in some subjects) encountered in the Localiser task for each subject, in each hemisphere independently. All ROIs were localised without regard to the activation in the Visual task.

Apart from the two different ways of positioning ROIs, the same procedure was used to quantify activation according to both definitions. A grid composed of a central box-ROI of 5 × 5 voxels surrounded by four adjacent boxes of the same size (see Fig. 2) was centred on the average and individual face areas. This ROI was defined in standardised (pseudo-Talairach) space and each voxel was 1.3mm × 1.7mm in plane (about 5.4 of the original voxels which were 3.125mm × 3.125mm each). The size of our face area ROI is thus comparable (and even smaller) to that defined in published studies (Kanwisher et al., 1997a, 1998). The size of the box was chosen to match the size of the face area in the Localiser task. The individual definition includes on average 23 out of 25 standardised voxels with signal higher for faces than objects. A larger area includes more voxels with a signal change in the other direction (objects > faces), while a smaller area leads to a less focused "face area", with activation for Faces minus Objects spilling to contiguous ROIs. In each of these box-ROIs, the per cent signal change was summed for voxels for the Faces minus Objects and the Visual Subordinate minus Basic comparisons. This measure is similar to that of Kanwisher et al. (1997a).

The Localiser results for some subjects showed bilateral activation in the midFG whereas for others the activation was more strongly lateralised. In order to define a single face area in each individual, the data for each subject were obtained only from the dominant hemisphere (for the Localiser task) or from both hemispheres combined if the Localiser

activation was bilateral. For both the individual and average definitions, subjects were defined as "lateralised" if they had at least 50% more voxels showing higher signal for faces than objects in one hemisphere than in the other. Because laterality was determined for each definition separately, a subject may be lateralised in one definition and bilateral in the other (see Fig. 2). Summed per cent signal change was measured in each of the five box-ROIs and averaged across hemisphere for bilateral subjects and only in the dominant hemisphere for lateralised subjects.

Results

The data were first combined across hemispheres for "bilateral" subjects, and a mean was computed for each box-ROI. Mean summed activation in each box-ROI for the individual and average definitions of the face area are shown in Fig. 3. For each definition, an ANOVA was performed on the summed activation with Task (Visual Subordinate minus Basic vs. Faces minus Objects) and ROI (1–5) as within-subjects factors.

For the average definition, none of the main effects was significant, nor was the interaction between ROI and Task (F < 1). There was no difference between the different box-ROIs for Faces minus Objects nor for Subordinate minus Basic,

even using the powerful Least Significant Difference (LSD) post hoc test (using $P < .05$). LSD tests showed a higher signal change for Faces minus Objects than Subordinate minus Basic judgements in box ROIs 1 and 4.

For the individual definition, only the main effect of ROI was significant [$F(4,10) = 2.75$, $P < .05$]. The Task by ROI interaction was not significant [$F(4,40) = 1.53$, n.s.]. LSD tests showed more activation in box-ROI 3 (face area) than all other box-ROIs for Faces minus Objects but no difference between box-ROIs for Subordinate minus Basic. The only task differences occurred in box-ROIs 1 and 4, with a stronger signal change for Subordinate minus Basic than for Faces minus Objects. Having defined ROI based on the Localiser results, we maximised the amount of activation in this area for the Faces minus Objects comparison: A regression to the mean would disadvantage the Visual task, playing against our hypothesis.

We also compared directly the individual and average definitions in each box-ROI, for each of the two tasks. The only significant difference occurred in the box-ROI 3 (face area) for Faces minus Objects, where the individual definition led to a stronger summed signal change than the average definition [mean difference = 4.3, $t(10) = 3.56$, $P = .005$]. Thus, the individual definition leads to both stronger and more focused activation in the

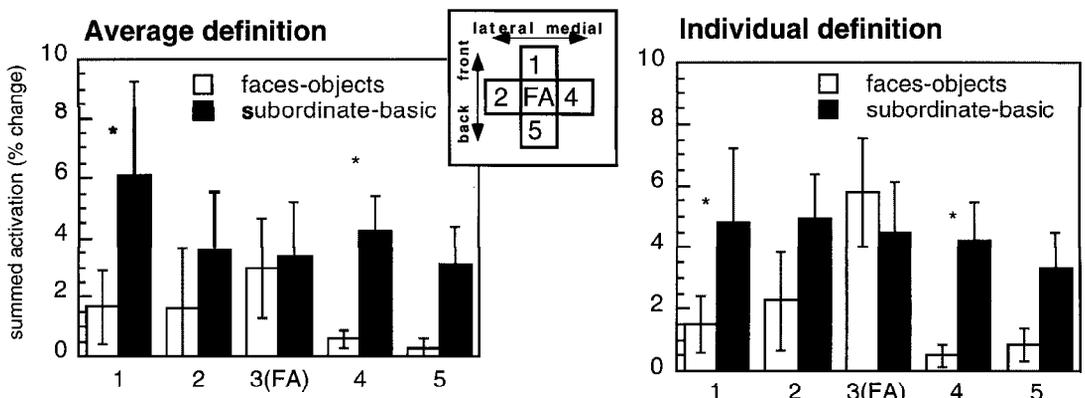**Fig. 3.** *Summed per cent signal change in each 5 × 5 standardized voxel square ROI (a volume of .38cm³) for the average and indiviudal definitions of the face area, for data combined across hemisphere for "bilateral" subjects and that from the dominant hemisphere in "lateralised" subjects. Asterisks indicate significant paired t-tests.*

face area for Faces minus Objects. This is not surprising, because it is optimised for each subject. However, it is worth pointing out that a mean individual definition is more similar to each individual face area, which typically shows a strong and focused peak of activation, and thus may be necessary in order to investigate the spatial distribution of the activation.

DISCUSSION

Categorisation Level in Visual and Semantic Processing

Categorisation level was found to be an important determinant for the recruitment of several occipital and temporal areas in both visual and semantic judgements. Our results are consistent with Jolicoeur et al. (1984) in that subordinate-level judgements required additional perceptual processing relative to basic-level judgements. Subordinate-level visual judgements engaged a much larger portion of the ventral pathway than their basic-level counterparts, including a large part of the fusiform gyri as well as the ventral and lateral aspects of the occipital gyri.

We also tested the hypothesis formulated by Gauthier et al. (1997), that subordinate-level semantic judgements may recruit early visual areas because of visual imagery. In Gauthier et al. this effect could have been the result of the differential word length in the Semantic task when subordinate-level and basic-level words were visually presented. Here, by using aurally presented words and asking subjects to close their eyes, the additional length of the subordinate-level words was carried in the auditory domain (presumably causing postMTG activation). This reduced the extent of activation as compared to the visual presentation. We obtained no strong evidence of early visual activation by semantic subordinate-level judgements. This result, however, should be considered cautiously because of two caveats: a possible lack of statistical power as well as the fact that the lingual gyrus was significantly more activated for Semantic than Visual subordinate-level judgements.

Finally, the recruitment of the left postITG in all three of the comparisons used in this study came as somewhat of a surprise. Both our Visual and Semantic Subordinate minus Basic comparisons could lead to differential linguistic/semantic processing. This would be consistent with the activation of the left inferior temporal gyrus in a study by Binder et al. (1997) for semantic judgements on animals' names when compared to judgements on tones. However, passive viewing of faces should not evoke any more linguistic/semantic processing than that of nonface objects, but the postITG was also recruited, only in the left hemisphere, in our Faces minus Objects comparison. We originally postulated that subordinate-level visual judgements would share perceptual processing with that of faces even during passive viewing of the latter. It is possible that the common denominator in all of our tasks, as well as Binder et al.'s study, is an aspect of visual processing that is elicited by concrete nouns more than by tones and by subordinate-level more than by basic-level judgements, regardless of whether they rely on perception or on mental images. In order for this to be resolved, a stronger test of visual imagery for subordinate semantic judgements is required.

The Role of Categorisation Level in Face Recognition

The single subtraction design used here was successful in replicating, within individual subjects, the pattern of results obtained by Gauthier et al. (1997). The present results also extend our earlier findings, providing additional information on the role of categorisation level in the comparison between faces and objects. Consistent with Gauthier et al., the anatomically defined mid-fusiform gyrus was found to be engaged by visual subordinate-level judgements with nonface objects. An individual functional definition of the face area was found to be preferable to an anatomical definition or to an average functional definition for measuring the focal signal change for faces minus objects. Critically, subordinate-level processing recruited the face area using both an average and an individual definition. This is evidence against the

hypothesis that the subordinate-level activation isolated by Gauthier et al. (1997) could be mutually exclusive from the face area within individual subjects.

Subordinate-level processing elicited activation of a magnitude comparable to that obtained for Faces minus Objects in the individually defined face area, *as well as* in surrounding ventral temporal cortex. Therefore, categorisation level, a dimension that we argue is often confounded with stimulus-class membership in experimental designs, could potentially account for the *magnitude* of the activation in the face area but not for the specificity in spatial distribution. In other words, the face area is a subset of the region of midFG cortex that is implicated in subordinate-level processing of objects. In reality, it is unlikely that level of categorisation can account for the entire magnitude of the faces minus objects effect, given that studies that compared face and object processing at the subordinate level obtained stronger activation for faces (Kanwisher et al., 1997a; Tong et al., 1998, but see our earlier discussion regarding a "1-back" task with rare target repetitions).

Before we accept the hypothesis that categorisation level accounts for the magnitude of activation in the face area, several issues need to be addressed. First, is this not merely the result of a difficulty effect? And if it is, what is the relation between difficulty and face recognition? Second, even if some of the Faces minus Objects effects are due to a level of categorisation difference, what accounts for the spatial specificity of the activation pattern?

There is no doubt that there is a difficulty effect intrinsic to our level of categorisation manipulation: We selected our stimuli so that each picture would be matched faster to its basic-level name than its subordinate-level name (Gauthier et al., 1997). The basic-level advantage is crucial to categorisation work (Rosch, 1978; Rosch et al., 1976; Tanaka & Taylor, 1991) and it would be meaningless to attempt to equate level of difficulty between basic and subordinate-level judgements. Consider that objects can often be categorised at the basic level merely by the presence of single features or configurations of features (e.g. the presence of wings is highly diagnostic of the category "bird"),

whereas objects within the same basic-level category share parts and their configuration (e.g. all cars have wheels, a bumper, and headlights in the same relative configuration). On the other hand, some perceptual dimensions have been found to be more important to the processing of subordinate-level tasks (including faces) as compared to basic-level tasks, for example, shading, texture, colour, surface detail, pigmentation, and the specific spatial arrangement of features (see Bruce & Humphreys, 1994, for review).

In this context, the activation of most of the ventral pathway obtained here for Subordinate minus Basic can be viewed as the result of the additional visual attention or more complex processing required for subordinate-level judgements. However, this does not mean that the fusiform and occipital gyri play the same undifferentiated role. Some evidence comes from Gauthier et al. (1997), who used a visual semantic control condition in which subordinate-level and basic-level words, matched in many respects (including a difficulty effect) with those used in the visual task, were presented visually (rather than aurally). When this Semantic Subordinate minus Basic control (which includes a difference in difficulty) was subtracted from the Visual Subordinate minus Basic comparison, what remained of the ventral temporal lobe activation was only the fusiform-ITG region. In addition, one can consider the effect of brain lesions in different parts of the visual system. Lesions anywhere in the ventral visual pathway can produce visual recognition impairments that may be worse for subordinate than basic-level recognition, again simply because subordinate-level judgements are more difficult. However, a recent study (Gauthier, Behrmann, & Tarr, 1999) reveals that prosopagnosic patients (prosopagnosia tends to be associated with fusiform damage [A.R. Damasio, Damasio, & Hoesen, 1982; A.R. Damasio, Tranel, & Damasio, 1990]) may be normal at basic-level recognition but show an increasing impairment as the recognition task becomes increasingly subordinate (e.g. discriminating a chair, a pelican, or a duck from another duck). In other words, although a level-of-categorisation manipulation leads to a significant difficulty effect in normal subjects (who

respond more slowly the more subordinate the comparison), prosopagnosic patients show a disproportionate sensitivity. Many studies report that prosopagnosic patients have configural processing difficulties (Arguin, Bub, & Dudek, 1996; Davidoff, Matthews, & Newcombe, 1986; Levine & Calvanio, 1989; Rentschler, Treutwein, & Landis, 1994; Williams & Behrmann, 1998), so it appears that for the most part, they can see local features but cannot integrate them to discriminate visually similar objects. Subordinate-level recognition probably requires additional processing of local features as well as combining them into more global configurations. Thus, it is possible that the additional occipital lobe activation for subordinate-level matching is due to additional featural processing and the additional fusiform activation to the additional configural processing.

Our initial claim was that the often-used Faces minus Objects task includes, among other things, a subordinate-level effect. How can this hold, when Faces minus Objects does not include a difficulty effect (given the absence of a task) nor lead to increased activation of the occipital lobe? It may be worth distinguishing two ways in which tasks may differ in difficulty. On the one hand, a given task may require more complex or additional computational steps, causing subjects to perform it more slowly and less accurately than another task. On the other hand, two tasks may include comparable computational demands but subjects may have much more experience solving one task than another, in practice making it less difficult (similar to an automaticity effect). Expertise in visual recognition leads to what has been called a "basic-level shift" (Tanaka & Gauthier, 1997; Tanaka & Taylor, 1991): While novices take longer identifying a robin as a robin than as a bird, expert birders are equivalently fast at both levels. Although we do not understand what mediates this shift or what neural substrates support such a shift, we do know that experts use configurational clues more than novices. Thus, expertise leads to well-established behavioural effects (configural processing) that are similar to those obtained with faces (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Gauthier, Williams, Tarr, & Tanaka, 1998).

Such configural processing may be reflected in a relative shift of the burden of processing from early visual areas to the midFG. In other words, it is possible, and consistent with the behavioural literature, that the relative absence of occipital lobe activation for faces (as compared to subordinate-level matching for nonface objects) reflects an expertise shift in difficulty. Faces recognised at the subordinate level may require configural processing similar to that necessary for subordinate-level object matching (and presumably performed in the midFG), but expert subjects may be more efficient in earlier computational steps that feed in the midFG and/or they may have learned to use relatively fewer processing resources than novices. Indeed, a recent experiment (Gauthier et al., 1999) demonstrated that when subjects are experts with a category of novel objects ("Greebles"), identity matching of upright minus inverted Greebles as well as passive viewing of Greebles minus Objects both produce focal activation comparable in spatial extent to that obtained with faces. Thus, the combination of categorisation level and expertise appears to provide a plausible explanation for the strong focal activation obtained when faces are compared to nonface objects.

One piece of evidence that may be hard to reconcile with our conclusions is the behavioural pattern exhibited by patient CK (Moscovitch, Winocur, & Behrmann, 1997). CK has selectively spared face recognition abilities, but fails at subordinate-level (even basic-level in some cases) recognition of nonface objects. Such results are consistent with nonexpert subordinate-level processing requiring much of the ventral pathway as well as the midFG. However, anecdotal evidence suggests that CK has lost his expertise in recognising toy soldiers, *inconsistent* with his face recognition abilities being spared because of expertise (although it is possible that knowledge of toy soldiers has more to do with local features—colours of uniforms, shape of hats, types of weapons, flags— than with configural cues). Ideally, we would want to know what would have happened if CK had premorbidly been an expert with objects, such as Greebles, that have been shown to produce behavioural configural effects and recruit the face area in

experts. Obviously, this particular question cannot be answered.

Our experiments tested directly what was already implicit in the neuroimaging literature: There is nothing unique about faces activating the fusiform gyrus, since many other tasks also do so (Kanwisher et al., 1997a; Lang et al., 1998; Schacter et al., 1995). In this study, however, we have established that midFG activation for nonface objects includes the face area even when measured in individual subjects. We have also demonstrated that an endogenous manipulation of categorisation level, using identical pictures of nonface objects, can lead to a signal change in the face area of a magnitude comparable to that of passive viewing of faces minus objects. Our results point to what may be special about the way faces activate the midFG: For each individual, faces lead to spatially focused activation in a small part of the fusiform gyrus. Categorisation level does not seem to be able to account for this spatial specificity. Rather than focusing only on this small area of the cortex (Kanwisher et al., 1997a, 1998; Tong et al., 1998), neuroimaging research may have to consider how the activation in the entire ventral pathway is modulated by different object categories, tasks, and levels of expertise. We may arrive at new and productive hypotheses by rephrasing the question "Why is the face area so strongly engaged by faces?" into "Why is the surrounding cortex so weakly engaged by faces?". Our results also highlight the importance of controlling for categorisation level in making comparisons across stimulus classes. Consider the proposal by McCarthy et al. (1997) that face-specific processing can only be isolated if the "general" object recognition system is occupied by concurrent object processing. To demonstrate this point, these authors devised a stimulus presentation design in which faces (or flowers) would appear periodically at one of several locations while nonface objects would occupy the other positions randomly. However, in light of our findings, it appears that there is a confound in this design—nonface objects (for instance the objects presented during passive viewing in our Localiser task) are *not* processed automatically at the subordinate level (Jolicoeur et al., 1984; Rosch et al., 1976; Schultz et al., 1997; Tanaka & Gauthier, 1997). Therefore, the background objects presented in this study in all likelihood *did not* "occupy" general subordinate-level processes to any significant degree. Moreover, it is unlikely that flowers are processed at the subordinate level without specific instructions to this effect.

The current debate on whether the putative face area is or is not a face-specific module has focused on contrasting two types of evidence. First, that the "face area" is engaged more strongly by faces than any other type of nonface objects. Second, that under some circumstances (e.g. expertise, subordinate-level processing), nonface objects may recruit the same area. This debate may be doomed to remain unresolved in that even if faces turn out to consistently engage a specific part of the midFG more than other objects, it would still be unclear what factors produce such specialisation. Thus, we suggest that a more profitable course is to explore how the "face area" becomes specialised by resolving the conditions under which nonface objects can also activate this area.

REFERENCES

Allison, T., Ginter, H., McCarthy, G., Nobre, A., Puce, A., Luby, M., & Spencer, D.D. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology, 71*(2), 821–825.

Arguin, M., Bub, D., & Dudek, G. (1996). Shape integration for visual object recognition and its implications in category-specific visual agnosia. *Visual Cognition, 3*(3), 221–275.

Binder, J.R., Frost, J., Hammeke, T., Cox, R.W., Rao, S.M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17*(1), 353–362.

Bruce, V., & Humphreys, G. (1994). Recognizing faces and objects. *Visual Cognition, 1*, 141–180.

Constable, R., Studlarski, P., Mencl, E., Pugh, K.R., Fulbright, R., Lacadie, C., Shaywitz, S.E., & Shaywitz, B.A. (1998). Quantifying and comparing ROI activation patterns in functional brain MR imaging: Methodology considerations. *Magnetic Resonance Imaging, 16*, 289–300.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11,* 671–684.

Damasio, A.R., Damasio, H., & Hoesen, G.W.V. (1982). Prosopagnosia: Anatomical basis and behavioral mechanisms. *Neurology, 32,* 331–341.

Damasio, A.R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience, 13,* 89–109.

Damasio, H. (1995). *Human brain anatomy in computerized images.* New York: Oxford University Press.

Davidoff, J.B., Matthews, W.B., & Newcombe, F. (1986). Observations on a case of prosopagnosia. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 279–290). Dordrecht, The Netherlands: Martinus Nijhoff.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107–117.

Dill, M., & Fahle, M. (1998). Limited translation invariance of human visual pattern recognition. *Perception and Psychophysics, 60*(1), 65–81.

Duvernoy, H.M. (1991). *The human brain. Surface, three-dimensional sectional anatomy and MRI.* Wien: Springer-Verlag.

Farrell, B. (1985). "Same"–"different" judgements: A review of current controversies in perceptual comparisons. *Psychological Bulletin, 98*(3), 419–456.

Gabrieli, J.D.E., Brewer, J.B., Desmond, J.E., & Glover, G.H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science, 276,* 264–266.

Gauthier, I., Anderson, A.W., Tarr, M.J., Skudlarski, P., & Gore, J.C. (1997). Levels of categorization in visual object studied with functional MRI. *Current Biology, 7,* 645–651.

Gauthier, I., Behrmann, M., & Tarr, M.J. (1999a). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience, 11,* 349–370.

Gauthier, I., & Tarr, M.J. (1997). Becoming a "Greeble" expert: Exploring the face recognition mechanism. *Vision Research, 37*(12), 1673–1682.

Gauthier, I., Tarr, M.J., Anderson, A., & Gore, J. (1999b). Activation of the middle fusiform "face area" increases with experience in recognizing novel objects. *Nature Neuroscience, 2,* 568–573

Gauthier, I., Williams, P., Tarr, M.J., & Tanaka, J.W. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38,* 2401–2428.

Haxby, J.V., Horwitz, B., Ungerleider, L.B., Maisog, J.M., Pietrini, P., & Grady, C.L. (1994). The functional organization of human extrastriate cortex: A PET/RCBF study of selective attention to faces and locations. *The Journal of Neuroscience, 14,* 6336–6353.

Ishai, A., Ungerleider, L.G., Martin, A., Maisog, J.M., & Haxby, J.V. (1997). fMRI reveals differential activation in the ventral object vision pathway during the perception of faces, houses, and chairs. *NeuroImage, 5,* S149.

Johnson, C.L., Paivio, A., & Clark, J.M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120*(1), 113–139.

Jolicoeur, P., Gluck, M., & Kosslyn, S.M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16,* 243–275.

Kanwisher, N., Chun, M.M., McDermott, J., & Ledden, P.J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research, 5,* 55–67.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997a). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*(1), B1–11.

Kanwisher, N., Woods, R., Iacoboni, M., & Mazziota, J. (1997b). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience, 9,* 133–142.

Kosslyn, S.M., Alpert, N.M., & Thompson, W.L. (1995). Identifying objects at different levels of hierarchy: A positron emission tomography study. *Human Brain Mapping, 3,* 107–132.

Lang, P.J., Bradley, M.M., Fitzsimmons, J.R., Scott, J.D., Moulder, B., & Nangia, V. (1998). Emotional arousal and activation of the visual cortex: An fmri analysis. *Psychophysiology, 35*(2), 199–210.

Levin, D.T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 1364–1382.

Levine, D.N., & Calvanio, R. (1989). Prosopagnosia: A defect in visual configural processing. *Brain and Cognition, 10,* 149–170.

Malach, R., Reppas, J., Benson, R., Kwong, K., Jiang, H., Kennedy, W., Ledden, P., Brady, T., Rosen, B., & Tootell, R. (1995). Object-related activity revealed by functional magnetic resonance imaging in

human occipital cortex. *Proceedings of the National Academy of Sciences, USA, 18,* 8135–8139.

McCarthy, G., Puce, A., Gore, J., & Allison, T. (1997). Face-specific processing in the fusiform gyrus. *Journal of Cognitive Neuroscience, 9*(5), 605–610.

Moscovitch, M., Kapur, S., Kohler, S., & Houle, S. (1995). Distinct neural correlates of visual long-term memory for spatial location and object identity: A positron emission tomography study in humans. *Proceedings of the National Academy of Sciences, USA, 92,* 3721–3725.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special in face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9*(5), 555–604.

Murphy, G., & Smith, E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behaviour, 21,* 120.

Murphy, G.L. (1991). Parts in object concepts: Experiments with artificial categories. *Memory and Cognition, 19,* 423–438.

Puce, A., Allison, T., Asgari, M., Gore, J.C., & McCarthy, G. (1996). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Neurophysiology, 74*(3), 1192–1199.

Puce, A., Allison, T., Gore, J.C., & McCarthy, G. (1995). Differential sensitivity of human visual cortex to faces, letter strings, and textures: A functional magnetic resonance imaging study. *Neurophysiology, 74*(3), 1192–1199.

Puce, A., Allison, T., Spencer, S.S., Spencer, D.D., & McCarthy, G. (1997). Comparisons of cortical activation evoked by faces by intracranial field potentials and functional MRI: Two case studies. *Human Brain Mapping, 5,* 298–305.

Rentschler, I., Treutwein, B., & Landis, T. (1994). Dissociation of local and global processing in visual agnosia. *Vision Research, 34,* 963–971.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19,* 473–497.

Rhodes, G., & McLean, I.G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding. *Perception, 19,* 773–794.

Rosch, E. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

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

Schultz, R.T., Gauthier, I., Fulbright, R., Anderson, A.W., Lacadie, C., Skudlarski, P., Tarr, M.J., Cohen, D.J., & Gore, J.C. (1997). Are face identity and emotion processed automatically? *Neuroimage, 5*(4), S148.

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

Sheskin, D.J. (1997). *Handbook of parametric and nonparametric statistical procedures.* Boca Raton, FA: CRC Press.

Skudlarski, P., Constable, R.T., & Gore, J.C. (1999). ROC analysis of statistical methods used in functional MRI. *Neuroimage, 9,* 311-329.

Stern, C.E., Corkin, S., Gonzalez, R.G., Guimaraes, A.R., Baker, J.R., Carr, C.A., Sugiura, R.M., Vedantham, V., & Rosen, B.R. (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences, USA, 93*(16), 8660–8665.

Talairach, J., & Tournoux, P. (1998). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportion system: An approach to cerebral imaging.* Stuttgart: G. Thieme.

Tanaka, J.W., & Gauthier, I. (1997). Expertise in object and face recognition. In R.L. Goldstone, P.G. Schyns, & D.L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 36, pp. 83–125). San Diego, CA: Academic Press.

Tanaka, J.W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (1998). *Response properties of the human fusiform face area.* Unpublished manuscript, Harvard University.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–193.

Williams, P., & Behrmann, M. (1998). *Acquisition and recognition of novel entry–level object categories in prosopagnosic patients.* Paper presented at the 69th annual meeting of the Eastern Psychological Association, Boston, MA.

Williams, P., & Tarr, M.J. (undated). *RSVP: Experimental control software for MacOS.* [Online]. Available: http://psych.umb.edu/rsvp/ [1998, September 9th].

Zeki, S., Watson, J.D., Friston, K., & Frackowiak, R. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience, 11*(3), 641–649.

Zola-Morgan, S. Squire, L.R., Amaral, D.G., & Suzuki, W.A. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *Journal of Neuroscience, 9,* 4355–4370.

APPENDIX A

List of Words Used in the Visual Task

Picture (Sub Matching)	Sub Distractor	Basic Matching	Basic Distractor	Picture (Sub Matching)	Sub Distractor	Basic Matching	Basic Distractor
Acoustic guitar	Electric guitar	Guitar	Phone	Lounge chair	Beach chair	Chair	Phone
Alarm clock	Mantle clock	Clock	Chair	Mantle clock	Alarm clock	Clock	Chair
Ant	Bee	Insect	Bird	Motorboat	Dinghy	Boat	Chair
Beach chair	Classroom chair	Chair	Dog	Mug	Tea cup	Cup	Chair
Bee	Fly	Insect	Car	Payphone	Rotary phone	Phone	Piano
Biplane	Boeing	Plane	Dog	Pelican	Sparrow	Bird	Dinosaur
Blue Whale	Beluga	Whale	Phone	Picnic table	Coffee table	Table	Guitar
Boeing	Bomber	Plane	Clock	Pine	Cedar	Tree	Hat
Bomber	Seaplane	Plane	Car	Police car	Honda	Car	Shoe
Camaro	Station wagon	Car	Tree	Poster bed	Bunkbed	Bed	Boat
City bus	School bus	Bus	Boat	Racing car	Honda	Car	Clock
Classroom chair	Lounge chair	Chair	Boat	Recliner	German Shepherd	Dog	Chair
Coffee table	Dining table	Table	Insect	Rotary phone	Touchtone phone	Phone	Car
Cruiser	Galleon	Boat	Car	Sandal	Sneaker	Shoe	Bus
Daisy	Rose	Flower	Car	School bus	Tour bus	Bus	Bird
Deck chair	Folding chair	Chair	Plane	Seaplane	Triplane	Plane	Car
Desk chair	Recliner	Chair	Car	Sneaker	Sandal	Shoe	Bus
Dinghy	Speedboat	Boat	Chair	Sparrow	Eagle	Bird	Fish
Dining table	Picnic table	Table	Piano	Speedboat	Cruiser	Boat	Piano
Diplodocus	Triceratops	Dinosaur	Bird	Station wagon	Lamborghini	Car	Bottle
Duck	Pelican	Bird	Car	T-rex	Diplodocus	Dinosaur	Table
Dumptruck	Tractor trailer	Truck	Whale	Tabasco bottle	Wine bottle	Bottle	Boat
Eagle	Jay	Bird	Dinosaur	Tea cup	Mug	Cup	Plane
Electric piano	Upright piano	Piano	Insect	Thunderbird	Police car	Car	Dog
Executive chair	Beach chair	Chair	Bird	Top hat	Cap	Hat	Table
Ferrari	Police Car	Car	Bus	Touchtone phone	Payphone	Phone	Table
Flamingo	Duck	Bird	Chair	Tour bus	City bus	Bus	Bird
Fluted glass	Water glass	Glass	Bird	Tractor trailer	Dumptruck	Truck	Whale
Fly	Ant	Insect	Piano	Triceratops	T-rex	Dinosaur	Fish
Folding chair	Lounge chair	Chair	Glass	Triplane	Bomber	Plane	Bed
Galleon	Motor boat	Boat	Truck	Tuna	Piranha	Fish	Cup
German shepherd	Retriever	Dog	Shoe	Upright piano	Electric piano	Piano	Plane
Goose	Eagle	Bird	Flower	Volkswagen	Thunderbird	Car	Bottle
Honda	Ferrari	Car	Bus	Water glass	Fluted glass	Glass	Bus
Jay	Pelican	Bird	Truck	Wine bottle	Tabasco bottle	Bottle	Car
Lamborghini	Camaro	Car	Insect				

APPENDIX B

Lists of Words Used in the Semantic Task

Basic	Subordinate	Basic	Subordinate	Basic	Subordinate	Basic	Subordinate
Automobile	Black bear	Bowl	Grand piano	Helicopter	Picnic basket	Pants	Siamese
Baby	Boxing glove	Box	Grizzly	Horse	Pingpong table	Pig	Ski goggles
Bag	Cobra	Bat	Hammerhead	House	Polar bear	Plant	Soda bottle
Basket	Cockatoo	Boat	Jeep	Knife	Pool table	Rabbit	Sunglasses
Bear	Corvette	Door	Ketchup bottle	Lamp	Pterodactyl	Rocket	Swan
Bell	Crane	Elephant	Lab coat	Leaf	Raincoat	Shark	Tow truck
Bike	Dagger	Fan	Maple leaf	Light	Rattlesnake	Ship	Traffic light
Bike[a]	Dalmation	Glasses	Mouthwash bottle	Motorcycle	Sailboat	Snake	Tricycle
Bone	Desk lamp	Glove	Oil lamp	Mouse	Seagull	Train	Victorian Chair
Book	Flash light	Gun	Owl				

[a]Repeated due to experimenter's error.
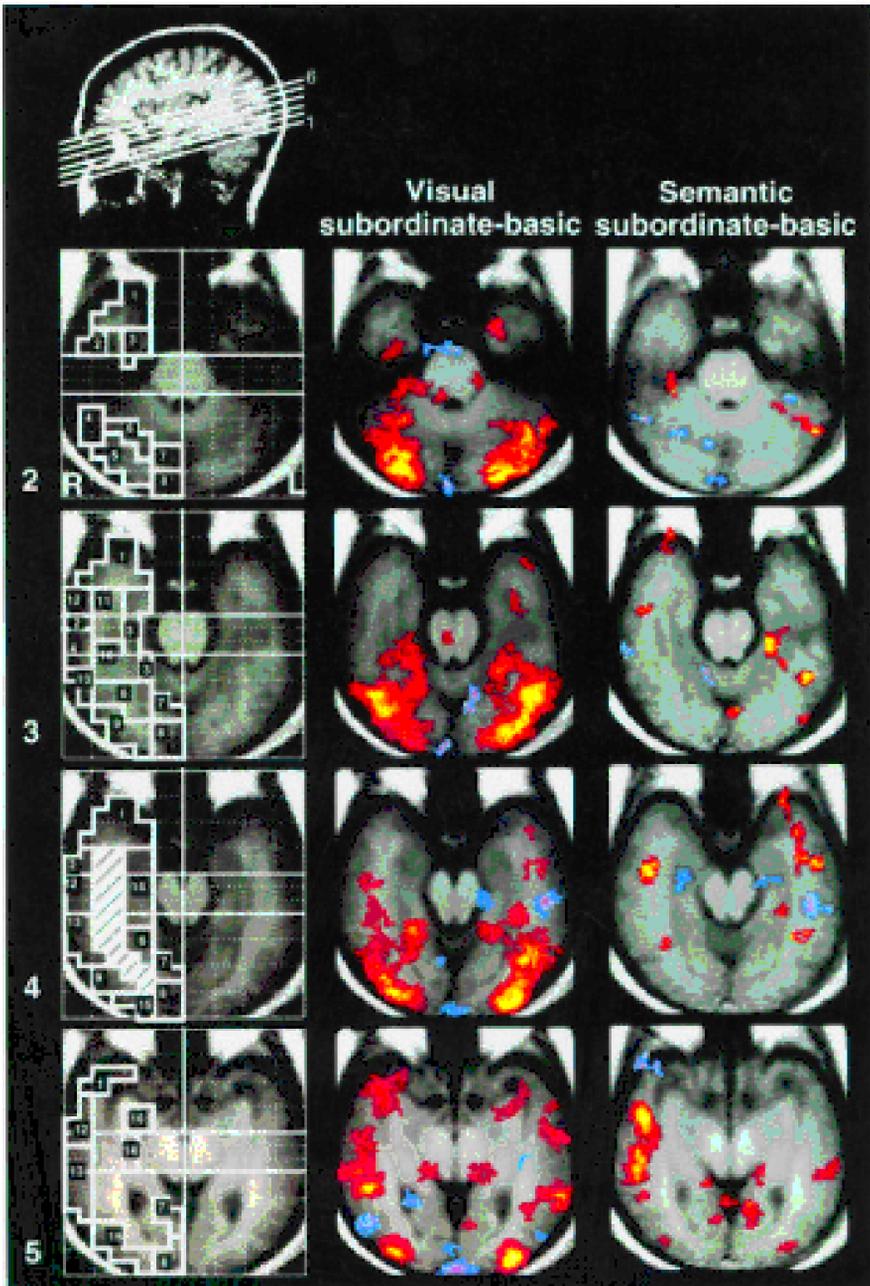
Plate 2 (Gauthier et al.). *Regions of interest defined anatomically, mapped on composite anatomical images for 11 subjects. 1, temporal poles; 2, anterior inferior temporal gyrus; 3, parahippocampal gyrus; 4, posterior inferior temporal gyrus; 5, inferior occipital gyrus; 6, posterior fusiform gyrus; 7, lingual gyrus; 8, cuneus; 9, lateral occipital gyrus; 10, middle fusiform gyrus; 11, anterior fusiform gyrus; 12, anterior middle temporal gyrus; 13, posterior middle temporal gyrus; 14, hippocampus; 15, superior occipital gyrus; 16, amygdala. Median composite image for the Visual (N = 11) and the Semantic (N = 8) Subordinate minus Basic comparisons, in four different slices through the temporal lobe. Red to yellow voxels show the voxels more activated for Subordinate than Basic level while blue to purple voxels indicate voxels activated for Basic than Subordinate level. The maps are thresholded at a t –value of 0.25 and a cluster filter of 11 voxels.*