Brain (2001), **124,** 1619–1634

The neural representation of nouns and verbs: PET studies

Lorraine K. Tyler, Richard Russell, Jalal Fadili and Helen E. Moss

Department of Experimental Psychology, University of Cambridge, UK

Correspondence to: Professor L. K. Tyler, Centre for Speech and Language, Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK
E-mail: lktyler@csl.psychol.cam.ac.uk

Summary

Neuropsychological studies of patients with selective deficits for nouns or verbs have been taken as evidence for the neural specialization of different word classes. Noun deficits are associated with lesions in anterior temporal regions while verb deficits arise from left inferior frontal lesions. However, neuroimaging studies do not unequivocally support this account, with only some studies supporting claims for regional specialization. We carried out two PET studies to determine whether there is any regional specialization for the processing of nouns and verbs. One study used the lexical decision task and the other used a more semantically demanding task, i.e. semantic categorization. We found robust activation of a semantic network extending from left inferior frontal cortex into the inferior temporal lobe, but no differences as a function of word class. We interpret these data within the framework of cognitive accounts in which conceptual knowledge is represented within a non-differentiated distributed system.

Keywords: PET; nouns and verbs

Abbreviations: fMRI = functional MRI; MNI = Montreal Neurological Institute; rCBF = regional cerebral blood flow; ROC = reviewer operating characteristic; SPM = statistical parametric mapping

Introduction

Functional neuroimaging studies, using PET and functional MRI (fMRI), suggest that conceptual knowledge is represented within an extensive network involving the left lateral temporal lobe, left posterior parietal lobe and left inferior frontal gyrus, possibly including some homologous areas in the right hemisphere (Vandenberghe *et al*., 1996; Mummery *et al*., 1998). Neuropsychological studies of patients with category-specific deficits suggest that this semantic network might be organized further as a function of factors such as word class. For example, patients have been reported with selective deficits for nouns or verbs, with lesions to anterior temporal regions associated with noun deficits, and left frontal lesions with verb deficits (e.g. Goodglass, 1966; Damasio and Tranel, 1993; Daniele *et al*., 1994). This neural differentiation is compatible with behavioural studies showing that children and adults appear to treat nouns and verbs differently. For example, they are acquired at different rates, with nouns being learned earlier than verbs (Gentner, 1981; Gleitman, 1994), and adults generally exhibit poorer performance with verbs than nouns on a variety of tests. These behavioural patterns are, in turn, consistent with linguistic analyses of nouns and verbs which claim that they differ in their structure and organization. Gentner, for example, has claimed that nouns have core invariant meanings, associated with denser connections between properties (Gentner, 1981), while Miller and Fellbaum argue that verbs have looser internal structure and greater polysemy (Miller and Fellbaum, 1991). Nouns may be organized into hierarchies with many shared correlated properties, while verbs may form a matrix-like structure, where many semantic properties are orthogonally related rather than correlated (Huttenlocher and Lui, 1979; Graesser *et al*., 1987).

However, not all of the data support the claim for neural specialization of conceptual knowledge as a function of form class. For example, there are several reports of patients with verb deficits whose lesions lie outside the left frontal regions (see Daniele *et al*., 1994) and some whose lesions primarily compromise the left temporal lobe (Damasio and Tranel, 1993). Another problem is that most of the available neuropsychological data come from naming and other lexical retrieval tasks, rather than from tasks which probe semantic representations *per se* (McCarthy and Warrington, 1985; Caramazza and Hillis, 1991; Breedin *et al*., 1998). It is possible that these patients suffer essentially from a naming deficit in which their major problem is an inability to produce the phonological form of the word, rather than from an impairment in the representation of semantic knowledge. Moreover, behavioural differences do not necessitate neural specialization. A number of recent cognitive models of conceptual knowledge claim that what appear to be selective deficits for specific categories/domains of conceptual knowledge can derive from random damage to a distributed semantic system. Category and domain structure emerges out of this system as a function of the structure and content of individual concepts (Gonnerman *et al*., 1997; Tyler *et al*., 2000).

Recent functional neuroimaging studies (PET and fMRI) investigating the neural representation of nouns and verbs have been interpreted as supporting the claim for regional specialization, with the left inferior prefrontal cortex being specialized for verbs and left temporal cortex specialized for nouns (Petersen *et al*., 1988; Perani *et al*., 1999). However, the results are, in fact, very mixed and do not support this straightforward dissociation unequivocally. Almost all of the imaging studies use a verb generation task in which subjects produce a verb in response to a target noun. One of the few imaging studies which did not use the verb generation task is the lexical decision study by Perani and colleagues comparing nouns and verbs (Perani *et al*., 1999). This provides the clearest evidence for cortical specialization for verbs. When subjects were asked to make lexical decisions to written words and non-words, the perisylvian language areas were activated (Wise *et al*., 1991) for both nouns and verbs. However, verbs produced additional activation over nouns in the left dorsolateral frontal cortex, superior parietal and anterior and middle temporal lobes. Other studies, primarily using the verb generation task (e.g. Petersen *et al*., 1988; Martin *et al*., 1995), do not replicate this finding. Although the initial study by Petersen and colleagues (Petersen *et al*., 1988) reported additional activation for verbs primarily in the left prefrontal cortex, this region could not be claimed to be specialized for verb processing since, in a subsequent study, the same pattern was obtained in a name generation task (Petersen *et al*., 1989). Similarly, even though Martin and colleagues found greater activation in left prefrontal cortex when subjects generated action words (verbs) compared with object names, this same region was also activated when colour words (adjectives) were generated, although to a lesser extent (Martin *et al*., 1995). Other studies report no differences between nouns and verbs (Warburton *et al*., 1996; Buckner *et al*., 2000). In the study of Warburton and colleagues, a large left lateralized network was activated for nouns and verbs, including temporal, parietal and prefrontal regions (Warburton *et al*., 1996). This network was activated more highly for verbs than nouns, but there was no regional specificity for one type of word compared with another. Similarly, Buckner and colleagues show that verb generation and word stem completion (involving nouns) activate the same cortical regions (Buckner *et al*., 2000). They found left prefrontal activation extending from both the inferior and dorsal parts of the inferior frontal gyrus into more anterior prefrontal regions and including left inferior temporal cortex.

Variation in the pattern of activations in functional imaging studies may be due partly to problems in the selection of experimental stimuli, and to the limitations of the verb generation task. We know from numerous behavioural cognitive studies that stimulus factors such as word length, familiarity, frequency and imageability play a crucial role in word recognition and language comprehension (Rubin, 1980; Gernsbacher, 1984). Since these are important variables in behavioural studies, it is likely that they will also affect performance in imaging studies and thus sets of stimuli need to be matched on these properties. However, since stimulus characteristics are rarely specified in imaging papers on verb–noun differences, it is impossible to determine whether materials were in fact matched in the appropriate ways, although there is no indication in most published reports that they are.

Another set of problems is raised by the use of the verb generation task. This task is used in most of the studies exploring noun–verb differences, where activation resulting from verb generation is often compared with noun generation or stem completion (e.g. Warburton *et al*., 1996; Bruckner *et al*., 2000). In the verb generation task, subjects initially are given a noun cue and asked to produce a related verb. Thus, in a single trial, they encounter both nouns and verbs, and both will contribute towards activation, so the task cannot be thought of as selectively tapping verb processing. Secondly, the relationship between the noun and verb is left unspecified, as is the means by which verb–noun pairs are selected; consequently, the sets of word pairs may not be sufficiently homogeneous. Finally, form–class ambiguity, i.e. the fact that many verbs also function as nouns and vice versa, is ignored. These methodological issues cast doubt on the reliability of the findings from previous studies, and provided the motivation for the two PET studies presented here.

The aim of the two PET experiments reported here was to determine whether there is neural specialization for nouns and verbs. We attempted to overcome some of the limitations of earlier studies by matching nouns and verbs on the relevant variables of letter length, familiarity and imageability. In addition, we used two different behavioural paradigms. In the first study, we used visual lexical decision since this task is used widely in activation studies of lexical and semantic processing, and is clearly sensitive to semantic processing (Price *et al*., 1996; Perani *et al*., 1999). In a second experiment, we used a semantic categorization task which requires more effortful control of semantic information and is thus predicted to activate the left inferior prefrontal cortex (Wagner *et al*., 1997; Buckner *et al*., 2000), the area that has been implicated most strongly in verb deficits and verb processing. In addition, the task demands of the semantic categorization task are more comparable with the verb generation task, while avoiding some of the pitfalls associated with verb generation. In the semantic categorization task, three written cue words are presented sequentially, and subjects make decision as quickly as possible about whether a fourth (target) word belongs to the same category as the cue words (Devlin *et al.*, 2000). Both the semantic categorization and verb generation tasks require effortful semantic processing. In the case of verb generation, response selection is required, while for semantic categorization subjects have to form a conceptual representation of the three cue words, hold this in memory and compare it with a target word. Since the semantic categorization task should activate left inferior prefrontal cortex more strongly than the lexical decision task, we predicted that we would then be in a position to see whether there was more activation in this area for verbs than nouns.

Methods

Subjects

Subjects were all native English speakers aged between 18 and 40 years. There were nine subjects in Experiment 1, ranging in age between 21 and 34 years (mean ± SD, 26 ± 5), and eight subjects ranging in age between 21 and 30 years (25 ± 3) in Experiment 2. All participants in the imaging studies were right-handed, neurologically normal males. Subjects had no prior knowledge of the study, and written informed consent was obtained from all subjects according to protocols approved by the Cambridge Local Ethical Research Committee.

Imaging methods

In both experiments, there were 12 scans per subject, eight test scans (two for each of the four conditions) and four baseline scans. Stimuli were presented using DMDX (Forster and Forster, 1990) running on a Dell PC (Pentium II, 400 MHz) running Windows 98. Items were presented in black 26 point Arial font against a white background on a video monitor ~1 m from the subjects' heads. The room was dimly lit. In all tasks, subjects responded with a right hand two-choice button press. Accuracy and reaction times for both types of task were recorded to the nearest millisecond.

Scans were performed at the Wolfson Brain Imaging Centre in Cambridge, UK on a GE Advance PET Scanner (General Electric Medical Systems, Milwaukee, Wisc., USA). It comprises 18 rings of crystals, which results in 35 image planes, each 4.25 mm thick. The axial field-of-view was 15.3 cm, thus allowing for whole brain acquisition. Each subject received a bolus of 300 MBq before each scan for a total radiation exposure of 4.2 mSv. Stimuli were presented during the first 45 s, to coincide with the critical period of tracer uptake (Silbersweig *et al.*, 1993). The emission data were acquired with the septa retracted (3D mode) and reconstructed using the PROMIS algorithm (Kinahan and Rogers, 1989) with an unapodized Colsher filter. Corrections were applied for randoms, scatter, attenuation and dead time. The voxel sizes were 2.34 mm × 2.34 mm × 4.25 mm.

Experiment 1: lexical decision task

Methods

Two tasks were used in this experiment, one for test conditions and another for baseline conditions. The test conditions used lexical decision, where subjects decide whether a string of letters forms a word in English. The baseline conditions used a letter detection task in which subjects decided whether a string of letters (which did not form a word in English) contained the letter 'x'. This task required subjects to make a timed response without any semantic processing.

Stimuli for lexical decision conditions consisted of words and orthographically legal, pronounceable non-word letter strings (e.g. hicton, blape). There were a total of 240 test words. These were divided into four equal groups: high imageability nouns, high imageability verbs, low imageability nouns and low imageability verbs. We included imageability as a factor in this study since verbs tend to be more abstract than nouns, and many behavioural studies have shown differences between abstract and concrete words, with concrete words showing an advantage relative to abstract words (we use the terms imageability and concreteness interchangeably here since they are highly correlated). Also, concrete and abstract words often dissociate following brain damage (Franklin, 1989), suggesting that they form distinct domains of knowledge, and imaging studies suggest that abstract words may impose greater processing demands on the semantic system (Kiehl *et al.*, 1999). By ensuring that nouns and verbs are matched on imageability, we avoid any potential confounding effects of this variable, and can also evaluate its contribution to levels of activation.

Words in the four conditions were between three and eight letters long (inclusive). Items within the two sets of concrete words were matched for imageability, as were items within the two sets of abstract words. Imageability ratings were obtained either from the Coltheart database (Coltheart, 1981) or from norms collected from 15 young normal controls from the Cambridge University Centre for Speech and Language subject pool who rated the 250 words from 1 (not imageable) to 7 (highly imageable). Because most of our pre-test words were of low imageability, an additional 38 items with ratings of 600 or higher from the Coltheart database were included in order to encourage the subjects to use the entire range of scores. The items had an average Coltheart rating of 628, and were given an average rating of 658 in the pre-test, suggesting rating consistency between the two sets. In total, 94 items from this locally collected set were used as experimental stimuli in our imaging study, 41 concrete verbs and 53 abstract verbs. We also matched the sets on their frequency of occurrence as a member of a specific word class (i.e. as nouns or verbs). Many words in English assume different

Table 1 *Statistics for the four sets of stimuli in the lexical decision task*

Scan condition	Example	Word length		Imageability*		CELEX frequency	
		Mean	SD	Mean	SD	Noun	Verb
Concrete noun	Sand	5.27	1.2	534	33	52	2
Concrete verb	Spill	5.23	1.2	529	38	3	51
Abstract noun	Duty	5.28	1.1	345	40	55	2
Abstract verb	Lend	5.27	1.1	341	36	2	51

*Range: 100 (minimum score)–700 (maximum score).

word classes in different contexts (cf. 'Don't *break* the china' and 'The blow made a clean *break*'). Because of this, it was important to select nouns which were unambiguously nouns or which had a much higher frequency as a noun than as a verb. Similarly, we selected verbs which were unambiguously verbs or which had a much higher frequency as a verb than as a noun according to the CELEX database (Baayen and Pipenbrook, 1995). The non-words were matched with the real words for number of letters. A summary of the stimuli statistics appears in Table 1.

Stimuli for the letter detection baseline condition were created by randomly generating sets of consonant strings according to a set of pre-specified parameters: mean string length, standard deviation of string length, number of strings and proportion of strings with 'x's. String length in the baseline conditions was matched with word length in the test conditions, and the proportions of strings containing 'x's to strings without 'x's was matched with the proportion of words to non-words in the test conditions.

Each of the four test conditions was split in half such that there were 30 words for each of the two scans of a given condition. The two scans for each condition were matched for word length, frequency and imageability. Each scan lasted 90 s, during which time 36 letter strings (30 words and six non-words) were presented at a rate of one every 2.5 s (we included only a small number of non-words in the test scanning period to maximize the semantic activation due to processing of the real words). Similarly, each baseline scan had 30 letter strings containing the letter 'x' and six letter strings without an 'x'. Stimuli appeared on screen for 500 ms, followed by 2000 ms of blank screen, during which time subjects made their responses. Of the six non-words, only two occurred during the initial 45 s when tracer uptake occurs. To reduce strategic effects due to this very high ratio of words to non-words, 24 items were presented before the 90 s scan block, and 12 items were presented afterwards. Of these 36 items, 24 were words and 12 were non-words; subjects were unaware of when the scan period began and ended. Similarly, the baseline conditions had 24 strings with an 'x' and 12 without.

Results

Behavioural data

Response times and error rates were recorded for all responses in both the letter detection and the lexical decision tasks to verify that the subjects were performing the tasks correctly. Response times for errors and time out responses were not included in the reaction time analyses. Subjects made very few errors (2% in the lexical decision conditions and 3% in the letter detection condition), indicating that they were performing the task correctly. Response time data were inverse transformed to reduce the influence of outlying data points (Ulrich and Miller, 1994) before conducting inferential statistics. There was a reliable effect of task on response time, with letter detection (mean = 503 ms) faster than lexical decision [mean = 526 ms; $F_2(1,358) = 14.40$, $P < 0.001$]. There was also a reliable effect of word class in the lexical decision conditions, with nouns (mean = 510 ms) faster than verbs [mean = 541 ms; $F_2(1,236) = 22.0$, $P < 0.001$], a finding which is consistent with previous behavioural studies. There was no effect of imageability ($F < 1$) and no interactions between word class and imageability ($F < 1$). We initially ran the experiment on 20 subjects out of the scanner, using exactly the same materials, timing parameters and blocked design. Letter detection was significantly faster than lexical decision and there was no significant difference in response times to nouns and verbs.

Image analysis

The imaging data were analysed in the following way. Functional images were realigned (Friston *et al.*, 1995*a*) as implemented in Statistical Parametric Mapping (SPM99b; Wellcome Institute of Cognitive Neurology; www.fil.ion.ucl.ac.uk). Translation and rotation corrections did not exceed 5 mm and 2°, respectively, for any of the participants. The mean image created for each subject by the realignment procedure was used to determine the parameters for transforming the images of that subject onto the Montreal Neurological Institute (MNI) mean brain. These normalization parameters were then applied to the functional images (Ashburner and Friston, 1997; Ashburner *et al.*, 1997). After normalization, the voxels were isotropic at 2 mm^3. Finally, each image was smoothed with a 16 mm FWHM (full-width half-maximum) Gaussian filter. The SPM software was used to compute a within-subjects analysis (i.e. a fixed effects model) using the general linear model (Friston *et al.*, 1995*b*). One scan was lost from each of two of the nine subjects, yielding 84 degrees of freedom. All results are reported at a

$P < 0.05$ level after correcting for multiple comparisons (Worsley, 1996), in order to reduce the possibility of false positives. (The issue of whether it is more appropriate to use corrected or uncorrected levels of significance is currently a hotly debated topic in neuroimaging research. We prefer to use corrected levels since this reduces the probability of false positives due to large numbers of comparisons over large data sets and gives a better estimate of robust effects. This kind of approach is more compatible with statistical analysis procedures in experimental psychology, and thus ensures that the cognitive behavioural studies and the cognitive activation studies are subject to similarly rigorous statistical analysis procedures.) The following contrasts were used in the data analysis.

(i) To determine regions common to all four lexical decision conditions compared with letter decision conditions, we masked the main effect of lexical conditions versus baseline with each of the lexical conditions individually versus baseline. This contrast is: (CV + AV + CN + AN) – B masked with (CV – B), (AV – B), (CN – B), (AN – B), where regions of the main effect are only considered if they surpass $P < 0.05$ in each of the masking contrast images (CV = concrete verbs; AV = abstract verbs; CN = concrete nouns; AN = abstract nouns; B = baseline). This enables us to determine which regions are commonly activated in all conditions. A less conservative approach is to analyse all of the conditions minus the baseline without masking. We carried out such analyses for the semantics−baseline and the form class and imageability contrasts and found no areas of significant activation that were not also present in the masked analyses. The only difference between the two types of analysis was that the unmasked analysis produced more significant activations with larger extents than in the masked analyses. However, we have only reported the masked analyses in the text since the potential problem with unmasked contrasts is that they can mis-identify regions of common activation which, in fact, are only activated by a single condition.

(ii) Main effect of word class: to determine whether there were any regions which were only activated in response to verbs, we contrasted the two verb categories (CV + AV) with the two noun categories (CN + AN) and masked this contrast with the four simple contrasts (CV–CN, CV–AN, AV–CN, AV–AN) and the two baseline contrasts (CV–B, AV–B). We carried out comparable analyses to look for regions specific to nouns.

(iii) Main effect of imageability: to assess the effects of abstract relative to concrete words, we contrasted the two abstract categories (AV + AN) with the two concrete categories (CV + CN) masking with the four simple contrasts as for the word class effects described above. Comparable analyses were carried out to investigate the effects of concrete words relative to abstract words.

(iv) Cross-over interactions between word class and imageability: (CV + AN) – (CN + AV) and the reverse.

(v) Other interactions between word class and imageability: CV – (AV + CN + AN), AV – (CV + CN + AN), CN – (CV + AV + AN) and AN – (CV + AV + CN). This collection of contrasts allowed us to consider overall effects of the experimental conditions versus baseline, main effects of word class and imageability, and interactions between them.

The lexical decision task relative to the baseline showed four areas of activity. These areas (described in Table 2 and shown in Fig. 1) were in the left inferior frontal gyrus, left fusiform gyrus, left middle and inferior temporal gyri and right middle temporal gyrus. [It is interesting to contemplate why the left fusiform is insignificant at the cluster level but significant at the voxel level. The multiple comparison correction is based on statistical threshold based on the probability of observing a cluster of height *s* and extent *k* by chance. This is directly related to random fields theory. An almost isolated voxel (very small cluster) with a relatively high signal height is likely to be an incorrectly activated area and thus reflect a false positive. Conversely, large clusters with low signal height are likely to be due to chance and are then rejected at the voxel level. This may explain what happened to the left fusiform which is rejected at the cluster level but is still significant at the voxel level.] After correction for multiple comparisons (Worsley *et al.*, 1992; Friston *et al.*, 1994), only the inferior frontal gyrus and fusiform gyrus activations were significant at the voxel and cluster level. The other two regions of activation were significant only at the cluster level. The areas of cortex activated by the lexical conditions in contrast to the baseline were quite robust and overlap with areas found in other studies (Mummery *et al.*, 1996, 1998; Devlin *et al.*, 2000).

Table 2 shows the results of the form class and imageability contrasts. There were *no* areas of activity that were significant at either the voxel or the cluster level after corrections for multiple comparisons. The areas of maximal change in signal intensity, although not significant, are also shown in Table 2. Thus, there is no evidence for cortical regions specific to the processing of nouns or verbs, or to abstract or concrete words. This was the case in spite of the behavioural data which showed a response time advantage for nouns over verbs, a result which is consistent with other behavioural experiments.

Power analyses

We carried out further analyses to ensure that the experiment had sufficient power to detect small differences in activation, if they were present in the data. In contrast to an introspective approach that uses the voxel scores (possibly outliers) as typical observations to calculate the power of the analysis (Van Horn, 1998), we used a more stringent retrospective approach using the whole data set to estimate the effect size and the proportion of active voxels. We applied a finite mixture-based model to fit a receiver operating characteristic (ROC) curve to the observed data (Gustard *et al.*, 2000).

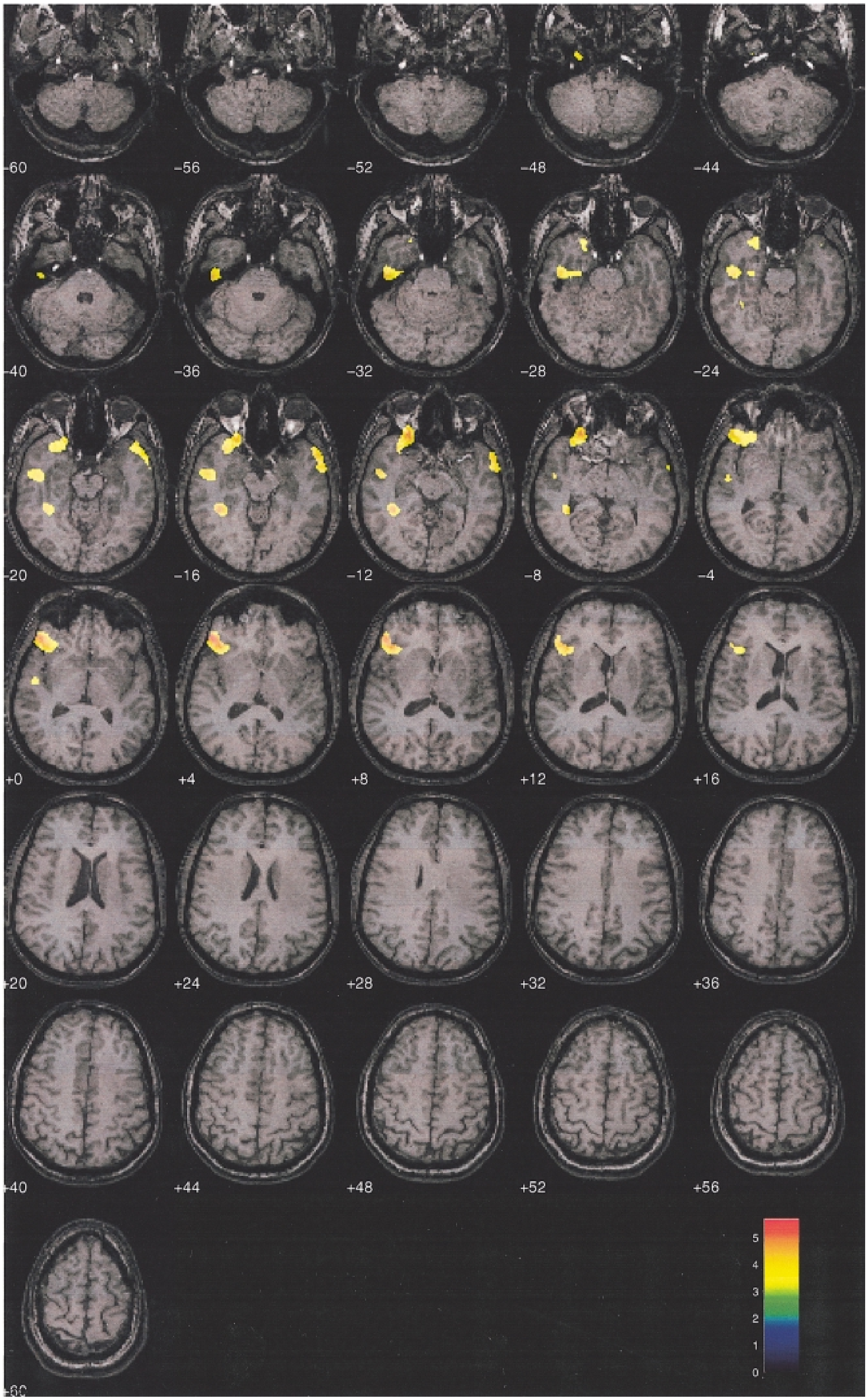

Fig. 1 The areas activated in the semantic–baseline contrasts in the lexical decision task (Experiment 1). Activations are shown superimposed on a structural image. Red and yellow areas were reliably ($P < 0.05$) active after statistical correction.

Table 2 *Brain areas of activity in the lexical decision task*

Region	Coordinates			Voxel level		Cluster level	
	x	y	z	$P_{\text{corrected}}$	t	$P_{\text{corrected}}$	Extent
Contrast (i), test conditions−baseline masked by each condition – baseline contrast							
Frontal							
L inf. gyrus	–52	32	4	0.002	5.68	0.000	1453
(BA 45, 47)	–28	32	–10	0.026	5.01		
	–40	16	–14	0.117	4.53		
Temporal							
L fusiform gyrus	–46	–48	–14	0.008	5.34	0.112	252
(BA 37)							
R mid. temp. gyrus	58	6	–14	0.135	4.49	0.063	315
(BA 21)	62	–6	–14	0.286	4.21		
L mid. and inf. temp gyrus	–60	–8	–10	0.144	4.46	0.004	670
(BA 21, 20)	–52	–14	–28	0.185	4.38		
	–58	–10	–2	0.408	4.06		
Contrast (ii), main effect of word class							
Verbs–nouns							
R substantia nigra	12	–22	–6	0.808	3.64	0.958	10
No effect of nouns–verbs							
Contrast (iii), main effect of imageability							
Concrete–abstract							
Superior frontal gyrus	–16	–64	–2	0.989	3.22	0.988	1
(BA 10)							
Abstract–concrete							
Cingulate gyrus	–18	–42	28	0.598	3.87	0.585	79
(BA 31)							

Results presented for both the voxel and cluster levels of significance were based on a height threshold of 0.001. The coordinates are in MNI (Montreal Neurological Institute) space, and the corrected P values, the SPM$\{t\}$ and the extent of the activation are presented. Multiple peaks within an extent are shown on subsequent lines. L = left; R = right; temp. = temporal; mid. = middle; inf. = inferior. We include Brodmann areas (BAs) of the active regions to assist the reader in quickly determining the locations of these activations. They are not intended as commentaries on the underlying cytoarchitecture of the active areas.

This model overcomes the binormal assumption of the classical parametric methods and can then be generalized to any statistical score, and it is more flexible than non-parametric models (Genovese *et al.*, 1997) since it does not require a minimal number of replications. Therefore, this model is directly applicable in our context using either the t or the F SPMs. Additional details are given in the Appendix I.

The data (provided by the t maps given in SPM) were fitted with a mixture of two t distributions representing the t values of individual voxels under the null hypothesis (H_0) and under an active hypothesis (H_1). The algorithm provides us with the mixing proportion (the estimated proportion of activated voxels) and effect size, which best fit the data. These were estimated using a maximum likelihood procedure described by Gustard and colleagues (Gustard *et al.*, 2000).

For the effect of words relative to the baseline, we found $\lambda = 0.0545$ and $\delta = 2.6466$, i.e. the mean effect size elicited an ~2.65% change in regional cerebral blood flow (rCBF) and 5.5% of voxels were activated in this condition. (For each analysis, we fit the mixture model to the SPM output relevant for that contrast. For example, for the analysis of word class contrasts, the data entered into the model were the masked analyses reported earlier.) The detection power was of 60% at a false positives fraction of 0.008. This point was detected automatically in order to maximize the weighted sum of true positives and true negatives in the SPM. For the verb–noun contrast, we obtained values of $\lambda = 0.1$ and $\delta = 1.9$. Consequently, the power deduced using the same criterion as above was lower (42%) with a higher false positives fraction (0.02). Interestingly, when this technique was applied to the nouns versus verbs contrast, the minimization algorithm failed to find a mixture which fitted the histogram. In other words, no significant alternative distribution was found, thus confirming that there are no significant differences in neural activation as a function of word class. One potential problem with these analyses is that we might not have sufficient statistical power to detect small differences. We think that it is unlikely that lack of power can explain our findings since the number of degrees of freedom (subjects, sessions) was sufficiently high to ensure reliable performance of our signal detection algorithm.

Table 3 *Sample stimuli for the semantic categorization task*

Cue 1	Cue 2	Cue 3	Target	Response
Semantic categorization				
bucket	basket	bin	TUB	Same
writhe	slither	wriggle	SQUIRM	Same
aisle	alley	lane	BADGE	Different
radiate	glisten	dazzle	SPIN	Different
Letter categorization				
aaaaaa	aaaaa	aaaa	AAAAAA	Same
sssss	ssss	sssssss	SSS	Same
lllll	llllll	lll	YYYYY	Different
ddd	ddddddd	dddddd	RRRR	Different

Experiment 2: semantic categorization task

Methods

Two tasks were used in this experiment, one for test conditions and the other for baseline conditions. For the test conditions, we used a semantic categorization task in which subjects read three cue words presented one after another on a computer screen and then made a decision as quickly as possible about whether a fourth (target) word belonged to the same category as the cue words. Cue words were in lower case and the target words were in upper case to signal to the participants when to make a response. For instance, subjects made a 'same' response to the target OTTER when it followed the three cue words 'dolphin, seal, walrus' by pressing the left mouse button, and a 'different' response to CUP when it followed 'moccasin, sandal, boot' by pressing the right mouse button (see Table 3). The mouse was always held in the subject's right (dominant) hand and both reaction times and accuracy were recorded.

For the baseline task, we used a letter categorization task which shared the same stimulus and response characteristics as the test task, but had no lexical or semantic component. Instead, subjects were presented with three strings of letters, matched in length to the word stimuli, and were asked whether a fourth string, in upper case, was made up of the same letters. For example, 'fffff, fff, ffffffff, FFFFFF' constituted a 'same' trial and 'ttttt, tttttt, tttt, HHHH' was a 'different' trial. Again, subjects signalled their response by pressing either the left or the right mouse button.

Since Experiment 1 established that imageability did not significantly affect activations, in this study we decided to manipulate word frequency rather than imageability. Word frequency is another major variable reputed to affect lexical processing in a wide range of psycholinguistic tasks (e.g. Rubenstein *et al*., 1970). We constructed four sets of stimuli: high frequency nouns, high frequency verbs, low frequency nouns and low frequency verbs. The test stimuli were matched across conditions for imageability and letter length. As in Experiment 1, verbs in the two verb conditions had a much higher frequency as verbs than as nouns, and stimuli for the two noun conditions were much more frequent as nouns than verbs. (We pre-tested the quadruplets to ensure that the component words were equally semantically related across the four conditions. In the pre-test, 15 subjects were presented with each quadruplet and asked to rate them for their semantic relatedness on a scale of 1–7, where 1 was very unrelated and 7 was highly related. Subjects were also asked to circle any word which they thought did not fit into the set.) Thus, we used a 2 × 2 design with word class and frequency as the main variables. Stimuli statistics are described in Table 4. The items in the baseline letter categorization task were matched on letter length with those in the semantic categorization task.

The experiment was first run in a pilot study outside the scanner in order to determine presentation rates and durations such that the task was challenging for participants but allowed fast and accurate responses. As a result of this pre-testing, in the final version of the experiment each cue word (or letter string) was displayed on a computer screen using DMDX software (Forster and Forster, 1999) for 200 ms with a 400 ms inter-stimulus period. The target word (or letter string) was also presented for 200 ms and response times were measured from target onset. There was a 1750 ms delay following the target word to allow time for participants to make a response. Thus each trial lasted 3750 ms.

The eight subjects participated in twelve 90-s scans, eight of the semantic categorization condition and four of the letter categorization condition. To coincide with the critical period of tracer uptake (Silbersweig *et al*., 1993), subjects received 45 s of stimuli (12 trials) followed by a blank screen for the remaining 45 s of the scan, during which they were asked to relax and clear their mind. The semantic and letter conditions were presented systematically such that no subject saw the conditions in the same order.

Results

Behavioural data

Response times were recorded and analysed in the same way as in Experiment 1. Subjects' responses to 'same' (mean response time = 581 ms) and 'different' trials (mean response time = 579 ms) were not statistically different ($F < 1$) and few errors were made in either type of trial (4.5 and 4% errors, respectively). Analyses of inverse transformed response times were performed on all stimuli (both same and different trials). There was a reliable effect of task on response time, with letter decision (mean = 527 ms) faster than semantic decision [mean = 703 ms; $F_2(1,140) = 309.63$, $P < 0.001$]. There was also a reliable effect of word class in the semantic decision task, with nouns (mean = 662 ms) faster than verbs [mean = 747 ms; $F_2(1,94) = 11.84$, $P < 0.01$]. There was no effect of frequency ($F < 1$) or interaction between word class and frequency ($F < 1$). These results indicate that the letter decision task was easier to perform than the semantic decision task, consistent with subject reports and previously reported data (Devlin *et al*., 2001). Consistent with the first experiment, the verb items

Table 4 *Statistics for the four sets of stimuli in the semantic categorization task*

Scan condition	SR	Word length		Imageability		CELEX frequency	
		Mean	SD	Mean	SD	Noun	Verb
High frequency nouns	5.6	5.5	0.9	523	28	111	
Low frequency nouns	5.4	5.4	0.7	531	32	26	
High frequency verbs	5.6	5.3	0.9	516	17		89
Low frequency verbs	5.8	5.6	0.9	520	28		21

SR = semantic relatedness.

were slightly more difficult than the noun items, even though the conditions were carefully matched.

Imaging data

The imaging data were analysed in the same way as for the previous study. Translation and rotation corrections did not exceed 3 mm and 1.5°, respectively for any of the participants. No scans were lost from the eight subjects, yielding 76 degrees of freedom. All results are reported at a $P < 0.05$ level after correcting for multiple comparisons (Worsley, 1996). The following contrasts were used in the data analysis.

(i) Regions common to all semantic categorization conditions in comparison with letter categorization conditions. This was effected by masking the main effect of the semantic conditions versus baseline with each of the semantic conditions individually versus baseline. This contrast is: (HN + LN + HV + LV) – B masked with (HN – B), (LN – B), (HV – B), (LV – B), where regions of the main effect are only considered if they surpass $P < 0.05$ in each of the masking contrast images (H = high frequency; L = low frequency; N = noun; V = verb; and B = baseline).

(ii) Main effect of word class: to look for effects of verbs relative to nouns, we contrasted the two verb categories (HV + LV) with the two noun categories (HN + LN) and masked this contrast with the four simple contrasts (HV–HN, HV–LN, LV–HN, LV–LN) and the two baseline contrasts (HV–B, LV–B). We carried out comparable analyses to look for effects of nouns relative to verbs.

(iii) Main effect of frequency: to assess the effects of high frequency relative to low frequency words, we contrasted the two high frequency categories (HV + HN) with the two low frequency categories (LV + LN) masking with the four simple contrasts as for the word-class effects described above. Comparable analyses were carried out to investigate the relative effects of low frequency words relative to high frequency words.

(iv) Cross-over interactions between word class and frequency: (HV + LN) – (HN + LV) and the reverse.

(v) Other interactions between word class and frequency: HV – (LV + HN + LN), LV – (HV + HN + LN), HN – (HV + LV + LN), LN – (HV + LV + HN). This collection of contrasts allowed us to consider overall effects of the experimental conditions versus baseline, main effects of word class and frequency, and interactions between them.

The semantic categorization task relative to baseline produced three areas of reliable activity, one in the right inferior frontal gyrus, another in the left middle temporal gyrus and a massive left perisylvian region extending from the inferior frontal gyrus to the inferior temporal gyrus (see Table 5 and Fig. 2). Within this large region, there were distinct peaks in the left inferior frontal gyrus and inferior temporal gyrus. All of these regions of activity were reliable at both voxel and cluster levels after correction for multiple comparisons.

In the verb–noun contrast, there was a single area of activation in the right inferior temporal gyrus. However, it disappeared when masked by the appropriate contrasts (see Image analysis above), as the area was not active in all the verb conditions. Similarly, there were no areas more active for all nouns than verbs. In the frequency analyses, there were also no areas more active for all high versus all low frequency words, or vice versa. In the high frequency nouns minus other test conditions contrast, there was a trend toward significant activation in a region of the left middle temporal gyrus. This activity remained after masking the contrast with each of the high frequency minus other conditions contrasts. Thus the region shows an interaction where effects of word class are modulated by frequency.

As with the first experiment, we carried out a power analysis to determine whether the task was sufficiently sensitive to detect differences in activation. For the main contrast of words relative to the baseline, we found $\lambda = 0.102$ and $\delta = 3.17$. Again, this means that the average effect size was a ~3.17% change in rCBF and 10% of voxels were activated in this condition. The detection power was high (80%) at a false positives fraction of 0.01. For the verbs–nouns contrast, the algorithm converged with $\lambda = 0.0944$ and $\delta = 2.97$. The power deduced using the same criterion as before was slightly lower (76.5%), with a false positives fraction of 0.012. When the algorithm was applied to the nouns–verbs contrast, the minimization procedure successfully converged, this time with a mixing proportion of 0.0664 and an effect size of 2.5. The power corresponding to the chosen reliability criterion was of 58% at a false positives fraction of 0.01. These results confirm that the

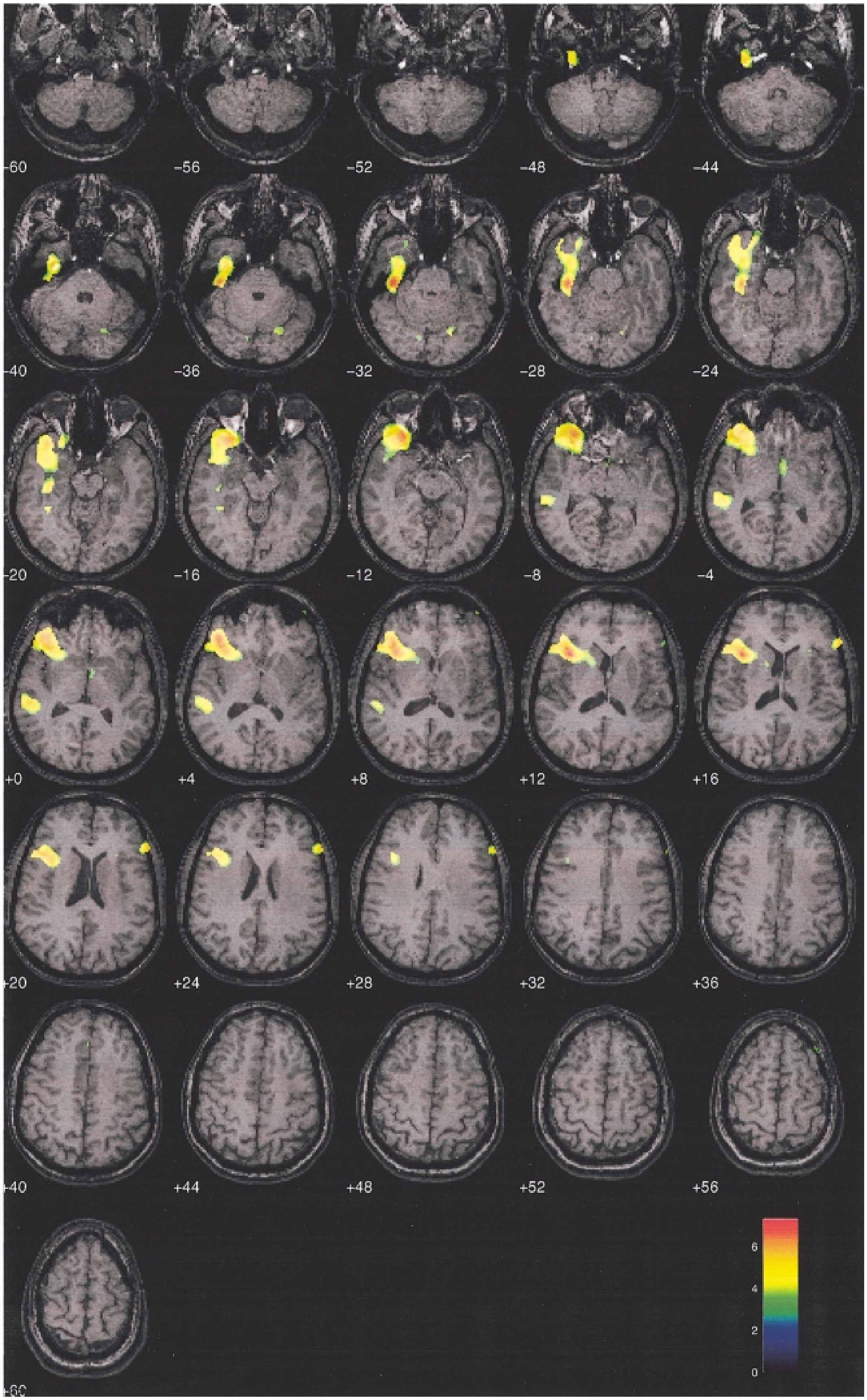

Fig. 2 The areas activated in the semantic–baseline contrasts in the semantic categorization task (Experiment 2). Activations are shown superimposed on a structural image. Red and yellow areas were reliably ($P < 0.05$) active after statistical correction.

Table 5 *Brain areas of activity in the semantic categorization task*

Region	Coordinates			Voxel level		Cluster level	
	x	y	z	$P_{\text{corrected}}$	t	$P_{\text{corrected}}$	Extent
Contrast (i), test conditions−baseline masked by each condition−baseline contrast							
Frontal and temporal							
L inf. frontal gyrus (BA 44, 45, 47)	–42	16	12	0.000	7.30	0.000	6171
L inf. temp. gyrus	–44	–24	–30	0.000	7.22		
(BA 20)	–46	30	4	0.000	6.98		
L mid. temp. gyrus (BA 21)	–62	–38	–2	0.008	5.42	0.006	560
R inf. frontal gyrus (BA 45)	60	26	20	0.008	5.43	0.160	202
Contrast (ii), main effect of word class							
Verbs–nouns							
BA 20/37	–48	–44	–26	0.649	3.87	0.803	42
Nouns–verbs: no effects							
Contrast (iii), looking at the main effect of frequency[†]							
Contrast (iv), non-cross-over interactions							
High frequency nouns – all others, masked with high frequency nouns – each other condition							
Temporal							
L mid. temp. gyrus (BA 21)	–62	–22	–10	0.098	4.66	0.056	308

Results presented for both the voxel and cluster levels of significance were based on a height threshold of 0.001. The coordinates are in MNI space, and the corrected P values, the SPM$\{t\}$ and the extent of the activation are presented. Multiple peaks within an extent are shown on subsequent lines. L = left; R = right; temp. = temporal; mid. = middle; inf. = inferior. [†]There were no significant effects of high versus low frequency or vice versa.

second experiment provided more robust activation and was thus more powerful than the first experiment for all the analysed contrasts.

Analyses comparing Experiments 1 and 2

Comparing Tables 2 and 5, we see that the semantic categorization task generally produced more robust activations with larger extents than the lexical decision task. To investigate statistically areas of overlap between Experiments 1 and 2, we analysed both data sets together. When the lexical decision–baseline contrast of Experiment 1 and the semantic decision–baseline contrast of Experiment 2 are analysed together, there are five reliable areas of activity. All of these areas are in the left hemisphere and are reliable at the voxel level (cluster analyses are not possible because the data from the two experiments do not have the same intrinsic smoothness) after correction for multiple comparisons. These regions are in the fusiform gyrus, inferior temporal gyrus, temporal pole, uncus and inferior frontal gyrus (see Table 6). These regions are largely the intersection of those of Experiments 1 and 2. Notably, there are no reliable effects of word class across both experiments.

Discussion

In the two PET studies reported here, we found robust activation in the left prefrontal cortex and inferior temporal lobe, associated with semantic processing of written words. These areas have been reported in previous studies of semantic processing and are thought to constitute a network which is activated when subjects are engaged in a variety of tasks involving semantics (Price *et al*., 1994; Rumsey *et al*., 1997).

Effects of lexical variables

We found no significant differential activation as a function of the two lexical variables—imageability and frequency—when we corrected for multiple comparisons. This suggests that the neural representation of conceptual knowledge is not specialized as a function of these variables; for example, there is no cortical region which is specialized for abstract words or high frequency words. This makes sense when one considers that the variables of frequency/familiarity and imageability are continuous rather than discrete and thus do not lend themselves to principles of categorical organization. While it might be simple to categorize a highly abstract or concrete word, many words are of medium imageability according to the norms that are used typically. It is more

Table 6 *A conjunction of the lexical decision–baseline contrast of Experiment 1 and the semantic decision–baseline contrast of Experiment 2*

Region	Coordinates			Voxel level		Cluster extent
	x	y	z	$P_{\text{corrected}}$	t	
Contrast (i): a conjunction of the lexical decision–baseline contrast of Experiment 1 and the semantic decision–baseline contrast of Experiment 2, masked by each of the test condition–respective baseline condition contrasts						
Frontal and temporal						
L inf. frontal gyrus (BA 45, 47)	–50	30	4	0.000	5.80	1373
	–30	30	–10	0.000	4.76	
	–38	20	–16	0.000	4.15	
L fusiform gyrus (BA 37)	–44	–46	–18	0.000	4.39	65
L. inf. temp. gyrus (BA 20)	–48	–18	–34	0.000	3.91	204
	–36	–14	–32	0.001	3.79	
	–48	–8	–22	0.093	2.92	
L temp. pole (BA 20, 38)	–34	0	–46	0.001	3.66	31
Contrast (ii): main effect of word class*						

The results are presented for the voxel level of significance based on a height threshold of 0.001. The coordinates are in MNI space, and the corrected P values, the SPM$\{t\}$ and the extent of the activation are presented. Multiple peaks within an extent are shown on subsequent lines. *No significant effects were found.

difficult to see how these words would be categorized if the neural representation of conceptual knowledge were organized by imageability.

Previous studies have reported additional activation in right inferior frontal gyrus for abstract words (Beauregard *et al*., 1997; Kiehl *et al*., 1999), challenging earlier cognitive claims for the differential representation of abstract and concrete words based largely on split visual field experiments, which propose that abstract words are only represented in the left hemisphere while concrete words are represented bilaterally (e.g. Paivio, 1991). However, left hemisphere activation (in inferior prefrontal cortex) found in imaging studies typically is associated with increased processing demands (Buckner *et al*., 2000), and there is evidence from cognitive studies that abstract words may be more difficult to process than concrete words. For example, in lexical decision tasks outside the scanner, abstract words generate slower latencies and more errors than concrete words (James, 1975; Kroll and Merves, 1986). Thus, the additional processing in the right prefrontal cortex associated with abstract words may be due to the extra demands involved in processing these words. This is consistent with the behavioural data from the imaging study of Kiehl and colleagues which showed longer latencies for abstract compared with concrete words (Kiehl *et al*., 1999). In our lexical decision PET study, there was no difference in the response latencies of abstract and concrete words, and no additional activation in right prefrontal cortex, suggesting that when abstract and concrete words are matched for processing difficulty, there is no differential activation associated with processing one or other type of stimulus. This in turn suggests that differential activation found in right inferior frontal gyrus is associated with processing difficulty.

Our data suggest that there is no neural specialization for abstract and concrete words. How does this comport with the claim that deep dyslexic patients, who typically have damage to left inferior frontal cortex, have greater impairments for abstract words compared with concrete words (e.g. Franklin, 1989)? We have argued (Tyler *et al*., 1995) that abstract word deficits in deep dyslexic patients tend to be found only when patients are required to make a verbal response. When tasks are used which do not require language production, the disadvantage for abstract words tends to disappear. For example, we tested two classic deep dyslexic patients, with extensive left inferior frontal damage, who showed the typical pattern of greater problems in reading aloud abstract than concrete words (Tyler *et al*., 1995). These same patients were then tested on a semantic priming task where we varied the imageability of the word pairs. The word pairs were either concrete (*street–road*) or abstract (*luck–chance*) synonyms. The patients showed significant semantic priming for both the abstract and concrete words, supporting the claim that their difficulties with abstract words were restricted to tasks requiring a verbal output. The association of phonological output deficits with damage to left inferior frontal cortex is consistent with claims for the role of this region in phonological processing.

Effects of word class

The main aim of our imaging studies was to determine whether processing the meaning of nouns and verbs is associated with activation in different cortical regions. We found no consistent evidence for the differential effects of word class; both nouns and verbs activated areas in left inferior frontal and temporal cortex. Thus, although our behavioural data and a number of cognitive studies suggest that nouns are processed more readily than verbs, this processing advantage does not appear to be correlated with differential neural activation. That is, differences in the acquisition and processing of nouns and verbs do not seem to be correlated with differential neural representation. Moreover, the power analyses showed that our experimental paradigms, especially the semantic categorization task, were sufficiently sensitive to have detected small differences in activation if they had been present.

This set of results is consistent with those cognitive accounts which postulate that the conceptual system is represented within an undifferentiated distributed network with no explicit category or domain structure. Indeed, our recent conceptual structure model (Durrant-Peatfield *et al.*, 1997; Tyler *et al.*, 2000) makes the explicit prediction that, to the extent that the cognitive architecture is reflected in the neural architecture, there will be no regional specialization for different categories/domains of conceptual knowledge. This prediction is supported by the results of both the current studies and other recent PET studies in which we looked for cortical specialization for the domains of living and non-living things (Devlin *et al.*, 2001). In these studies, we found, much as we do in the present studies, an extensive semantic network which was activated by all concepts and no regional specialization as a function of domain. These findings are also compatible with Damasio and Tranel's claims that concepts are not represented in a permanent form in one neural site, but 'they depend on many interacting networks that hold the potential for reactivation of components of concepts . . .' (Damasio and Tranel, 1993).

How does the lack of regional specialization for verbs and nouns relate to the neuropsychological findings that noun deficits are associated with damage to left temporal lobe whereas verb deficits are associated with damage to left inferior frontal cortex? First, as discussed in the Introduction, there is considerable lack of consistency in these reports. Secondly, many neuropsychological dissociations are restricted to naming deficits and do not involve impairments of conceptual knowledge (e.g. Damasio and Tranel, 1993), thus they do not directly address the issue of how different categories of knowledge are represented in the neural substrate. Instead, these impairments may reflect the consequences of damage to phonological processes (Miceli *et al.*, 1984) or to the processes involved in the mapping between phonology and semantics (e.g. a mediation system; Damasio and Tranel, 1993). Thirdly, not all studies of noun–verb dissociations stimuli matched on relevant variables. Thus, the neuropsychological data do not provide unequivocal evidence that the semantic representations of nouns and verbs are located in distinct cortical regions.

Left inferior prefrontal cortex

We found reliable activation in the left inferior frontal cortex, supporting claims that this cortical region is involved in semantic processing (Fiez, 1997; Buchner *et al.*, 2000). In addition, the semantic categorization task generated more activation in this region than did the lexical decision task. We observed the same pattern in two other PET studies we have reported recently, comparing activations for words denoting living and non-living things using the lexical decision and semantic categorization tasks (Devlin *et al.*, 2001). In the lexical decision task, there was no significant left inferior frontal activation, in contrast to the semantic categorization task which generated highly significant activation across the left inferior frontal cortex. Similarly, using the same semantic categorization task in an fMRI study, we also found robust activation in left inferior frontal cortex (Devlin *et al.*, 2000). These findings are consistent with the proposal that the left inferior frontal cortex is more activated in semantic tasks which require effortful, explicit processing, or when more semantic information has to be held in memory while a subject is deciding upon a response (Gabrieli *et al.*, 1998). Our data are less compatible with the view, recently proposed by Thompson-Schill and colleagues, that the left inferior frontal cortex is associated with selection amongst competing semantic alternatives rather than with retrieval of semantic knowledge (Thompson-Schill *et al.*, 1997). The semantic categorization task, which generated additional activation in the left inferior prefrontal cortex, did not involve selection between alternatives. Instead, subjects had to decide whether a target word was a member of the same grouping as three preceding cue words. This involved accessing the meaning of each word, constructing an appropriate semantic grouping/category and determining whether the target was a member of the same category. This task does not involve the activation of multiple competing alternatives followed by a process of selection, but rather requires subjects to maintain in memory the meaning of each word and the category to which they belong. However, instead of the inferior frontal cortex being specialized for one type of cognitive process, it may be the case that it is involved in a diverse array of cognitive functions, as has been argued recently (Duncan and Owen, 2000).

Conclusions

The results of the two PET studies reported here do not support the claim that nouns are represented in the temporal lobes and verbs in left inferior frontal cortex. Instead, they are more compatible with the alternative hypothesis that the meanings of nouns and verbs are represented within an undifferentiated cortical network which is not divided by category or domain. However, lack of regional specialization does not preclude the possibility that specific cortical regions may be involved to different degrees in the processing of subsets of concepts. These studies focused on the meaning representations of nouns and verbs. However, nouns and verbs differ most strongly in their syntactic function. It remains to be seen whether, when the task emphasizes the syntactic roles of nouns and verbs, we find any evidence of regional specialization.

Acknowledgements

We wish to thank Emma Williams at the WBIC, Joe Devlin and Elizabeth Dick for their help with this research. This research was funded by an MRC programme grant to L.K.T. and W. D. Marslen-Wilson, a McDonnell-Pew grant to L.K.T. and a Wellcome Trust Fellowship to H.E.M.

References

Ashburner J, Friston K. Multimodal image coregistration and partitioning—a unified framework. Neuroimage 1997; 6: 209–17.

Ashburner J, Neelin P, Collins DL, Evans A, Friston K. Incorporating prior knowledge into image registration. Neuroimage 1997; 6: 344–52.

Baayen R, Pipenbrook R. The Celex Lexical Database. Philadelphia: Linguistic Data Consortium, University of Pennsylvania; 1995.

Beauregard M, Chertkow H, Bub D, Murtha S, Dixon R, Evans A. The neural substrate for concrete, abstract, and emotional word lexica: a positron emission tomography study. J Cogn Neurosci 1997; 9: 441–61.

Breedin SD, Saffran EM, Schwartz MF. Semantic factors in verb retrieval. An effect of complexity. Brain Lang 1998; 63: 1–31.

Buckner RL, Koutstaal W, Schacter DL, Rosen BR. Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. Brain 2000; 123: 620–40.

Caramazza A, Hillis AE. Lexical organization of nouns and verbs in the brain. Nature 1991; 349: 788–90.

Coltheart M. The MRC Psycholinguistic Database. Q J Exp Psychol 1981; 33A: 497–505.

Damasio AR, Tranel D. Nouns and verbs are retrieved with differently distributed neural systems. Proc Natl Acad Sci USA 1993; 90: 4957–60.

Daniele A, Giustolisi L, Silveri MC, Colosimo C, Gainotti G. Evidence for a possible neuroanatomical basis for lexical processing of nouns and verbs. Neuropsychologia 1994; 32: 1325–41.

Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, et al. Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. Neuroimage 2000; 11: 589–600.

Devlin J, Russell R, Davis M, Price C, Moss H, Fadili J, et al. Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. Neuropsychologia. In press 2001.

Dorfman D, Alf E. Maximum likelihood estimation of parameters of signal-detection theory and determination of confidence intervals-rating method data. J Math Psychol 1969; 6: 487–96.

Duncan J, Owen A. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci 2000; 23: 475–82.

Durrant-Peatfield MR, Tyler LK, Moss HE, Levy JP. The distinctiveness of form and function in category structure: a connectionist model. In: Shafto MG, Langley P, editors. Proceedings of the 19th Annual Conference of the Cognitive Science Society. Mahwah (NJ): Lawrence Erlbaum; 1997. p. 193–8.

Everitt BS, Bullmore ET. Mixture model mapping of the brain activation in functional magnetic resonance images. Hum Brain Mapp 1999; 7: 1–14.

Fiez JA. Phonology, semantics, and the role of the left inferior prefrontal cortex [editorial]. [Review]. Hum Brain Mapp 1997; 5: 79–83.

Forster KI, Forster JC. The DMASTR display system for mental chronometry. Tucson (AZ): University of Arizona; 1990.

Franklin S. Dissociations in auditory word comprehension: evidence from 9 'fluent' aphasic patients. Aphasiology 1989; 3: 189–207.

Friston KJ, Worsley KJ, Frackowiak RS, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1994; 1: 214–20.

Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RS. Spatial registration and normalization of images. Hum Brain Mapp 1995a; 3: 165–89.

Friston KJ, Holmes AP, Worsley KJ, Poline J-B, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 1995b; 2: 189–210.

Gabrieli JD, Poldrack RA, Desmond JE. The role of left prefrontal cortex in language and memory. [Review]. Proc Natl Acad Sci USA 1998; 95: 906–13.

Genovese CR, Noll DC, Eddy WF. Estimating test–retest reliability in functional MR imaging. I. Statistical methodology. Magn Reson Med 1997; 38: 497–507.

Gentner D. Some interesting differences between nouns and verbs. Cognit Brain Theory 1981; 4: 161–78.

Gernsbacher MA. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. J Exp Psychol Gen 1984; 113: 256–81.

Gleitman LR. Words, words, words . . . [Review]. Philos Trans R Soc Lond B Biol Sci 1994; 346: 71–7.

Goodglass H, Klein B, Carey P, Jones K. Specific semantic word categories in aphasia. Cortex 1966; 2: 74–89.

Gonnerman LM, Andersen ES, Devlin JT, Kempler D, Seidenberg MS. Double dissociation of semantic categories in Alzheimer's disease. Brain Lang 1997; 57: 254–79.

Graesser AC, Hopkinson PL, Schmid C. Differences in interconcept organization between nouns and verbs. J Mem Lang 1987; 26: 242–53..

Gustard S, Fadili MJ, Williams E, Hall LD, Carpenter AT, Bretty M, et al. Effect of slice orientation on reproducibility of fMRI motor activation at 3T. Magn Reson Imag. In press 2001.

Huttenlocher J, Lui F. The semantic organisation of some simple nouns and verbs. J Verb Learn Verb Behav 1979; 18: 141–62.

James CT. The role of semantic information in lexical decisions. J Exp Psychol Hum Percept Perform 1975; 104: 130–6.

Kiehl KA, Liddle PF, Smith AM, Mendrek A, Forster BB, Hare RD. Neural pathways involved in the processing of concrete and abstract words. Hum Brain Mapp 1999; 7: 225–33.

Kinahan PE, Rogers JG. Analytic 3D image reconstruction using all detected events. IEEE Trans Nucl Sci 1989; 35: 680–4.

Kroll JF, Merves JS. Lexical access for concrete and abstract words. J Exp Psychol Learn Mem Cognit 1986; 12: 92–107.

Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG. Discrete cortical regions associated with knowledge of colour and knowledge of action. Science 1995; 270: 102–5.

McCarthy R, Warrington EK. Category specificity in an agrammatic patient: the relative impairment of verb retrieval and comprehension. Neuropsychologia 1985; 23: 709–27.

Metz CE. ROC methodology in radiological imaging. Invest Radiol 1986; 21: 720–33.

Miceli G, Silveri MC, Villa G, Caramazza A. On the basis for the agrammatics' difficulty in producing main verbs. Cortex 1984; 20: 207–20.

Miller GA, Fellbaum C. Semantic networks of English. Cognition 1991; 41: 197–229.

Mummery CJ, Patterson K, Hodges JR, Wise RJ. Generating 'tiger' as an animal name or a word beginning with T: differences in brain activation. Proc R Soc Lond B Biol Sci 1996; 263: 989–95

Mummery CJ, Patterson K, Hodges JR, Price CJ. Functional neuroanatomy of the semantic system: divisible by what? J Cogn Neurosci 1998; 10: 766–77.

Paivio A. Dual coding theory: retrospect and current status. Can J Psychol 1991; 45: 255–87.

Perani D, Cappa SF, Schnur T, Tettamanti M, Collinia S, Rosa M, et al. The neural correlates of verb and noun processing: a PET study. Brain 1999; 122: 2337–44.

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.

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME. Positron emission tomographic studies of the processing of single words. J Cogn Neurosci 1989; 1: 153–70.

Price CJ, Wise RJ, Watson JD, Patterson K, Howard D, Frackowiak RS. Brain activity during reading: the effects of exposure duration and task. Brain 1994; 117: 1255–69.

Price CJ, Wise RJ, Frackowiak RS. Demonstrating the implicit processing of visually presented words and pseudowords. Cereb Cortex 1996; 6: 62–70.

Rubenstein H, Garfield L, Millikan JA. Homographic entries in the internal lexicon. J Verb Learn Verb Behav1970; 9: 487–94.

Rubin DC. 51 properties of 125 words: a unit analysis of verbal behavior. J Verb Learn Verb Behav 1980; 19: 736–55.

Rumsey JM, Horwitz B, Donohue BC, Nace K, Maisog JM, Andreason P. Phonological and orthographic components of word recognition: a PET–rCBF study. Brain 1997; 120: 739–59.

Silbersweig DA, Stern E, Frith CD, Cahill C, Schnorr L, Grootoonk S, et al. Detection of thirty-second cognitive activations in single subjects with positron emission tomography: a new low-dose H2(15)O regional cerebral blood flow three-dimensional imaging technique. J Cereb Blood Flow Metab 1993; 13: 617–29.

Skudlarski P, Constable RT, Gore JC. ROC analysis of statistical methods used in functional MRI: individual subjects. Neuroimage, 1999; 9: 311–29.

Sorensen JA, Wang X. ROC methods for evaluation of fMRI techniques. Magn Reson Med 1996; 36: 737–44.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci USA 1997; 94:14792–7.

Tyler LK, Moss HE, Jennings F. Abstract word deficits in aphasia: evidence from semantic priming. Neuropsychology 1995; 9: 354–63.

Tyler LK, Moss HE, Durrant-Peatfield M, Levy J. Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. Brain Lang 2000; 75: 195–231

Ulrich R, Miller J. Effects of truncation on reaction time analysis. J Exp Psychol Gen 1994; 123: 34–80.

Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS. Functional anatomy of a common semantic system for words and pictures. Nature 1996; 383: 254–6.

Van Horn JD, Ellmore TM, Esposito G, Berman KF. Mapping voxel-based statistical power on parametric images. Neuroimage 1998; 7: 97–107.

Wagner AD, Desmond JE, Demb JB, Glover GH, Gabrieli JDE. Semantic repetition priming for verbal and pictorial knowledge: a functional MRI study of left inferior prefrontal cortex. J Cogn Neurosci 1997; 9: 714–26.

Warburton E, Wise RJ, Price CJ, Weiller C, Hadar U, Ramsay S. Noun and verb retrieval by normal subjects. Studies with PET. [Review]. Brain 1996; 119: 159–79.

Wise R, Chollet F, Hadar U, Friston K, Hoffner E, Frackowiak, R. Distribution of cortical neural networks involved in word comprehension and word retrieval. Brain 1991; 114: 1803–17.

Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. J Cereb Blood Flow Metab 1992; 12: 900–18.

Received July 24, 2000. Revised February 2, 2001.
Accepted April 5, 2001

Appendix I

Overview of ROC methods

Estimation of the receiver operating characteristic (ROC) curve is a well-known methodology for quantifying the detection accuracy of a given statistical test. A ROC curve is a plot of $1 - \beta$ versus α, where α and β are, respectively, the type I (false-positive) and type II (false-negative) error probabilities. The area under the ROC curve is commonly regarded as a good single criterion for characterizing detection accuracy (Sorenson and Wang, 1996; Skudlarski *et al.*, 1999); the larger the area under the ROC curve, the better the detection accuracy of the test. ROC curves can be fitted to an observed data set by assuming certain forms for the probability distributions of the test statistic under the null and alternative hypotheses. The parameters of the distributions in such a finite mixture model can then be estimated by maximum likelihood (Dorfman and Alf, 1969; Metz, 1986, Everitt and Bullmore, 1999).

More formally, the mixture model for the observed test statistics can be written

$$f(x) = \lambda f_{H1} = (1 - \lambda) f_{H0}$$

where λ is the mixing proportion (the proportion of truly activated voxels); $f(x)$ is the probability of observing a voxel as being activated; and f_{H0} and f_{H1} are the probability density functions under the null and the alternative hypotheses, respectively.

If the test statistic is a t statistic, as it is in this study, these probability density functions are assumed to be respectively the central and non-central t distributions with the same number of degrees of freedom; the non-centrality parameter for the probability density function under the alternative hypothesis is denoted δ. Assuming the voxels are independent, the model parameters λ and δ can be estimated conveniently by maximizing the log likelihood function of the mixture model, for example using a modified Newton method for iterative optimization. For each value of α, we can then calculate the corresponding power $(1 - \beta)$ using the alternative probability density function parameterized by δ. The ROC curve is then plotted and the area under the curve is calculated.

Additionally, the ROC curve can be used to identify rational probability thresholds for activation mapping. For example, it can be shown that the threshold which maximizes the fraction of truly activated voxels identified as activated (true positive fraction) and the fraction of truly non-activated voxels identified as non-activated (true negative fraction) is given by the point on the ROC curve with gradient $= (1 - \lambda)/\lambda$ (Gustard *et al.*, 2001). This approach, whereby the key decision concerning which probability threshold to use for activation mapping is informed by consideration of false-negative as well as false-positive error rates, provides an alternative to the traditional emphasis in brain mapping on false-positive error rate as the sole criterion for threshold selection.