NeuroImage **11,** 589–600 (2000)
doi:10.1006/nimg.2000.0595, available online at http://www.idealibrary.com on IDEAL®

Susceptibility-Induced Loss of Signal: Comparing PET and fMRI on a Semantic Task

Joseph T. Devlin,* Richard P. Russell,* Matt H. Davis,* Cathy J. Price,† James Wilson,‡ Helen E. Moss,* Paul M. Matthews,‡ and Lorraine K. Tyler*

**Centre for Speech and Language and Wolfson Brain Imaging Centre, Cambridge University, Cambridge, England; †Wellcome Department of Cognitive Neurology, University College London, London, England; and ‡Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, Oxford, England*

Received October 11, 1999

Functional magnetic resonance imaging (fMRI) has become a popular tool for investigations into the neural correlates of cognitive activity. One limitation of fMRI, however, is that it has difficulty imaging regions near tissue interfaces due to distortions from macroscopic susceptibility effects which become more severe at higher magnetic field strengths. This difficulty can be particularly problematic for language tasks that engage regions of the temporal lobes near the air-filled sinuses. This paper investigates susceptibility-induced signal loss in the temporal lobes and proposes that by defining *a priori* regions of interest and using the small-volume statistical correction of K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans (1996, *Hum. Brain Mapp.* 4: 58–83), activations in these areas can sometimes be detected by increasing the statistical power of the analysis. We conducted two experiments, one with PET and the other with fMRI, using almost identical semantic categorization paradigms and comparable methods of analysis. There were areas of overlap as well as differences between the PET and fMRI results. One anticipated difference was a lack of activation in two regions in the temporal lobe on initial analyses in the fMRI data set. With a specific region of interest, however, activation in one of the regions was detected. These experiments demonstrate three points: first, even for almost identical cognitive tasks such as those in this study, PET and fMRI may not produce identical results; second, differences between the two methods due to macroscopic susceptibility artifacts in fMRI can be overcome with appropriate statistical corrections, but only partially; and third, new data acquisition paradigms are necessary to fully deal with susceptibility-induced signal loss if the sensitivity of the fMRI experiment to temporal lobe activations is to be enhanced. © 2000 Academic Press

INTRODUCTION

Both positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are noninvasive techniques for investigating the neural correlates of cognitive processes. However, fMRI offers a number of important advantages over PET that make it the modality of choice for many applications. One significant advantage is that fMRI provides for greater spatial and temporal resolution. This has the potential to reduce or eliminate many of the problems of stimulus blocking faced by PET, allowing the use of event-based and trial-based experimental designs (Josephs and Henson, 1999; Zarahn *et al.*, 1997). Furthermore since fMRI does not involve radiation exposure it is well suited to longitudinal studies of progressive neurological patients where repeated scanning sessions will be valuable. Before fMRI can replace PET, however, it is important to confirm that these two different modalities provide a consistent picture of the neural activations involved in a range of cognitive and behavioral tasks.

Although many studies have demonstrated that PET and fMRI produce similar results in motor (Joliot *et al.*, 1999; Kinahan and Noll, 1999), perceptual (Clark *et al.*, 1996), and higher-level cognitive activity such as attention (Coull and Nobre, 1998), other studies have been unable to fully replicate PET results in fMRI (e.g., Ojemann *et al.*, 1998; Price *et al.*, 1999b). One factor that is likely to contribute to these findings is the well-known difficulty of using fMRI to image regions

1053-8119/00 $35.00

near tissue interfaces such as the air-filled sinus/brain interfaces, particularly at higher magnetic field strengths. In these areas, rapid changes in magnetic susceptibility induce large internal static local field gradients. These macroscopic field gradients result in both geometric distortion of the image and, at higher local field gradients, loss of BOLD signal (Jezzard and Clare, 1999). Both effects are referred to as a susceptibility artifact (Ojemann *et al.*, 1997). The result of susceptibility artifacts is a reduced signal-to-noise ratio (SNR) in the affected regions. Any such reduction in activation contrast-to-noise reduces sensitivity to resolve the signal changes due to cognitive activity, which are typically slight, even at higher magnetic fields (Menon and Kim, 1999).

This poses a particular problem for language studies as two of the left-hemisphere areas commonly implicated in language processing, the temporal pole and the posterior-lateral surface of the middle and inferior temporal gyri, are regions prone to large susceptibility artifacts due to their proximity to air-filled sinuses. For instance, two very similar studies of lexical-semantic processing using PET (e.g., Perani *et al.*, 1999) and fMRI (e.g., Kiehl *et al.*, 1999) demonstrated very comparable patterns of activation overall except for a lack of temporal lobe activation in the fMRI study. It is important, however, to note that not all temporal regions are equally vulnerable to susceptibility-induced reduction in signal. Indeed, many studies using fMRI have identified temporal lobe activations in linguistic tasks in regions outside of the expected susceptibility artifacts (e.g., Binder *et al.*, 1997; Price *et al.*, 1999a; Pugh *et al.*, 1996). Nonetheless, areas of the temporal lobes proximal to the sinuses are likely to be difficult to image using fMRI.

Given the decreased SNR in certain temporal lobe regions in fMRI, statistical comparisons in these areas are likely to be significantly less powerful than in other regions of the brain. One technique for increasing the statistical power of an analysis is to reduce the number of comparisons by defining an *a priori* region-of-interest (ROI). Recently, Worsley and colleagues (1996) have proposed an algorithm for generating small-volume corrections based on the smoothness of the random field, the three-dimensional shape of the volume, and the size of the region. If this correction is applied to an analysis of echo-planar imaging (EPI) data, it may indeed be possible to increase the sensitivity of the data analysis in regions likely to be affected by a susceptibility-induced reduction of signal. This is particularly relevant to semantic tasks which regularly activate anterior and posterior-lateral temporal areas in the left hemisphere. In cases such as this where temporal lobe involvement is well established previous imaging studies can be used to generate fairly precise anatomical hypotheses and therefore to define appropriate ROIs.

In this study we investigated this possibility directly by developing a semantic categorization task designed to yield reliable temporal lobe activations, which we then performed using both PET and fMRI to assess brain activations. In the first experiment, a group of subjects performed the task in PET to determine areas of activation using methods that permit detection of blood flow changes in all brain regions including the anterior and posterior-lateral portions of the temporal lobes. Results from PET allowed us to define a precise ROI for use in the fMRI analysis. The second experiment was a replication of the first using fMRI. In an initial analysis we compared the areas of activation between the two experiments with a neuroanatomically unconstrained analysis of the fMRI data. Subsequently, we applied a statistical correction from our PET-based ROI to determine whether the susceptibility problem in the temporal lobes in fMRI could be reduced using an *a priori* anatomical hypothesis to constrain the analysis. The results highlighted three issues. Although the PET and fMRI experiments identified some common regions of activation, there were also differences between the two sets of results. Within regions of susceptibility artifacts activation could be identified with an appropriate statistical correction only when the signal loss in the area was fairly small. Where large local field gradients existed, however, the loss of signal was too severe to be recovered by post-processing techniques.

MATERIALS AND METHODS

Two categorization tasks were used in this study. The first was a semantic task designed to elicit robust temporal lobe activations and the second was a letter categorization task which was used as the baseline. The tasks were designed to be used with both PET and fMRI with only minimal differences due to the blocking requirements of the modality. Otherwise, the experimental manipulations were identical.

In the semantic categorization task, subjects read three cue words presented one after another on a computer screen and then made a speeded decision about whether a fourth (target) word belonged to the same category as the cue words. Cue words were in lowercase and the target was in uppercase to signal participants when to make a response. For instance, subjects made a “same” response to “dolphin, seal, walrus, OTTER” by pressing the left mouse button and a “different” response to “moccasin, sandals, boot, CUP” by pressing the right mouse button (see Table 1). The mouse was always held in the subject’s right (domi-

TABLE 1

Sample Stimuli Used in the Two Categorization Tasks

Cue 1	Cue 2	Cue 3	Target	Response
Semantic categorization				
bookcase	cabinet	bench	COUCH	"same"
lobster	mussel	shrimp	CLAM	"same"
squirrel	wolf	fox	LIME	"different"
knife	spoon	fork	MARBLES	"different"
Letter categorization				
aaaaaa	aaaaa	aaaa	AAAAAA	"same"
sssss	ssss	sssssss	SSS	"same"
llllll	llllll	lll	YYYYY	"different"
ddd	ddddddd	dddddd	RRRR	"different"

nant) hand and both reaction times and accuracy were recorded.[1]

The letter categorization task shared the same stimulus and response characteristics, 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 capital letters, contained the same letter. For example, "fffff, fff, fffffffff, FFFFFF" constituted a "same" trial and "ttttt, ttttttt, tttt, HHHH" was a "different" trial. Again, subjects signaled their response by pressing either the left or the right mouse button.

The stimuli in the semantic task were matched on relevant dimensions including word frequency, familiarity, and letter length using the Celex (Baayen and Pipenbrook, 1995) and MRC Psycholinguistic databases (Coltheart, 1981). Trials with "same" responses had a mean (±SD) familiarity of 497 (±42.1) and frequency of 13.6 (±11.6) while items with "different" responses had a mean (±SD) familiarity and frequency of 502 (±37.8) and 16.9 (±12.8), respectively. The items in the letter task were matched on letter length with those in the semantic task. In both conditions, stimuli ranged from 3 to 9 letters in length with a mean (±SD) of 5.4 (±1.0) letters. There were 192 semantic trials and 96 letter trials.

The experiment was first run in a pilot study outside the scanners to determine presentation rates and durations such that the task was challenging for participants but allowed fast and accurate responses. Each cue word (or letter string) was displayed for 200 ms with a 400-ms delay between them. The target word (or letter string) was also presented for 200 ms. 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. Subjects ($n = 12$) were not significantly faster to respond to "same" trials (mean reaction time (RT) = 712 ms) than to "different" trials (mean RT = 736 ms) and very few errors were made in either type of trial (9 and 8% errors, respectively).

As mentioned previously, the type of stimuli, presentation durations and rates, and most aspects of the image preprocessing were identical in the PET and fMRI experiments. Where there were differences between the two experiments, they were uniformly the result of optimizing the signal-to-noise ratio in each modality with the intention of making the results of the two experiments more, rather than less, comparable.

EXPERIMENT 1: CATEGORIZATION IN PET

The purpose of this experiment was to identify areas of reliable activation in the semantic task relative to the baseline letter task. These regions were then used as the standard for comparison with fMRI. In addition, any temporal lobe activations in this experiment will be used as anatomical hypotheses for temporal activation in the fMRI experiment, so it is important that some activations in this experiment occur in areas prone to signal loss in EPI.

Method

Eight right-handed, healthy male volunteers aged 21–47 (mean 28), all of whom spoke British English as their first language, participated in this experiment. Each gave informed consent after the experimental methodology was explained. Volunteers were medically screened for PET prior to entering the scanning room.

The subjects participated in twelve 90-s scans, eight of the semantic categorization condition and four of the letter categorization condition. 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. Because only 45 s of stimuli was presented, to coincide with the critical period of tracer uptake (Silbersweig *et al.,* 1993), only half (96) of the semantic trials from the initial behavioral test were included in this experiment. The semantic and letter conditions were presented systematically such that no subject saw the conditions in the same order.

Scans were performed at the Wolfson Brain Imaging Centre in Cambridge, England, on a GE Advance PET Scanner (General Electric Medical Systems, Milwaukee, WI). It comprises 18 rings of crystals, which results in 35 image planes, each 4.25 mm thick. The axial field-of-view is 15.3 cm, thus allowing for whole-brain

[1] In fact the semantic task was divided into two conditions, natural kinds and artifacts, and was also used to investigate questions of specialization within the semantic system (e.g., Damasio *et al.,* 1996; Martin *et al.,* 1996; Mummery *et al.,* 1996). This distinction, however, is irrelevant to this study and will not be discussed further. For full details see Devlin *et al.* (2000).

acquisition. Each subject received a bolus of 300 MBq before each scan for a total radiation exposure of 4.2 mSv. The emission data were acquired with the septa retracted (3D mode) and reconstructed using the PROMIS algorithm (Kinahan and Rogers, 1989) with an unapodised Colsher filter. Corrections were applied for randoms, scatter, attenuation, and dead time. The voxel sizes were 2.34 × 2.34 × 4.25 mm.

Functional images were realigned (Friston *et al.*, 1995a) 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 4°, respectively for any of the participants. The mean image created by the realignment procedure was used to determine the parameters for transforming the images onto the Montreal Neurological Institute (MNI) mean brain. The 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 at full width half-maximum (FWHM) Gaussian filter. The SPM software was used to compute a within-subjects analysis (i.e., a fixed-effects model) using the general linear model (GLM) (Friston *et al.*, 1995b) where the only effect of interest was the comparison of the semantic and letter categorization tasks. There were 12 scans for each of the eight subjects yielding 78 degrees of freedom, a point we return to later.

Results and Discussion

The subjects' mean (±SD) reaction time and error rate in the semantic categorization task were 774 (±101) ms and 8.6 (±6.8)%, respectively. In the letter categorization task the mean RT and error rate were 652 (±152) ms and 4.7 (±6.26)%, respectively. These data indicated that the subjects found the letter categorization task easier than the semantic categorization task as both their mean reaction times and their error rates were lower for the letter task. This suggests that the subjects were attending to the task and performing it adequately.

The semantic categorization task produced two areas of reliable activation relative to the baseline, an area located in the right cerebellum and a massive region of more than 5000 voxels in the left hemisphere extending from the middle temporal region, through the temporal pole, and into the inferior and middle frontal areas (see Fig. 1). To identify individual peaks within this large volume we increased the height threshold (the uncorrected voxel-level *P* value) from the default of 0.001 to 0.0001. This effectively broke the single large volume of activation into three separate regions (see Table 2). These included a large left frontal activation which included Broca's area (BA 44/45) and extended dorsally into the middle frontal gyrus (BA 9). In addition, there were two left temporal activations, one with a peak in the inferior temporal gyrus (BA 20) and another on the anteromedial temporal pole (BA 38). All activations were significant at both the cluster and voxel levels after corrections for multiple comparisons (Friston *et al.*, 1994; Worsley *et al.*, 1992), except the temporal pole activation which was reliable at the cluster level ($P < 0.05$) but only a trend at the voxel level ($P < 0.1$).

There are two important points to note in these data. First, the volumes activated by the semantic task are a subset of those reported in previous studies of lexico-semantic localization in both the neuropsychological (Damasio *et al.*, 1996; Gainotti *et al.*, 1995) and the neuroimaging (Demonet *et al.*, 1992; Price *et al.*, 1997; Vandenberghe *et al.*, 1996) literature. Second, and perhaps more importantly for our purposes, both regions of temporal lobe activation were in areas likely to be affected by EPI susceptibility artifacts. The anteriomedial activation was adjacent to the sphenoid sinus while the more posterior activation was on the lateral and ventral surfaces of the middle and inferior temporal gyri just anterior to the petrous bone and the mastoid air cells. Thus these two regions provide an appropriate test case for comparisons with fMRI.

EXPERIMENT 2: CATEGORIZATION IN fMRI

The purpose of the second experiment was to determine whether the PET results could be replicated using fMRI. Many studies have demonstrated that for nonlinguistic tasks, fMRI is capable of reproducing patterns of results similar to those seen in PET studies (Clark *et al.*, 1996; Coull and Nobre, 1998; Kinahan and Noll, 1999). However, these studies chose tasks that did not produce activations in areas of the brain where susceptibility artifacts are expected from fMRI. The present experiment was specifically designed to investigate whether fMRI could be used to replicate PET findings that included activation in regions prone to large static local field gradients.

Because we anticipated a loss of signal in the two temporal regions, we defined an *a priori* ROI based on the results of the first experiment. The ROI included only those temporal lobe voxels in extents that reached corrected significance in the PET data. In other words, our anatomical hypothesis was that we should see activation in the fMRI experiment in the same two areas, the posterior inferior temporal gyrus and the anteromedial temporal pole, which were present in the PET experiment. Thus we explored whether the small-volume statistical correction of Worsley *et al.* (1996) could overcome the reduction in signal in susceptible regions and replicate the findings from the PET study. Meth-

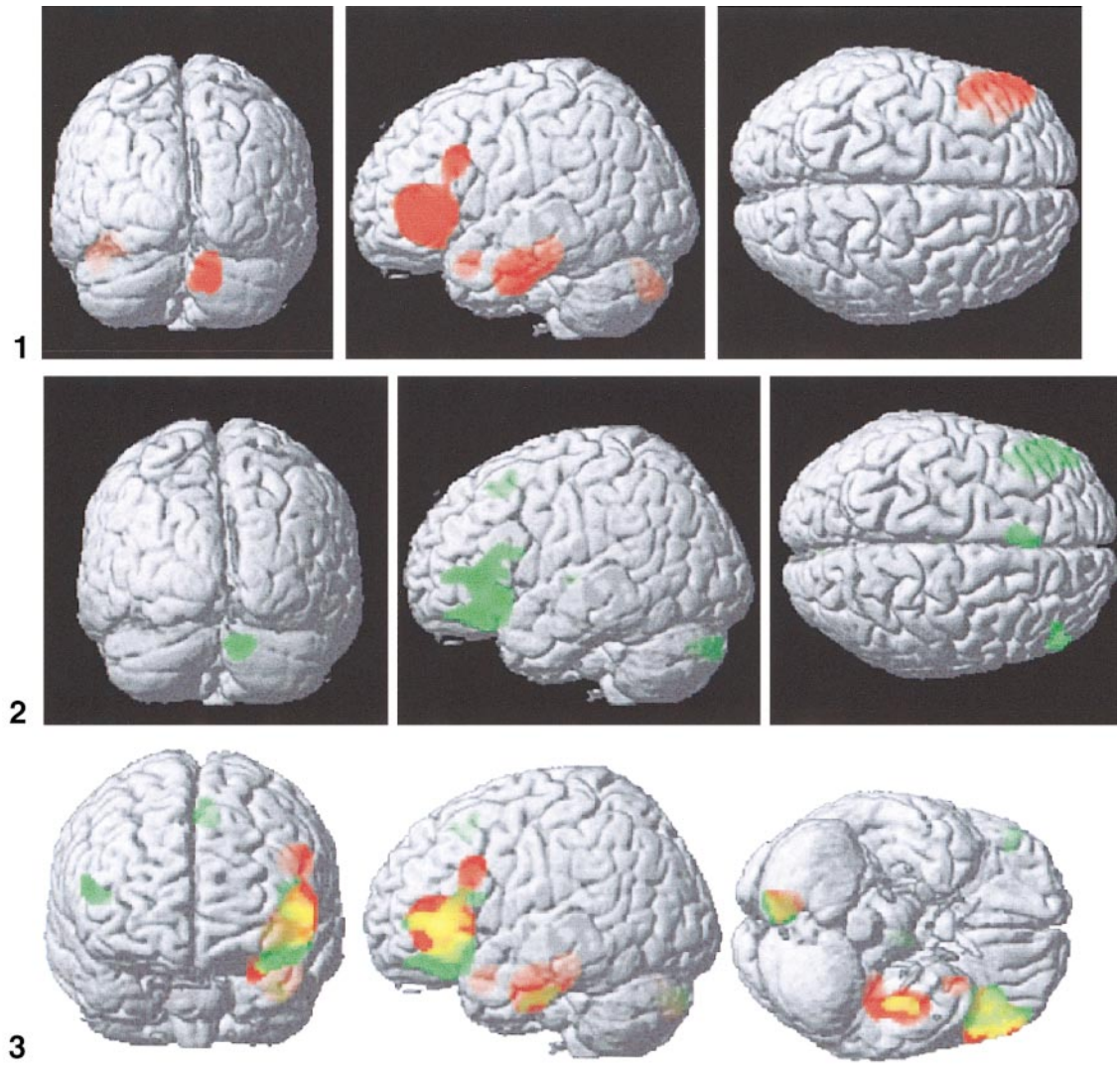

FIG. 1. Areas of activation in the Semantic–Letter categorization comparison from the PET experiment. All activations were reliable at a corrected $P < 0.05$, with extents shown at a height threshold of $p_{\text{uncorrected}} < 0.0001$. The brightness of active areas is proportional to their distance from the brain surface.

FIG. 2. Areas of activation in the Semantic–Letter categorization comparison from the fMRI experiment as in Fig. 1.

FIG. 3. Comparison between the semantic activation in the PET and fMRI experiments. PET activations are shown in red and fMRI activations in green. Where the two overlap is displayed in yellow. Note that the area in the temporal lobes that was reliably present in the ROI analysis is included in the reliable fMRI activations.

odologically this experiment was identical to the previous experiment except for differences in stimulus blocking between PET and fMRI and minor differences in image preprocessing.

Method

Eight right-handed, healthy volunteers aged 22–64 (mean 34), all of whom spoke British English as their first language, participated in this experiment. There were three women and five men. Each gave informed consent after the experimental methodology was explained. Volunteers were screened for magnetic resonance compatibility prior to entering the scanning room.

The subjects participated in two 9-min sessions where stimuli were presented in 30-s blocks. As in the previous experiment, there were three conditions: two semantic conditions (natural kinds and artifacts) and one baseline condition (letter categorization). This permitted 8 trials (at 3750 ms each) per block rather than 12, as in the PET experiment. Consequently, each subject saw a total of 192 semantic trials (8 trials/block with 12 semantic blocks/session and two sessions), twice as many trials as occurred in the PET experi-

TABLE 2

Comparison of the Semantic and Letter Categorization Tasks in the Group PET Study[a]

Region	Coordinates (*x*, *y*, *z*)	Voxel level		Cluster level	
		$p_{corrected}$	*T*	$p_{corrected}$	Extent
Frontal					
Left Broca's area (BA 45)	(−52 34 2)	0.000	6.92	0.000	1787
Left inferior and medial gyri (BA 44/9)	(−48 18 30)	0.026	5.02		
Temporal					
Left inferior temporal gyrus (BA 20)	(−42 −14 −28)	0.000	7.12	0.000	999
	(−36 −32 −16)	0.083	4.65		
Left anteromedial temporal lobe (BA 38)	(−28 10 −24)	0.067	4.72	0.030	122
Cerebellum					
Right posterior cerebellum	(14 −78 −32)	0.001	5.96	0.001	409

[a] The results are presented for both the voxel and cluster levels of significance based on a height threshold of 0.0001. The coordinates are in MNI space and the corrected *P* values, the SPM{*t*}, and the extent of activation are presented. Multiple peaks within an extent are shown on subsequent lines.

ment. Thus, all of the trials that were used in the behavioral pretest were present in this experiment.

During each 9-min imaging session 180 images were collected. An additional 4 dummy volumes were collected at the start of each session to allow for T1 equilibrium before the test trials started. It is worth noting that the trial duration (3.75 s) was not an integer multiple of the TR (3 s) and consequently the data were acquired at four different points within the peristimulus time, decreasing the potential contribution of artifactual increases or decreases in the signal intensity (cf. Price *et al.*, 1999a).

All scans were carried out using the Varian-Siemens 3-T MRI scanner at the Functional Magnetic Resonance Imaging of the Brain (FMRIB) Centre in Oxford. A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage head radiofrequency coil tuned to 127.4 MHz. A gradient-echo EPI sequence was used for image collection (TR 3 s, TE 30 ms, 64 × 64 resolution, 256 × 256-mm FOV). Twenty-one slices were employed to cover the brain with 6-mm slice thickness and in-plane resolution of 4 mm. Because of the high field strength of the magnet (3 T), a manual shim was set up for each subject using eight terms (three linear and five quadratic) to reduce magnetic field inhomogeneties and a TE of 30 ms was used to jointly optimize BOLD contrast-to-noise and image signal-to-noise while minimizing intravoxel dephasing.

Functional images were processed as for PET. Images were realigned with translation and rotation corrections less than 3 mm and 2°, respectively. The images were then normalized to the EPI template transforming them onto the MNI mean brain. Finally, each image was smoothed with an 8-mm FWHM Gaussian filter.

The fMRI data were analyzed using a within-subject (i.e., fixed effects) analysis as was carried out with the PET data. To make this as comparable as possible to the PET analysis, it was necessary to reduce the degrees of freedom and to limit the number of statistical comparisons in the fMRI analysis. Rather than use all 2520 scans in the analysis (which leads to very large number of degrees of freedom), we reduced the number of scans to 42 (7 subjects × 2 sessions × 3 conditions) by creating mean images for each condition per session per subject. This produced an analysis with 26 degrees of freedom which was more comparable to the PET analysis.

Note that despite an effort to keep the PET and fMRI experiments identical, there were many differences between the two. These included the total number of semantic trials (96 vs 192), the length of stimulus blocks (45 s vs 30 s), the normalization of the images (PET template vs EPI template), the smoothing of the functional images (16 mm vs 8 mm), and the specification of the GLM (including temporal autocorrelations in the fMRI analysis and 78 *df* vs 26 *df*). Each of these was necessary due to important differences between the properties of the two imaging modalities. By doing appropriate analyses for each experiment we attempted to make the two analyses more, rather than less, comparable. We return to this point in the General Discussion.

Results and Discussion

Technical difficulties resulted in the loss of the behavioral data from one subject and consequently the functional images for that subject were not included in any of the analyses because we could not verify that he was adequately performing the task. The other seven subjects, however, performed the task comparably to the PET group. Specifically, as in the PET experiment they were slower in the semantic task than the letter

TABLE 3

Comparison of the Semantic and Letter Categorization Tasks in the Group fMRI Study[a]

Region	Coordinates (x, y, z)	Voxel level		Cluster level	
		$p_{corrected}$	T	$p_{corrected}$	Extent
Frontal					
Left BA 47 through BA 44/45	(−36 30 −20)	0.001	8.05	0.000	1223
	(−52 36 4)	0.001	7.91		
	(−46 24 −8)	0.002	7.50		
Right BA 45/46	(52 42 16)	0.042	6.10	0.017	85
	(46 40 10)	0.310	5.07		
Left BA 8	(−6 20 50)	0.103	5.66	0.006	118
Cerebellum					
Right posterior cerebellum	(10 −84 −32)	0.018	6.50	0.000	176

[a] The results are presented for both the voxel and cluster levels of significance based on a height threshold of 0.0001. The coordinates are in MNI space and the corrected P values, the SPM{t}, and the extent of activation are presented. Multiple peaks within an extent are shown on subsequent lines.

task (818 ms vs 678 ms) and also less accurate (9% vs 2% errors). These results indicate that the subjects were adequately performing the task.

The results of the image analysis are displayed in Fig. 2. They revealed four reliable areas of activation for the Semantic–Baseline comparison. There was a large active extent in the left frontal lobe from the inferior frontal gyrus (BA 47) into Broca's area (BA 44/45), a similar activation in the right hemisphere (BA 44/45), and the medial surface of the superior frontal cortex (BA 8) was active on the left (see Table 3). The fourth area was in the right cerebellum. There was also activation in the left thalamus but this did not reach corrected levels of significance at either the voxel or cluster level ($0.05 < P < 0.1$). There were no reliable activations in the temporal lobes.

These findings demonstrate areas in common with the PET results as well as differences between the two experiments (see Fig. 3). In Fig. 3, the regions of reliable semantic versus letter categorization activation in the PET experiment are displayed in red while those from the fMRI experiment are shown in green. The overlap between the two is displayed in yellow. Both experiments produced frontal activation in the left inferior and middle frontal gyri as well as in the right cerebellum. In the PET experiment, however, there were two areas of reliable temporal lobe activity, the inferior temporal gyrus and the anteromedial temporal pole, that were not present in the fMRI images, presumably due to air–tissue interfaces near these regions. In addition, the fMRI experiment revealed two areas of activation not present in the PET data, a left premotor region and the right-hemisphere homologue of Broca's area.

First consider the similar results from the two experiments. In both cases the semantic categorization task produced reliable activations of the left inferior frontal lobe as well as a posterior-medial activation in the right cerebellum. Both regions overlapped considerably between the two experiments. Nonetheless, there were minor differences between the frontal activations; namely, the fMRI activation was more inferior than the activation seen in the PET experiment and had less of a dorsal extent. The right cerebellar activations, on the other hand, were essentially identical.

One important difference between the two experiments was the presence of strong temporal lobe activations in the PET but not the fMRI data. Because both the posterior lateral temporal lobe (BA 20/21) and the anteromedial temporal pole (BA 38) activations were in regions of probable susceptibility artifacts it is not surprising that these activations were not found in the fMRI data. To determine whether these activations were actually missing or were simply more difficult to detect because of the reduced signal-to-noise ratio we increased the statistical power of our comparison by applying a ROI correction. Using the Worsley *et al.* (1996) small-volume correction calculation, we determined the corrected t threshold to be $t > 3.85$. We then masked the SPM{t} map from the fMRI data with the ROI and looked for voxels with SPM{t} values greater than 3.85. The result was a single region of reliable activation in the left inferior temporal lobe of 57 voxels with a peak at (−40 −24 −30) and an SPM{t} value (at the peak) of 4.87. There was no significant activation in the left anteromedial temporal cortex. Thus, using the small-volume correction for an *a priori* anatomical hypothesis was sufficient to overcome some, but not all, of the signal reduction in the temporal lobe.

Finally, there were two regions of significant activation present in the fMRI results that were not present in the PET data, an area in the superior frontal gyrus (BA 8) and a right inferior frontal region (BA 44/45). The former had a SPM{t} value of 5.7 at its peak voxel

(−6 20 50) which was not significant at the voxel level ($P > 0.1$) although the large extent (118 voxels) was reliable at the cluster level ($P < 0.01$, both P values corrected for multiple comparisons). In the PET data this area was activated by the Semantic–Baseline comparison with a peak activation of SPM{t} = 4.4 ($P > 0.1$) at (−8 32 44) and an extent of 74 voxels ($P < 0.1$). Thus the same activation was present in both studies although in neither case did individual voxels reach corrected levels of significance. Instead, the cluster as a whole was reliable in the fMRI data but was only a trend in the PET data. The area in the right inferior frontal lobe was more active in the fMRI experiment (SPM{t} = 6.1, $P < 0.05$ at the peak (52 42 16)) than in the PET experiment (SPM{t} = 3.4, n.s. at its peak (44 28 22)) although there was an increase in activity in this region in both experiments. It is not clear why it reached significance in the fMRI experiment but not in the PET experiment. One possible explanation for these differences suggested by Veltman *et al.* (2000) is that the differences in stimulus presentation (i.e., alternating blocks in fMRI versus independent blocks in PET) may lead to increased frontal activation in fMRI as the subjects constantly switch between tasks.

To review then, the PET and fMRI experiments produced similar (though not identical) results. Both studies activated bilateral inferior frontal regions, left superior frontal cortex, left inferior temporal areas, and the right cerebellum. In the fMRI experiment, the inferior/middle temporal gyrus activation became evident only after applying a small-volume correction to examine the specific regions of interest as defined by our PET experiment. No activity was observed in any analysis of the fMRI data within the left temporal pole. In the PET data, the activations in the left superior and the right inferior frontal gyri did not reach reliable levels of significance although both were present at lower thresholds. We drew two conclusions from these results. First, although the PET and fMRI experiments identified some common regions of activation, there were also important differences. Second, the differences within expected regions of rapid macroscopic susceptibility change were only partially reduced when an appropriate statistical correction based on an *a priori* anatomical constraint of the fMRI analysis was applied.

GENERAL DISCUSSION

In this paper we have addressed the question of the extent to which fMRI can be used for linguistic paradigms given that susceptibility artifacts reduce the signal in many (but not all) relevant regions. We presented semantic and letter categorization tasks designed to induce activity in the temporal lobes in two experiments, one using PET and another using fMRI. The PET data confirmed that the semantic task produced robust activations in left extra-Sylvian regions including the temporal lobe. The fMRI data partially replicated the PET findings although there were differences in both the frontal and temporal lobes. These differences were partially reduced when we used an *a priori* anatomical hypothesis to constrain the number of statistical comparisons. We demonstrated that although susceptibility artifacts may have interfered with our ability to detect signal changes in the temporal lobes in fMRI, the posterior lateral temporal lobe activation could be reliably established but the anteromedial temporal pole activation could not, at least with similar-sized data sets.

In fMRI, magnetic field gradients are used for the localization of signals. Air–tissue interfaces, however, introduce steep local macroscopic gradients that cause a number of effects. The first is spatial distortion of the image data, due to the fact that these local gradients are not modeled in the image reconstruction (Jezzard and Clare, 1999). The second is signal loss from the affected regions. A large local field gradient across a voxel dimension causes the phases along this dimension to disperse during the echo time, partially canceling each other out and thus resulting in a reduction in signal. Due to the long echo time used to obtain the echo planar images the second of these effects can become substantial in areas of the temporal lobe. These effects may help to explain why our initial statistical analysis failed to detect activations in the left anteromedial temporal lobe (BA 38, near the sphenoid sinus) and in the left posterior inferior temporal lobe (BA 20/21, near the petrous bone, auditory canal, and mastoid air cells).

We attempted to directly test possible alternative mechanisms for the reduced sensitivity of the fMRI experiment for temporal lobe activation. There are two possibilities. Either a regionally specific increase in noise or a reduction in signal size would reduce the size of the t statistic in the temporal regions. To determine which of these factors contributed to our reduced ability to detect activations in the temporal lobes, we generated masks corresponding to the four active regions identified by the PET experiment: left inferior frontal, left anteromedial temporal, and left posterior-lateral temporal cortices and right cerebellum.[2] We then applied each mask to the residual error maps (ResMS) and the contrast size maps (con_*) to calculate the mean (SD) noise and effect sizes in each area for the two experiments. The results are presented in Table 4.

The noise levels in both the PET and fMRI data were fairly consistent across areas (although very different

[2] The masks were created with MRIcro software (http://www.mrc-cbu.cam.ac.uk/personal/chris.rorden/mricro.htm).

TABLE 4

Comparison of the Signal (Effect Size) and Noise (Residual Errors) in the PET and fMRI Experiments[a]

	Left anterior temporal (−28 10 −24) N = 122	Left posterior temporal (−42 −14 −28) N = 999	Left inferior frontal (−52 34 2) N = 1787	Right cerebellum (14 −78 −32) N = 409
PET (Experiment 1)				
Effect size	4.21 (0.29)	4.12 (0.78)	4.32 (0.81)	4.02 (0.48)
Residual error	5.13 (0.47)	4.12 (1.12)	4.07 (0.79)	3.89 (0.33)
fMRI (Experiment 2)				
Effect size	−0.11 (−0.16)	0.02 (0.41)	0.79 (0.32)	0.81 (0.25)
Residual error	0.19 (0.05)	0.12 (0.10)	0.12 (0.07)	0.16 (0.14)

[a] The means (SD) for each region are shown. Regions were defined as those areas of reliable activation in the PET experiment and the coordinate of the peak activation is listed for each. N specifies the number of 2-mm^3 voxels in the volume.

across modalities). What varied from region to region was the size of the effect. In the PET data, the effect size was roughly equivalent in all four areas. In the fMRI data, on the other hand, the effect size was clearly reduced in the two temporal lobe regions relative to the both the frontal lobe and the cerebellum. The result was a loss of statistical power in the medial temporal pole area and the posterior lateral temporal cortex.

To determine whether this reduction in signal strength was due to macroscopic susceptibility effects from the nearby air–tissue interfaces, we modeled the signal loss that would occur due to the induced susceptibility artifacts. To begin we transformed each subject's structural (T1) scan into the isotropic MNI standard brain space. Next we generated a phase map for each subject and unwrapped it to give continuous phase variation across the image. Because phase is directly proportional to the strength of the magnetic field and the echo time, this provided three-dimensional information on the local spatial variation of the magnetic field within each individual brain. From these field maps, gradient maps in the x, y, and z directions were calculated. The gradient map in the x direction was found by shifting the field map by one pixel in the positive x direction and subtracting this from the original field map. This was repeated in the negative x direction and the average of the two was taken. Field gradient maps in the y direction and z direction were calculated in the same way. Finally, by transforming the gradient maps into EPI space and masking the gradient maps with the four volumes of interest defined previously, the signal loss experienced by the volume of interest was computed. The signal loss that a single voxel suffered due to the internal local field gradients across it was estimated assuming a linear phase variation across the voxel. The true phase variation across a voxel in the vicinity of an air–tissue interface is generally nonlinear and becomes more so the closer the voxel is to the interface (Chen and Wyrwicz, 1999). Computer simulations showed that the greater the nonlinearity of the phase variation the greater the signal loss. Thus by assuming a linear phase variation we have chosen a conservative estimate of the actual signal loss.

For each region, the number of voxels experiencing less than 10, 10–25, 26–50, and more than 50% signal loss were calculated. Because these values were calculated in EPI space, the voxels were 4 × 4 × 6 mm (i.e., 12 times larger than those reported in Table 4). The results are shown in Table 5. The most severe signal loss occurred in the anterior temporal area with 7 of 11 voxels (64%) losing at least half of their signal due to local static field gradients. The posterior temporal area

TABLE 5

Estimated Signal Loss Based on Individual Subjects' Field Maps in the fMRI Experiment[a]

Percentage of region with an estimated loss of signal	Left anterior temporal area (−28 10 −24) N = 11	Left posterior temporal area (−42 −14 −28) N = 84	Left inferior frontal area (−52 34 2) N = 150	Right cerebellum (14 −78 −32) N = 34
<10%	9%	29%	71%	79%
10–25%	9%	18%	24%	21%
26–50%	18%	14%	3%	0%
>50%	64%	39%	2%	0%

[a] The regions are identical to those in the previous table. N specifies the number of 4 × 4 × 6-mm voxels in the volume in EPI space.

also experienced significant signal loss, although it was less severe than the more anterior region. Neither the frontal nor the cerebellar areas were seriously affected by susceptibility-induced signal loss, although some of the most inferior frontal voxels near the sinuses did have reduced signal intensities.

These results therefore could explain why the small-volume statistical correction was successful for identifying activation in the posterior but not the anterior temporal area. Although the left inferior temporal lobe activation was significantly degraded due to susceptibility effects from the nearby petrous bone, auditory canal, and mastoid air cells, there was still signal from a sufficiently large volume to obtain a significant activation. However, almost all of the left anteromedial temporal activation experienced greater than 25% signal loss, resulting in no significant activation in this area. With such large levels of signal loss, it is unlikely that any postprocessing technique would suffice to identify activation. Instead, new imaging methods such as tailored RF pulses (Chen and Wyrwicz, 1999) or *z* shimming (Yang *et al.*, 1997, 1998) likely will be necessary to obtain reliable signal in these areas. It also has been suggested that use of spiral scanning could reduce susceptibility artifacts in the temporal lobe (Crelier *et al.*, 1999).

Although we have demonstrated that macroscopic susceptibility effects are one important source of differences between our PET and fMRI data, it is worth noting that many other factors could contribute as well. These include underlying physiological differences between changes in rCBF and BOLD contrast, differences between temporally integrated whole-brain versus serial slice data acquisition, and the remaining differences between PET and fMRI experimental conditions.

To directly compare the results of the PET and fMRI experiments, we sought to make the experimental conditions and the analysis methods as similar as possible. It should, however, be noted that although as many factors as possible were held constant, the effects of changes in individual factors can vary across the two modalities (Rees *et al.*, 1997). The minor differences between the PET and fMRI paradigms in the total number of stimuli (96 vs 192), the blocking of stimuli (45 s vs 30 s), the normalization of the images (PET template vs EPI template), the smoothing of the functional images (16 mm vs 8 mm), and the data analysis (particularly the degrees of freedom) were made with the intention of optimizing the data and its comparability. For example, we used the most appropriate template when normalizing each data set. Similarly, the different smoothing kernels produced images with similar numbers of independent comparisons.

It may, however, be the case that by making these factors as comparable as possible we have artificially reduced the power of our fMRI data. For instance, by analyzing mean condition images rather than the complete time series we greatly reduced our degree of freedom (from 1229 to 26) and consequently may have decreased our statistical sensitivity. Interestingly, when we performed a full fixed-effect analysis on the time series data, it did not identify any reliable signal in the temporal lobes. By applying the small-volume correction, however, we again found activation in the posterior lateral temporal lobe. One of the main differences between the mean condition image and results of the fixed-effect analysis was in the spatial smoothness of the data. The mean images were smoother than the time series data and consequently resel sizes were larger in the mean image analysis (315 vs 287 voxels per resel). This meant that the SPM{*t*} map from the time series analysis had more independent statistical comparisons and thus a larger *t* threshold ($t > 3.90$ vs $t > 4.48$) for identification of reliable activations at a corrected voxel level significance within the ROI. As a result, the analysis was no more sensitive to temporal lobe activations than the mean image analysis.

In our study, the regions of interest for the fMRI analysis were based on areas of activation for the same task as observed in the preliminary PET experiment. In general, of course, it can be more valuable to define a region of interest based on previous neuroimaging studies. It also is possible to use the neuroimaging literature to generate precise anatomical hypotheses and then calculate the correct statistical threshold for the hypothesis rather than simply using an arbitrary default value. This is best illustrated with an example. Three studies of category specificity have identified a lateral posterior temporal area as being more active for artifacts than natural kinds (Damasio *et al.*, 1996; Moore and Price, 1999; Mummery *et al.*, 1998).[3] The reported peaks in the three studies were at (−57 −54 0), (−58 −60 −6), and (−54 −54 0), respectively, all of which were within a typical FWHM (12–16 mm) of a PET smoothing kernel and thus presumably represent overlapping extents. By using the mean of the coordinates as the center of a sphere and a FWHM smoothing value as its radius, one can easily identify a precise ROI for testing further hypotheses. The shape and volume of the ROI can be combined with the smoothness of the SPM{*t*} map under consideration to calculate an exact *t* threshold to use for significance testing within this region (Worsley *et al.*, 1996).

We have shown that although susceptibility artifacts reduce the BOLD contrast signal in the parts of the temporal lobes, in some cases activation can still be

[3] It should be noted that this same region was also preferentially active for nonobjects in the Moore and Price study and therefore can not truly be considered preferentially sensitive to artifacts. Nonetheless, it suffices for illustrative purposes.

reliably detected using a small volume correction if one has a reasonably specific anatomical hypothesis. However, signal from other areas will be so limited that no postprocessing technique will suffice to detect activations. To minimize the substantial effects of susceptibility artifacts, new imaging methods (or perhaps larger data sets) are required. For example, tailored RF pulses (Chen and Wyrwicz, 1999) can be used to refocus the spin phases, thereby recovering the signal loss due to internal field gradients.

The fact that susceptibility artifacts can be partially overcome is particularly important for language studies as the anterior and posterior lateral temporal lobes are consistently implicated in language processing (e.g., Damasio *et al.*, 1996; Demonet *et al.*, 1992; Mummery *et al.*, 1998; Warrington and Shallice, 1984). Thus with the appropriate regions-of-interest correction it may be possible to use event-related and trial-based fMRI designs in investigating the time course of language processing in the temporal lobe (cf. Menon and Kim, 1999). In addition, the use of fMRI opens a new avenue for studying the relation between neurological and behavioral deficits in patients with semantic deficits. Because no single task adequately measures the integrity of semantic processing, it is necessary to develop converging evidence from numerous experiments to assess a patient's residual semantic abilities (Shallice, 1988). One clear advantage of fMRI over PET is the potential to run multiple experiments with individuals showing linguistic deficits. However, as both our work and that of Veltman *et al.* (2000) hopefully illustrate, there are still significant differences between measures of neural activity obtained in PET and fMRI that need to be reconciled to confidently interpret the significance of patterns of fMRI activation in studies of language processing.

ACKNOWLEDGMENTS

We thank the members of the WBIC in Cambridge and the FMRIB Centre in Oxford for their invaluable assistance in collecting and analyzing these data; Steve Smith, Peter Jezzard, and Matthew Brett for many helpful discussions, both theoretical and technical; Chris Rorden for creating and supporting MRIcro; and David Plaut for suggesting the semantic categorization task. Finally, we thank our volunteers for their participation. This work was supported by an MRC program grant to L. K. Tyler and W. D. Marslen-Wilson and an MRC program grant in support of the Oxford FMRIB Centre, and by a McDonnell- Pew grant to L. K. Tyler.

REFERENCES

Ashburner, J., and Friston, K. 1997. Multimodal image coregistration and partitioning: A unified framework. *NeuroImage* **6:**209–217.

Ashburner, J., Neelin, P., Collins, D. L., Evans, A. C., and Friston, K. J. 1997. Incorporating prior knowledge into image registration. *NeuroImage* **6:**344–352.

Baayen, R. H., and Pipenbrook, R. 1995. *The Celex Lexical Database.* Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. 1997. Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* **17:**353–362.

Chen, N., and Wyrwicz, A. M. 1999. Removal of intravoxel dephasing in gradient-echo images using a field-map based RF refocusing technique. *Magn. Reson. Med.* **42:**807–812.

Clark, V. P., Keil, K., Maisog, J. M., Courtney, S., Ungerleider, L. G., and Haxby, J. V. 1996. Functional magnetic resonance imaging of human visual cortex during face matching: A comparison with positron emission tomography. *NeuroImage* **4:**1–15.

Coltheart, M. 1981. The MRC Psycholinguistics database. *Q. J. Exp. Psychol.* **33A:**497–505.

Coull, J. T., and Nobre, A. C. 1998. Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *J. Neurosci.* **18:**7426–7435.

Crelier, G. R., Hoge, R. D., Munger, P., and Pike, G. B. 1999. Perfusion-based functional magnetic resonance imaging with single-shot RARE and GRASE acquisitions. *Magn. Reson. Med.* **41:**132–136.

Damasio, H., Grabowski, T., Tranel, D., Hichwa, R., and Damasio, A. 1996. A neural basis for lexical retrieval. *Nature* **380:**499–505.

Demonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., and Frackowiak, R. 1992. The anatomy of phonological and semantic processing in normal subjects. *Brain* **115:**1753–1768.

Devlin, J. T., Russell, R. P., Price, C. J., Davis, M., Moss, H. E., and Tyler, L. K. 2000. Neuroimaging evidence for a unitary semantic system. *Cognitive Neuroscience Society Annual Meeting*, 2000.

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. 1994. Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* **1:**214–220.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., and Frackowiak, R. S. J. 1995a. Spatial registration and normalization of images. *Hum. Brain Mapp.* **2:**165–189.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., and Frackowiak, R. S. J. 1995b. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2:**189–210.

Gainotti, G., Silveri, M. S., Daniele, A., and Giustolisi, L. 1995. Neuroanatomic correlates of category specific semantic disorders: A critical survey. *Memory* **3:**247–264.

Jezzard, P., and Clare, S. 1999. Sources of distortion in functional MRI data. *Hum. Brain Mapp.* **8:**80–85.

Joliot, M., Papathanassiou, D., Mallet, E., Quinton, O., Mazoyer, N., Courtheoux, P., and Mazoyer, B. 1999. fMRI and PET of self-paced finger movement: Comparison of intersubject stereotaxic averaged data. *NeuroImage* **10:**430–447.

Josephs, O., and Henson, R. N. 1999. Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philos. Trans. R. Soc. London B* **354:**1215–1228.

Kiehl, K. A., Liddle, P. F., Smith, A. M., Mendrek, A., Forster, B. B., and Hare, R. D. 1999. Neural pathways involved in the processing of concrete and abstract words. *Hum. Brain Mapp.* **7:**225–233.

Kinahan, P. E., and Noll, D. C. 1999. A direct comparison between whole-brain PET and BOLD fMRI measurements of single-subject activation response. *NeuroImage* **9:**430–438.

Kinahan, P. E., and Rogers, J. G. 1989. Analytic 3D image reconstruction using all detected events. *IEEE Trans. Nucl. Sci.* **35:** 680–684.

Martin, A., Wiggs, C., Ungerleider, L., and Haxby, J. 1996. Neural correlates of category-specific knowledge. *Nature* **379:**649–652.

Menon, R. S., and Kim, S.-G. 1999. Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends Cogn. Sci.* **3:**207–216.

Moore, C. J., and Price, C. J. 1999. A functional neuroimaging study of the variables that generate category specific object processing differences. *Brain* **122:**943–962.

Mummery, C. J., Patterson, K., Hodges, J., and Wise, R. J. 1996. Generating 'tiger' as an animal name or a word beginning with T: Differences in brain activation. *Proc. R. Soc. London B* **263:**989–995.

Mummery, C. J., Patterson, K., Hodges, J. R., and Price, C. J. 1998. Functional neuroanatomy of the semantic system: Divisible by what? *J. Cogn. Neurosci.* **10:**766–777.

Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinstry, R. C., Raichle, M. E., and Conturo, T. E. 1997. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *NeuroImage* **6:**156–167.

Ojemann, J. G., Buckner, R. L., Akbudak, E., Snyder, A. Z., Ollinger, J. M., McKinstry, R. C., Rosen, B. R., Petersen, S. E., Raichle, M. E., and Conturo, T. E. 1998. Function MRI studies of word stem completion: Reliability across laboratories and comparison to blood flow imaging with PET. *Hum. Brain Mapp.* **6:**203–215.

Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., and Fazio, F. 1999. The neural correlates of verb and noun processing: A PET study. *Brain* **122:**2337–2344.

Price, C. J., Moore, C. J., Humphreys, G. W., and Wise, R. J. 1997. Segregating semantic from phonological processes during reading. *J. Cogn. Neurosci.* **9:**727–733.

Price, C. J., Veltman, D. J., Ashburner, J., Josephs, O., and Friston, K. J. 1999a. The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. *NeuroImage* **10:**36–44.

Price, C. J., Veltman, D. J., and Friston, K. J. 1999b. PET and fMRI studies of visual word processing. *NeuroImage* **9:**S1043.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Shankweiler, D. P., Katz, L., Fletcher, J. M., and Gore, J. C. 1996. Cerebral organization of component processes in reading. *Brain* **119**(Pt. 4):1221–1238.

Rees, G., Howseman, A., Josephs, O., Frith, C. D., Friston, K. J., Frackowiak, R. S. J., and Turner, R. 1997. Characterizing the relationship between BOLD contrast and regional cerebral blood flow measurements by varying the stimulus presentation rate. *NeuroImage* **6:**270–278.

Shallice, T. 1988. Specialization within the semantic system. *Cogn. Neuropsychol.* **5:**133–142.

Silbersweig, D. A., Stern, E., Frith, C. D., Cahill, C., Schnorr, L., Grootoonk, S., Spinks, T., Clark, J., Frackowiak, R., and Jones, T. 1993. Detection of thirty-second cognitive activations in single subjects with positron emission tomography: A new low-dose H2O15 region cerebral blood flow three-dimensional imaging technique. *J. Cereb. Blood Flow Metab.* **13:**617–629.

Vandenberghe, R., Price, C. J., Wise, R., Josephs, O., and Frackowiak, R. S. J. 1996. Functional anatomy of a common semantic system for words and pictures. *Nature* **383:**254–256.

Veltman, R., Friston, K. J., Sanders, G., and Price, C. J. 2000. Test–retest reliability of language paradigms in PET and fMRI. *NeuroImage* **11:** 575–588.

Warrington, E., and Shallice, T. 1984. Category specific semantic impairments. *Brain* **107:**829–853.

Worsley, K. J., Marrett, S., Neelin, P., and Evans, A. C. 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* **12:**900–918.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* **4:**58–83.

Yang, Q. X., Dardzinski, B. J., Li, S. Z., Eslinger, P. J., and Smith, M. B. 1997. Multi-gradient echo with susceptibility inhomogeneity compensation (MGESIC): Demonstration of fMRI in the olfactory cortex at 3.0 T. *Magn. Reson. Med.* **37:**331–335.

Yang, Q. X., Williams, G. D., Demeure, R. J., Mosher, T. J., and Smith, M. B. 1998. Removal of local field gradient artifacts in T-2*-weighted images at high fields by gradient-echo slice excitation profile imaging. *Magn. Reson. Med.* **39:**402–409.

Zarahn, E., Aguirre, G., and D'Esposito, M. 1997. A trial-based experimental design for fMRI. *NeuroImage* **6:**122–138.