PERGAMON

Neuropsychologia 40 (2002) 54–75

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI

Joseph T. Devlin [a,*], Richard P. Russell [b], Matthew H. Davis [b], Cathy J. Price [c], Helen E. Moss [b], M. Jalal Fadili [b], Lorraine K. Tyler [b]

[a] *Centre for Functional Magnetic Resonance Imaging of the Brain and the Department of Experimental Psychology, Oxford University, John Radcliffe Hospital, Headley Way, Oxford, UK*

[b] *Department of Experimental Psychology, University of Cambridge, Cambridge, UK*

[c] *Wellcome Department of Cognitive Neurology, University College London, UK*

Received 16 October 2000; received in revised form 9 March 2001; accepted 30 March 2001

Abstract

Patients with semantic impairments sometimes demonstrate category-specific deficits suggesting that the anatomical substrates of semantic memory may reflect categorical organisation, however, neuroimaging studies have failed to provide consistent data in support of a category-based account. We conducted three functional neuroimaging experiments to investigate the neural correlates of semantic processing, two with positron emission tomography (PET) and a third with functional magnetic resonance imaging (fMRI). The first experiment used a lexical decision task to search for brain regions selectively activated by concepts from four different categories—animals, fruit, tools, and vehicles. The second experiment used a semantic categorisation task to increase the demands on the semantic system and to look for evidence of consistent activations for the domains of natural kinds or man-made items. The final experiment was a replication of the semantic categorisation task using fMRI to increase the spatial resolution and statistical sensitivity of the experiment. The results of these experiments reliably identified a distributed neural system common to both natural kinds and artifacts but failed to find robust evidence of functional segregation by domain or categories. Category effects were neither reliable nor consistently present across experiments although some were consistent with previous studies. We discuss the implications of these findings, arguing that they are most consistent with a semantic system undifferentiated by category at the neural level.

Keywords: Semantic memory; Category specificity; PET; fMRI

1. Introduction

Patients with semantic memory deficits subsequent to brain injury or disease sometimes demonstrate category-specific deficits. That is, there appears to be a distinction between the semantic domains of natural kinds and man-made objects, where some patients are preferentially impaired on categories of natural kinds such as animals and fruit relative to man-made items such as tools and vehicles, while the reverse pattern can also occur (see [21,64,70] for reviews). Although such a dissociation would appear to implicate distinct neural substrates for different categories of knowledge, the nature of these neural substrates remains unclear.

One interpretation of these deficits, suggested by Warrington and colleagues [77–79], proposed that natural objects such as animals, fruits, vegetables, etc. are distinguished primarily on the basis of their visual semantic properties while man-made items such as tools and vehicles are distinguished primarily by their functions. Thus damage to perceptual (especially visual) information will result in a preferential impairment of natural kinds, while damage to functional information will lead to a deficit for man-made items, or artifacts. The existence of patients with natural kind deficits but without a corresponding deficit for perceptual information [11,20,43,49] challenges this account and has fuelled the development of alternative explanations for category-specific deficits. These include damage to cate-

* Corresponding author. Fax: +44-1865-222717.
E-mail address: devlin@fmrib.ox.ac.uk (J.T. Devlin).

0028-3932/01/$ - see front matter
PII: S0028-3932(01)00066-5

gory-specific neural processing regions [10,36] and widespread damage interacting with the structure of concepts within a distributed and non-fractionated semantic system [17,19,33,71]. These three accounts have very different implications concerning the underlying neural substrates for conceptual knowledge in different categories and domains.

Implicit in Warrington and colleagues' account is the assumption that perceptual and functional semantic information can be damaged independently in brain injury and thus it is reasonable to assume that there must be separable neural correlates of perceptual and functional semantic knowledge. Caramazza and Shelton's hypothesis goes even further, postulating individual brain regions specific to evolutionarily salient categories of objects such as animals, fruit, and tools [11,65]. In contrast, Tyler, Moss and colleagues suggested that segregation of the semantic system at the anatomical level is unnecessary since category-specific deficits can arise without such specialisation [19,71].[1] In this account, conceptual representations are described at a cognitive level in terms of an interactive system of semantic features (see also [10,45]). Features which tend to co-occur across concepts (e.g. ⟨*has fur*⟩, ⟨*has legs*⟩, ⟨*has a tail*⟩) support each other and thus are more resistant to damage than features without this support. Because these correlated features are not distributed equally across domains or categories of knowledge, some categories are more susceptible to semantic damage than others, leading to category-specific semantic impairments. The important point to note about this hypothesis is that it assumes that semantic representations are not anatomically differentiated by semantic content. The idea is that concepts are represented and processed in a distributed neural system which involves a number of brain areas including frontal, temporal, parietal and occipital regions. The components of this highly interactive system will be more or less involved depending on a number of factors, such as the nature of the input (spoken, written, pictures), the nature of the task (naming, reading, matching) and the additional non-linguistic cognitive demands required. Thus, for example, occipital cortex will be more involved when the stimuli consist of pictures than when they are spoken words, and inferior pre-frontal cortex will be more involved as processing demands increase [6] compared to when subjects are required to merely covertly recognise a stimulus. Nevertheless, no specific region or set of regions is dedicated to processing concepts in any specific conceptual domain (natural vs. man-made) or category (animals vs. fruits).

These competing accounts can be evaluated according to a range of criteria, but the most direct evidence for the neural substrates of conceptual knowledge comes from the functional neuroimaging literature.

1.1. Evidence from functional neuroimaging

Recently a number of studies have used positron emission tomography (PET) to look for an anatomical basis for differences between natural kinds and artifacts in normals (see [59] for a review). In nine PET studies a total of 22 individual areas have been identified: twelve were associated with natural kinds while eleven were associated with artifacts (see Table 1). One area, the left lingual gyrus, was associated with both natural kinds and artifacts, albeit in different studies (natural kinds in [44,56] and artifacts in [47,57]). Moreover, there is very little consistency across studies—sixteen regions were seen in only a single study. Consequently, although each study found regionally-specific neural correlates for either natural kinds or artifacts, this variability across studies makes it difficult to draw meaningful conclusions regarding anatomical specialisation. Interestingly, the most consistent finding in this literature is that a region of the left posterior middle temporal gyrus responds preferentially to tools [44] and this has been observed in seven (out of nine) other PET studies [16,44,47,50,51,56,72].[2] In addition four other areas have been present in two or more studies. These include parts of the anterio-medial temporal poles bilaterally and the right inferior parietal lobe for natural kinds as well as the left inferior frontal cortex for man-made items. Thus, the most robust evidence of category-specific neural activity appears to be specific to an individual category (tools) but there is some evidence implicating a small set of additional areas as well.

Category-specificity has also been investigated using functional magnetic resonance imaging (fMRI) with similarly diverse findings. In addition, these studies had a number of methodological limitations which make their interpretation particularly difficult. For instance, one study limited the data acquisition to four transverse planes parallel to AC-PC line and reported statistical results for individual subjects rather than for the group [67]. They found areas of activation specific to either animals or furniture bilaterally in the middle frontal gyrus and the superior temporal gyrus although not consistently across subjects for either category. Similarly, Chao et al. [12] reported only individual subject

[1] Note that a similar hypothesis by Gonnerman and colleagues [17,18] accepted the idea of semantic specialisation for perceptual and functional information and demonstrated that category specific semantic deficits could arise from *either* selective damage to perceptual or functional features *or* from widespread damage interacting with the relations between features.

[2] [35] re-analysed data from the [16] study focusing specifically on the frontal lobes. Thus although the paper does not report activity for tools in this area, it was present in the original analysis.

Table 1
A summary of the findings from nine PET studies of category specificity[a]

	[56]	[44]	[16]	[50]	[51]	[9]	[35]	[57]	[47]	Totals
Natural kinds										
Left fusiform	×									1
L. lingual gyrus	×	×								2
L. mid inf. temporal gyrus			×							1
L. ant temporal pole				×					×	2
R. ant temporal pole				×					×	2
R. inf. Parietal lobe				×	×					2
L. mid frontal gyrus					×					1
R. mid frontal gyrus						×				1
R. fusiform						×				1
L. thalamus								×		1
R. sup parietal lobe								×		1
R. post temporal cortex									×	1
Artifacts										
L. inf. Frontal cortex	×	×					×			3
L. post mid temporal gyrus		×	×	×	×	×		×	×	7
L. parahippocampal gyrus					×					1
L. supramarginal gyrus						×				1
R. sup temporal gyrus						×				1
R. thalamus						×				1
L. precentral gyrus							×			1
L. lingual gyrus								×	×	2
R. lingual gyrus								×		1
L. precuneus								×		1
R. cuneus								×		1

[a] Studies which showed a relative increase for either natural kinds relative to artifacts (at any level of significance) or the opposite contrast are marked with a cross (×).

analyses which primarily demonstrated inter-subject variability even for the findings they claimed were most consistent. Thompson-Schill et al. [69] did acquire whole brain images and report group statistics but only for a 3 cm^3 region-of-interest in the left fusiform area which they claimed was sensitive to perceptual information particularly when the stimulus was a natural kind. Thus no fMRI study of category-specificity has yet included a whole brain analysis for a group of subjects. Furthermore, the findings from the three existing studies are as inconsistent as those from PET.

What is the cause of this variability? Two methodological factors which may contribute to this inconsistency include stimuli and analysis differences:

1. *Stimuli:* Differences in the stimuli used to elicit semantic processing (written words, spoken words, or pictured objects) may contribute to the variable findings. In the neuropsychological literature, several studies have convincingly demonstrated that apparent category specific semantic deficits can arise from improperly controlled stimuli.[3] In particular, because natural kinds tend to be less familiar than artifacts, an uncontrolled set of stimuli can lead to a spurious natural kinds deficit cf. [27,55,68]. Similarly, differences between natural kinds and artifacts in terms of their visual complexity [28] and the structural similarity among members of the category [39] can also exaggerate deficits for natural kinds. Thus the differences observed in neuroimaging studies which have not controlled their stimuli may be due to confounding effects of one or more of these variables.
2. *Analysis:* Another factor likely to contribute to the growing body of unreplicated results is the fact that many studies do not correct their statistical maps for the number of independent comparisons made. In a typical whole brain PET study, there are on the order of 200 000 voxels each analysed independently, resulting in approximately 500 independent observations (or resels, see [81]). Consequently, an uncorrected alpha level of 0.05 ($Z > 1.96$, one-tailed) corresponds to a family-wise false positive value approaching unity ($P < 0.999$). Even when studies use an uncorrected $P < 0.001$ ($Z > 3.09$) there is still a false positive rate greater than 0.9. As a result, many of the reported differences between categories may be Type I errors (i.e. false positives) due to the liberal statistical thresholds adopted and this could certainly contribute to the variable findings across studies.

[3] Not all cateory-specific deficits, however, can be explained in terms of stimuli confounds. A number of studies have used carefully controlled stimuli and shown reliable category-specific semantic impairments (e.g. [7,11,49]).

1.2. Current studies

The purpose of the studies reported here was to investigate the neural correlates of different categories and domains of knowledge using both PET and fMRI. Our aim was to evaluate two main claims about the neural structure of conceptual knowledge. Is there neural specialisation according to category, domain or feature-type as [77] and [11] predict, or is conceptual knowledge represented within a distributed and non-differentiated neural system as has been previously argued by Tyler, Moss and colleagues [71]?

To this end we ran three studies, each using visually presented words. An important aspect of these studies is that words within categories were carefully matched on a large number of variables (e.g. familiarity, concreteness, frequency, number of letters, etc) to avoid confounds. The first experiment used lexical decision in PET to determine whether there were brain regions selectively activated by either specific categories (i.e. animals, fruit, tools, or vehicles) or by domains (ie, natural kinds or man-made items). The second experiment also used PET and focused on domain-specific (natural/man-made) activations using a semantic categorisation task to increase the processing demands on the semantic system. Finally, we used the semantic categorisation task with fMRI to increase the spatial resolution and the statistical sensitivity of the experiment. The two competing accounts outlined above generate different predictions for the outcome of these studies. The neural specialisation account [11,79] predicts differential activation in different brain regions as a function of category/domain/type of feature. In contrast, the conceptual structure account [48,71] predicts overlapping activation in the same temporal lobe regions for all categories and domains of knowledge.

2. Experiment 1: Lexical decision in PET

The purpose of this experiment was to investigate the neural basis of semantic processing at three levels: 1) areas of activation common to all categories of conceptual knowledge, 2) differences between domains of knowledge (i.e. natural kinds and man-made items), and 3) activations specific to individual categories.

2.1. Method

2.1.1. Participants

Twelve right-handed healthy male volunteers aged 21–51 (mean 30), 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.

2.1.2. Stimuli and design

Each subject participated in ten 90 s scans from five different conditions. In four of the conditions, the participants performed a visual lexical decision task (the test conditions) while in the fifth, they performed a letter detection task (the baseline). Lexical decision involves deciding whether each stimulus is a real word of English or a non-word.

In the test conditions all of the real words came from a single category, either animals, fruit, tools, or vehicles. To maximize the signal from word events relative to non-words, the first 45 s of the task (corresponding to maximal tarcer uptake, [66]) had ten words and only five non-words, presented in a pseudo-random order. In the second half of the scan, the same words appeared again in a different pseudo-random order with five non-words cf. [47]. The words were matched across scans, categories, and domains on the following variables, all of which are known to be important factors affecting semantic and/or lexical processing e.g. [26,53]:

1. Familiarity: This is a measure in which participants rate how often they have thought about or experienced the object referred to by the word. Ratings were taken either from the MRC psycholinguistic database [14] or were collected at the Centre for Speech and Language (C.S.L.), Cambridge.
2. Concreteness: This is a measure of the degree to which the referent of a word can be experienced through the senses (e.g. *cat* is a highly concrete word, while *honor* is highly abstract). Ratings were taken from the MRC database supplemented by our own norms collected at the C.S.L.
3. Neighborhood size: A measure of how many similar words exist in the language. The value refers to the number of words that differ from the source word by only one letter in any position. Values were from the Macquarie neighborhood database program.
4. Number of letters.
5. Number of syllables.
6. Written word frequency: The number of times per million that the word occurs in a large sample of written text. Written word frequencies came from the Celex database [3] and reflect British English usage.

Mean (and standard error) values for each category on these six variables are given in Table 2. All means were closely matched ($F < 1$) with the exception of written word frequency ($F(3,\ 76) = 2.81,\ P < 0.05$). However, post-hoc comparisons between categories did not reveal any significant differences.

Because many studies have demonstrated that pseudo-words (e.g. HICTION, BLATE) can produce greater activation than real words [37,62,63], the non-words used in this experiment all contained orthographically illegal letter strings in English (e.g., RFSTEN) to minimize their contribution to the detected signal.

Table 2
Experiment 1: Means (± S.E.M.) for familiarity, concreteness, frequency, neighbourhood size, number of letters and the number of syllables in each of the four categories of knowledge[a]

	Animals	Fruit	Tools	Vehicles	Overall
Familiarity	515	513	517	529	518
	(9.1)	(12.1)	(10.6)	(15.0)	(5.9)
Concreteness	609	611	601	607	607
	(4.0)	(6.6)	(6.9)	(8.0)	(3.2)
Frequency	28	10	11	45	23
	(8.1)	(3.2)	(2.6)	(17.3)	(5.1)
Neighbourhood	6	5	5	5	5
	(1.5)	(1.4)	(1.1)	(1.2)	(0.6)
Number of letters	5	6	6	5	5
	(0.4)	(0.4)	(0.3)	(0.4)	(0.2)
Number of syllables	2	2	2	2	2
	(0.2)	(0.2)	(0.1)	(0.2)	(0.1)

[a] Familiarity and concreteness ratings range from 100 (unfamiliar, abstract) to 700 (very familiar, concrete).

The baseline condition was a letter detection task where the subjects viewed consonant letter strings and were told to press the right button if the string contained the letter "*x*" and the left button if it did not. The purpose of the baseline was two-fold. First, we wanted a non-lexical task, which incorporated as many of the cognitive components of the lexical decision task as possible. Because this baseline included orthographic visual stimulation, an identical motor response, and presumably similar demands on sustained attention and working memory, it allowed us to minimise the difference in test and baseline signal due to these non-lexical properties of the task. The second reason for including a baseline was to allow us to identify *common* areas of activation across the four semantic categories. Without a baseline, the design would only allow one to look for *differences* between the conditions. In other words, one of the explicit goals of this experiment was to identify regions which were activated by all categories of conceptual knowledge.

2.1.3. Procedure

Stimuli were presented using the experimental software DMDX [22] running on a PC under Windows 98. Items were presented for 500 ms in 26 point Ariel font and subjects had 2500 ms between trials to respond. Stimuli were presented on a video monitor approximately 1 metre from the subject's head. All stimuli were displayed in black on a white background in a dimly lit room. Participants were asked to press the right button to indicate a word in the lexical decision task or the presence of an "*x*" in the baseline task. They pressed the left button to indicate a non-word or the absence of an "*x*," respectively. Participants were encouraged to proceed as quickly and accurately as possible. Accuracy and reaction times (to the nearest millisecond) were recorded.

Scans were performed at the Wolfson Brain Imaging Centre in Cambridge, England on a GE Advance PET Scanner (General Electric Medical Systems, Milwaukee, Wisconsin). 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 acquisition. Each subject received a bolus of 300 MBq of H_2O^{15} before each scan for a total radiation exposure of 4.2 mSv. The emission data was acquired with the septa retracted (3D mode) and reconstructed using the PROMIS algorithm [42] with an unapodised Colsher filter. Corrections were applied for randoms, scatter, attenuation and dead time. The voxel sizes were $2.34 \times 2.34 \times 4.25$ mm.

Functional images were realigned [23] as implemented in Statistical Parametric Mapping (SPM99, Wellcome Institute of Cognitive Neurology, www.fil.ion.ucl.ac.uk). Translation and rotation corrections did not exceed 6 mm and 2.5°, 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. These parameters were then applied to the functional images [1,2] and the image was resampled into isotropic 2 mm^3 voxels. 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 (ie a fixed-effects model) using the general linear model [24]. Results are reported at a $P < 0.05$ level after correcting for multiple comparisions [82] although the statistical threshold was then lowered to $P < 0.001$ (uncorrected) for comparision to the existing literature.

Table 3
Experiment 1: Mean (± S.E.M.) response times (in msec) and error rates for the four lexical decision conditions and the letter detection condition

	Animals	Fruit	Tools	Vehicles	Letter Detection
RTs	552 (22)	553 (17)	552 (20)	538 (17)	612 (24)
Error rates	1% (0.4%)	1% (0.4%)	2% (0.8%)	0% (0.3%)	5% (0.9%)

2.2. Results and discussion

2.2.1. Behavioural data.

Table 3 shows the mean reaction times (and the standard error) and error rates for each of the five conditions. The consistently low error rates indicate that the subjects were correctly performing both tasks, but are too low to analyse. Response time data were inverse transformed to reduce the influence of outlying data points before conducting inferential statistics [73]. There was a reliable effect of task on reaction time demonstrating that participants responded more quickly in the lexical decision task than the letter detection task ($F_2(84) = 49.3$, $P < 0.0001$). A one-way ANOVA with category as the independent variable (Animals, Fruit, Tools, Vehicles) found no effect of condition ($F_2(3, 236) < 1$ indicating that the four test conditions were of comparable difficulty, and therefore that overall difficulty can not be a confounding factor in interpreting the patterns of activation for each category.

2.2.2. Imaging data

2.2.2.1. Common semantic activation. To examine those activations which were common to the four categories of knowledge, we computed a main effect of semantics relative to baseline (i.e. [A]nimals + [F]ruit + [T]ools + [V]ehicles − [B]aseline) and inclusively masked it with each of the individual contrasts (A-B, F-B, T-B, V-B) at an uncorrected $P < 0.05$ threshold. In this fashion we guaranteed that the effects were present in all four of the semantic conditions and were not the result of large activations present in only a subset of conditions.

The results are presented in Table 4 and Fig. 1a. Activations which were reliable at either the voxel or cluster level after correcting for multiple comparisons are reported in the top portion of the Table (and shown in red in the figure). All other activations (shown in yellow in the figure) only reached a voxel level significance of $P < 0.001$, uncorrected. For each activation, the Table presents the co-ordinates of the peak, the t-statistic at that co-ordinate, the size of the cluster in 2 mm^3 voxels, and the number (n) of individual subjects (out of 12) showing the effect at an uncorrected $P < 0.1$ level.

There were two areas of reliable activation at a corrected significance level. In the left hemisphere one very large area included both medial and lateral aspects of the temporal lobe and extended into the inferior frontal lobe. This activation extended from the inferior aspect of the medial temporal lobe (including the uncus, amygdala, and hippocampus) to the medial anterior temporal pole; spreading into the lateral surface of the anterior and posterior middle temporal gyrus, the inferior frontal gyrus and frontal operculum. In the right hemisphere, there was also activation in the anterio-medial temporal pole which extended from the inferior surface to the superior aspect.

There were nine other activations at an uncorrected $P < 0.001$ level. These included left hemisphere peaks on the superior frontal, orbito-frontal and inferior frontal gyri. In the right hemisphere there were peaks in the insula, the anterio-medial temporal pole, anterior and posterior regions of the superior temporal gyrus, and the lingual gyrus.

2.2.2.2 Domain-specific activations. Increases for natural kinds relative to man-made items were evaluated by contrasting the two natural kinds categories, animals and fruit, with the two man-made object categories, tools and vehicles (i.e. (A + F)–(T + V)). This contrast was masked with six simple contrasts to ensure that each of the natural kind categories was more active than both of the man-made object categories (i.e. A-T, A-V, F-T, F-V) as well as the baseline (i.e. A-B and F-B). Similarly, the effects of man-made items relative to natural kinds were evaluated using a contrast of (T + V)–(A + F) masked (at $P < 0.05$ inclusive)[4] with T-A, T-F, V-A, V-F, T-B and V-B.

There were no reliable effects of domain after correcting for multiple statistical comparisons but there were domain-specific activations at an uncorrected $P < 0.001$ level (see Table 5). Natural kinds preferentially activated the right inferior precentral gyrus. Man-made items, on the other hand, activated regions in the left middle temporal gyrus, the right inferior frontal gyrus

[4] All masking in this paper was inclusive and used a threshold of $P < 0.05$, uncorrected. In other words, an effect had to be present in each masking contrast at least with an uncorrected probablilty of 0.05.

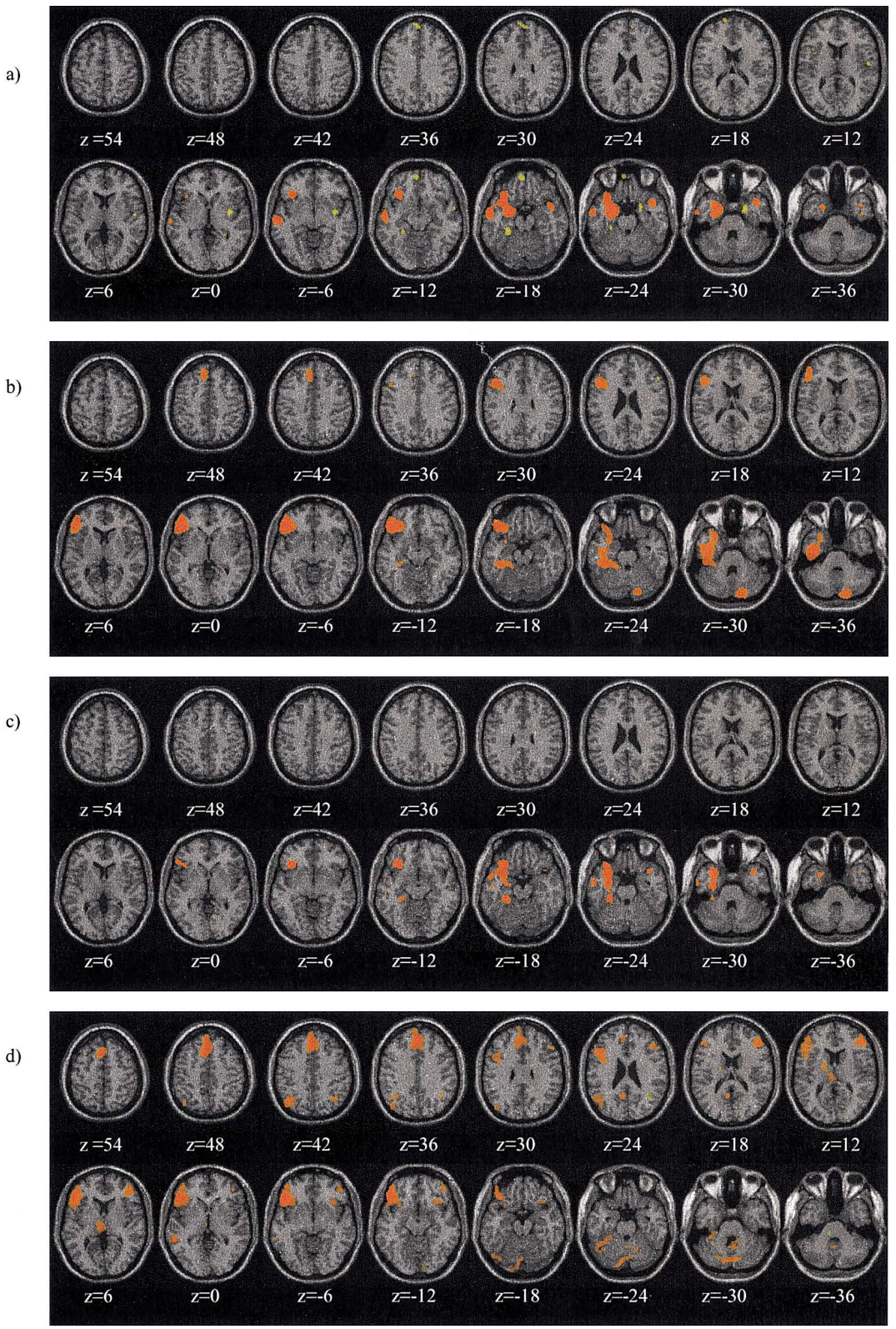

Fig. 1. The areas activated in the Semantic—Baseline contrasts in the (a) PET lexical decision experiment, (b) PET semantic categorisation experiment, (c) conjunction analysis of the two PET experiments and (d) in the categorisation experiment in fMRI. Activations are displayed superimposed on a structural image. Red areas were reliability ($P < 0.05$) active after statistical correction while yellow areas were activate at an uncorrected $P < 0.001$ threshold. Note that in the fMRI data there is less activation in the temporal lobes than in the PET data due to BOLD signal loss in areas adjacent to air-filled sinuses. The image is in neurological convention ($L = L$).

Table 4
Experiment 1: The results of the analysis of common semantic areas[a]

Description	x	y	z	SPM$\{t\}$	Extent	n/N
Activations reaching corrected significance levels						
Temporal Lobes						
L. anterio-medial temporal lobe	−28	2	−26	5.82	2516	9/12
	−34	16	−20	4.94		8/12
	−62	−20	−6	4.77		6/12
R. anterio-medial temporal pole	42	12	−26	4.34	404	7/12
	32	6	−50	3.76		5/12
	44	−8	−40	3.44		3/12
Activations significant at an uncorrected $P<0.001$ level						
Temporal Lobes						
R. anterio-medial temporal pole	22	0	−28	4.40	165	6/12
R. superior temporal gyrus	52	−10	10	3.61	36	4/12
R. superior temporal gyrus	62	2	−12	3.64	22	6/12
R. lingual gyrus	30	−44	0	3.49	10	3/12
Frontal Lobes						
L. medial superior frontal gyrus	−6	58	38	3.80	151	4/12
L. superior frontal gyrus	−12	66	18	3.45	43	5/12
L. orbito-frontal gyrus	−4	56	−16	3.75	155	5/12
L. inferior frontal gyrus	−40	14	10	3.25	5	6/12
R. insula	36	−4	−2	4.15	116	6/12

[a] The upper half of the table shows the two activations which were reliable after correcting for multiple comparisons ($T>4.74$, extent ≥ 320) while the lower half shows activations reliable at an uncorrected $P<0.001$ level. For each area the co-ordinates of the peak voxel are reported along with the SPM$\{t\}$ value at that voxel, the size of the activation in 2 mm^3 voxels, and the number (n) of individual subjects (out of N) showing the effect at an uncorrected $P<0.1$ level.

(the homologue of Broca's area in the right hemisphere), and the right cuneus.

2.2.2.3. Category-specific activations. Finally, the effects of individual categories were evaluated in an analogous fashion. To identify areas where a category was more active than all other categories, the category was contrasted with the other three categories and masked with each effect. For example, to determine whether there were any activations specific to the animal category, we contrasted animals to the other three categories [A − (F + T + V)] and then used masking to verify that the category was more active than each of the other categories (A-F, A-T, A-V) as well as the baseline (A-B). Each of the four categories was analysed in this fashion and the results are presented in Table 6.

There were no reliable effects of either domain or category after correcting for multiple statistical comparisons. The only differences identified were observed with an uncorrected α-level of 0.001 (after correction, $P<0.995$). In spite of this very high false positive rate, it is worth noting that one of these activations has also been demonstrated in two previous studies. When comparing natural kinds to artifacts, Mummery et al. [50] observed a left anterio-medial temporal pole activation in a verbal category fluency task while Moore and Price [47] identified this same region for both naming and matching black and white line drawings of natural kinds relative to artifacts, although this effect disappeared when the pictorial stimuli were appropriately coloured. Thus, when we used a lenient statistical threshold, the activation we observed in the left anterio-medial temporal pole for animals was consistent with two other studies. The remaining domain- and category-specific activations, however, were inconsistent with the existing literature and consequently may simply be Type I errors.

2.2.2.4. Power analysis. To establish that the experiment had sufficient sensitivity to detect differences between domains or categories, we conducted a two-stage power analysis. The first step looked for distinct distributions of t-values corresponding to inactive and active voxels to estimate the mean effect size [31] while the second used a non-central F-test to evaluate the sensitivity of the experiment based on this effect size [74]. First, the t-statistic for all voxels in a particular contrast was plotted in a histogram. Then, the curve was fitted with a mixture of two Gaussian distributions representing the t-values of individual voxels under the null hypothesis (H_0) and under an active hypothesis (H_1), shown in Fig. 2. The mixing ratio (λ) and effect size (δ) which best fit the data were estimated using maximum likelihood estimation. For the main effect of lexical decision relative to

Table 5
Experiment 1: The results of the domain specific analyses (ie, natural kinds vs. artifacts and vice-versa)[a]

Description	x	y	z	SPM$\{t\}$	Extent	N/N
Activations significant at an uncorrected $P<0.001$ level						
Natural kinds relative to man-made items						
R. inferior pre-central gyrus	56	−18	18	3.80	21	6/12
Man-made items relative to natural kinds						
R. inferior frontal gyrus	56	26	16	3.41	3	4/12
L. middle temporal gyrus	−64	−34	−12	3.58	16	4/12
R. cuneus	14	−74	14	3.25	2	2/12

[a] All activations were present at the uncorrected $P<0.001$ level and none reached corrected significance ($P<0.05$). Activations are reported as in the previous table.

letter detection, we found $\lambda = 0.09$ and $\delta = 2.44$. That is, the mean effect size elicited an approximately 2.5% mean change in regional cerebral blood flow (rCBF) and 9% of the voxels in the brain were activated in this condition. Interestingly, when this technique was applied to the domain-specific and category-specific contrasts, the minimisation algorithm failed to find a mixture which fit the histogram. In other words, all of the values in the histogram fit a single Gaussian distribution under the null hypothesis. This result is consistent with the claim that the activations observed in the domain and category specific contrasts at an uncorrected level may represent false positives.

To determine whether our paradigm was suitably sensitive we assessed its power to detect rCBF changes in the second step. Based on the estimated mean effect size of 2.5% in the words versus baseline contrast, we used a non-central F test [74] to calculate that our experiment had a 61% power for detecting true positives of this size or greater at an uncorrected statistical threshold of $P < 0.001$. In other words, if individual domains or categories of words produced rCBF changes of $\geq 2.5\%$, those effects would be detected 61% of the time. If, on the other hand, these effects were smaller, then the experiment had less power.

2.2.3. Summary of results of Experiment 1

This experiment identified several areas of reliable activation common to all four categories of knowledge. These included a large left hemisphere activation extending from the medial and lateral surfaces of the temporal lobe into the inferior frontal lobe as well as a more focal activation in the right anterio-medial temporal pole. In addition, there were a few areas that were differentially activated as a function of domain or of category, but only at an uncorrected level of significance. Although the animal-specific activation in the left anterio-medial temporal lobe has been seen previously [50,47], none of the other areas were consistent with the existing literature. Furthermore, the power analysis suggested that the activity in these contrasts is consistent with the null hypothesis. In addition, it revealed a 61% likelihood of detecting effects of 2.5% or more. Taken together, these findings suggest that either there are no differences in activation across domains or categories or that such differences are small and cannot de detected by the current experiment or others like it.

3. Experiment 2: Semantic categorisation using PET

Experiment 1 revealed no reliable differences in neural activation for the domains of natural kinds and man-made items at a corrected statistical threshold. There were, however, some differences at an uncorrected level. Although previous studies [52,61,63] have demonstrated that our task—lexical decision—activates both semantic and phonological areas, it could be argued that category specific effects only arise when the demands on the semantic system are increased. Therefore, we carried out a second PET experiment, using a semantic categorisation task designed to place greater demands on the semantic system than lexical decision, thus maximising the probability of detecting any small but robust effects. In addition, this second experiment enabled us to determine whether the activations observed in the first study were robust and replicable. Finally, in this study we increased the statistical power of the experiment by increasing the repetitions of each condition.

3.1. Methods

3.1.1. Participants

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.

Table 6
Experiment 1: The results of the category-specific analyses (i.e. animals, tools, fruits and vehicles)[a]

Description	*X*	*y*	*z*	SPM{*t*}	Extent	*N*/N
Activations significant at an uncorrected level						
Animals relative to fruit, tools, and vehicles						
L. middle temporal gyrus	−60	−68	18	3.74	43	5/12
L. anterio-medial temporal lobe	−32	8	−16	3.23	6	4/12
R. medial temporal lobe	42	−40	−2	3.47	15	4/12
Brainstem	−2	−18	−40	3.53	79	5/12
Fruit relative to animals, tools, and vehicles						
L. middle frontal gyrus	−42	44	18	4.09	49	6/12
L. post-central gyrus	−60	−20	24	3.24	9	3/12
R. primary visual cortex	24	−102	−14	3.35	2	3/12
R. primary visual cortex	16	−98	12	3.33	4	5/12
Tools relative to animals, fruit, and vehicles						
L. cuneus	−12	−80	34	4.22	22	6/12
R. inferior frontal gyrus	60	22	14	3.74	66	7/12
Vehicles relative to animals, fruit, and tools						
L. middle temporal gyrus	−64	−16	−16	3.66	72	3/12
	−62	−28	−22	3.26		
R. middle temporal gyrus	64	−16	−14	3.77	42	5/12
R. fusiform gyrus	50	−24	−24	3.40	19	2/12
L. middle frontal sulcus	−40	52	0	3.54	31	5/12
R. medial superior frontal gyrus	16	32	50	3.26	2	4/12
R. cerebellum	32	−48	−44	3.29	8	2/12

[a] All activations were present at the uncorrected $P<0.001$ level and none reached corrected significance ($P<0.05$). Activations are reported as in the previous table.

3.1.2. Stimuli and design

This experiment used a semantic categorisation task in which 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. For instance, subjects made a "same" response to the sequence, "dolphin, seal, walrus—OTTER" and a "different" response to "dolphin, seal, walrus—BANANA" (see Table 7 for further examples).

There were two semantic conditions, natural kinds and man-made items, as there were too few items to include separate conditions for individual categories after matching along relevant dimensions. In each condition, there were equal numbers of "same" and "different" trials. Importantly, in the "different" trials, the target word came from the same domain (but a different category) as the cue words. Thus in each condition all of the words were from a single domain.

The baseline in this experiment was a letter categorisation task which shared the same stimulus and response characteristics as the semantic categorisation 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 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.

The stimuli in the semantic task were matched on the dimensions of word frequency, familiarity, and letter length, as in Experiment 1, using the Celex [3] and MRC Psycholinguistic databases [14] and C.S.L norms. Natural kind words had a mean ($\pm$ SD) familiarity of

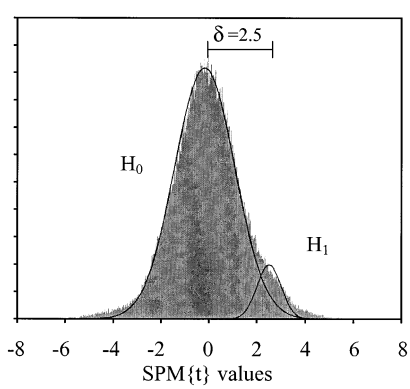

Fig. 2. A histogram of the SPM{*t*} values from the Lexical decision—Baseline contrast. Superimposed on this are two Gaussian distributions. The one on the left represents *t*-values from voxels under the null hypothesis (H_0) while the one on the right represents values under the alternative (H_1) hypothesis, namely activated voxels. The mixing ratio (λ) is reflected by the relative amplitudes of the Gaussians. The effect size (δ) is shown as the distance between their means.

Table 7
Experiment 2: 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"

502 ($\pm$ 53.0) and a mean frequency of 17 ($\pm$ 25.7) and artifact words had a mean familiarity of 502 ($\pm$ 87.1) and a mean frequency of 17 ($\pm$ 15.8). The items in the letter task were matched on letter length with those in the semantic task. In both conditions, stimuli ranged from 3–9 letters in length with a mean ($\pm$ S.D.) of 5.4 ($\pm$ 1.0) letters. There were 96 semantic trials and 48 letter trials.

3.1.3. Procedure

Presentation and timing of stimuli was controlled by the DMDX software. Cue words were in lower case and the target was in upper case to signal a response. Subjects made a "same" category response by pressing the left mouse button and a "different" response by pressing the right mouse button. The mouse was always held in the subject's right (dominant) hand and both reaction times and accuracy were recorded.

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 the response. Thus each trial lasted 3750 ms. In pilot testing outside the scanner with a different set of participants, these timing parameters were determined to yield quick and accurate responses (see [18]). Within a scan, subjects saw 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. Only 45 s of stimuli were presented to coincide with the critical period of tracer uptake and thus optimise the sensitivity of the design [66]. All subjects participated in twelve 90 s scans, four from the natural kinds categorisation condition, four from the artifacts categorisation condition, and four from the letter categorisation baseline. The conditions were presented systematically such that no subject saw them in the same order. Scans were performed at the Wolfson Brain Imaging Centre in Cambridge, England.

Functional images were pre-processed as in Experiment 1 with translation and rotation corrections of less than 5 mm and 3°, respectively. Each image was normalised and smoothed with a 16 mm FWHM kernel. Finally, the SPM software was used to compute a within-subjects analysis using the general linear model.

3.2. Results and discussion

3.2.1. Behavioural data

The subjects' mean ($\pm$ SD) reaction time and error rate in the semantic categorization task were 779 ($\pm$ 81.6) ms and 9 ($\pm$ 12.5)%, respectively. In the letter categorization task the mean RT and error rate were 657 ($\pm$ 59.8) ms and 5 ($\pm$ 7.6)%, respectively. Thus the participants were reliably faster in the letter task relative to the semantic task ($F_2(1, 142) = 185.9$, $P < 0.0001$) and their low error scores indicated that they were performing the task adequately. This suggests that the subjects found the baseline task easier than the semantic task, which makes it an appropriate baseline for identifying the semantic system. An analysis of the two semantic conditions revealed no reliable differences between domains (natural kinds = 763 $\pm$ 72.7 ms, artifacts = 794 $\pm$ 87.3 ms; $F_2(1, 92) = 3.079$, $P > 0.1$).

3.2.2. Imaging data

3.2.2.1. Common semantic activations. To examine the activation common to both domains, we computed a main effect of semantic categorisation relative to letter categorisation (i.e. [N]atural [K]inds + [Art]ifacts − [B]aseline) and masked it with the individual contrasts (NK-B, Art-B).

The semantic categorisation task produced three areas of activation at a corrected statistical threshold relative to the baseline: an area located in the right cerebellum, another on the medial surface of the left superior frontal gyrus, and finally a massive region of over 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. 1b). To identify individual peaks within this large volume the height threshold (the uncorrected voxel-level P-value) was increased from the default of 0.001 to 0.0001. This effectively broke the single large volume of activation into four separate regions (see Table 8). These included two frontal activations; one in the inferior frontal gyrus (the frontal operculum) extending into Broca's area and one slightly more anterior and posterior extending from Broca's area into the middle frontal gyrus. In addition, there were two left temporal activations, one with a peak in the inferior temporal gyrus and another on the anterio-medial temporal pole. All activations were significant at either the cluster and/or voxel levels after corrections for multiple comparisons [25,81] except the temporal pole activation which reached a trend at the cluster level ($P < 0.1$) but was not significant at the voxel level.

Table 8
Experiment 2: The results of the analysis areas common to both natural kind and artifact conditions[a]

Description	X	y	z	SPM{t}	Extent	n/N
Activations reaching corrected significance levels						
Temporal lobes						
L. inferior temporal gyrus	−40	−12	−34	6.68	906	7/8
	−36	−32	−20	4.95		
L. anterio-medial temporal pole	−28	12	−30	4.42	84	4/8
Frontal lobes						
L. inferior frontal gyrus	−48	36	−4	7.01	1824	7/8
	−44	18	26	4.91		5/8
L. medial superior frontal gyrus	−6	34	46	4.45	90	6/8
Cerebellum						
R. medial posterior surface	18	−84	−32	5.63	305	6/8

[a] The extents are shown for a height threshold of $P<0.0001$. All activations are reliable at either a voxel level ($T>4.8$) and/or a cluster level (extent ≥ 90 voxels) except the peak in the left anterio-medial temporal pole which is only a trend ($P<0.1$) at the cluster level. Activations are reported as in previous tables.

3.2.2.2. Domain-specific activations. Activations specific to natural kinds were calculated as NK—Art masked with NK—Baseline while activations specific to artifacts were Art – NK masked by Art—Baseline.[5] The results are shown in Table 9. As in the previous experiment, no activations were reliable at a corrected threshold although there were differences at an uncorrected threshold ($P < 0.001$). Relative to artifacts, natural kinds induced small regions of activation in a middle region of the left superior temporal sulcus, the right amygdala, and right cerebellum. The Artifacts—Natural Kinds contrast yielded activity in a left posterior middle temporal region that has previously been associated with tools and man-made kinds (see Table 1). In addition there were two other left hemisphere activations in the cuneus and anterior fusiform gyrus, a region associated with semantics.

Activation in the left cuneus for artifacts has also previously been reported [57], but the peak was more medial and ventral (approximately 4 cm distant) than the co-ordinates we observed. Thus, other than the left posterior middle temporal gyrus activation, the domain-specific activations observed in this experiment were inconsistent with both the existing literature and the previous experiment. Moreover, these activations did not approach significance when appropriate statistical corrections were applied.

To assess the probability of obtaining false positives we again estimated the mixture of the null and alternative hypotheses from our data. For the semantics versus baseline contrast, there was an estimated mixing ratio (λ) of 0.076 and an effect size (δ) of 2.9. In other words, there was a mean increase in rCBF of approximately 3% in the activated voxels and these constituted almost 8% of the total brain volume. These results are similar to those observed in Experiment 1. The same technique was applied to the domain contrasts and as before, the algorithm failed to find a mixture of Gaussians which fit the histogram. Thus the data from the two domain specific contrasts were consistent with the null hypothesis indicating that the activations observed (with an uncorrected statistical threshold) were likely to be false positives.[6] As a final step, we assessed the power of this experiment to detect rCBF changes to determine whether our paradigm was suitably sensitive. Based on the observed mean effect size of 3% in the words versus baseline contrast and a $P < 0.001$ uncorrected α-level, we calculated our power $(1-\beta)$ as 0.72, or as a 72% power for detecting true positives of 3% or more.

3.2.2.3. Comparing lexical and semantic decisions. The main difference between Experiments 1 and 2 was the additional activation in the frontal and cerebellar regions in Experiment 2, possibly due to the additional working memory, attentional, and semantic demands of the categorisation task relative to lexical decision (cf. [46,54]). A direct statistical comparison between Lexical—Baseline and Semantic—Baseline in the two experiments, however, revealed no significant difference in activation patterns. The only differences

[5] Masking in this case guaranteed that an effect was due to increases in activation rather than decreases. For instance, areas identified as more active for natural kinds than artifacts could be due to either increased activation for natural kinds or decreased activation for artifacts. To identify increases, the effects had to be present relative to the baseline as well.

[6] Even so, the left posterior middle temporal gyrus activation for artifacts may be a true activation although it is not detectable in this analysis due to its small size (i.e. 1 voxel out of 249988). The same is true for the left anterio-medial temporal activation in the previous experiment (6 voxels out of 242402).

Table 9
Experiment 2: The results of the domain specific analyses[a]

Description	x	y	z	SPM$\{t\}$	Extent	n/N
Activations at uncorrected level of significance						
Natural kinds relative to man-made items						
R. cerebellum	34	−64	−50	3.62	15	3/8
L. superior temporal sulcus	−70	−24	−4	3.28	4	3/8
R. amygdala	26	−8	−14	3.27	5	3/8
Man-made items relative to natural kinds						
L. anterior fusiform	−42	−28	−14	4.09	115	4/8
L. cuneus	−28	−90	38	3.68	9	4/8
L. posterior middle temporal gyrus	−56	−62	−4	3.27	1	3/8

[a] The table shows areas of activation for natural kinds relative to artifacts and vice-versa. All activations were present at the uncorrected $P<0.001$ level and none reached statistical significance ($T>4.8$, extent ≥ 325). Activations are reported as in the previous table.

were at an uncorrected $P<0.001$ level with greater activation in the categorisation experiment in the left middle frontal gyrus ($x=-46$, $y=18$, $z=28$, SPM$\{t\}=3.75$), left anterior inferior frontal gyrus ($x=-46$, $y=46$, $z=0$, SPM$\{t\}=3.67$) and right medial cerebellum ($x=10$, $y=-84$, $z=-34$, SPM$\{t\}=4.1$). There were no differences in the temporal lobe, even at uncorrected $P<0.001$. Therefore, as previously demonstrated, lexical decision and semantic decision both activate common semantic areas in the temporal lobes.

To determine the extent to which common activations were obtained in the two studies, we used a conjunction analysis [58]. To look for common semantic—baseline activations, we computed the conjunction of the two semantic versus baseline contrasts (A + F + T + V-B_1, NK + Art-B_2) and masked it by the six individual contrasts (A-B_1, F-B_1, T-B_1, V-B_1, NK-B_2, Art-B_2), where B_1 refers to the letter detection baseline in the first experiment and B_2 refers to the letter categorisation baseline in the second. The analysis indicated reliable activation after correcting for multiple comparisons in the anterio-medial temporal lobes bilaterally with a more posterior and lateral extent on the left and in the left inferior frontal cortex (see Fig. 1c). The two areas which only reached an uncorrected threshold ($P<0.001$) were proximal to the larger activations. Thus these experiments identified anterior portions of the temporal lobes bilaterally as well as regions of the left inferior frontal cortex and left posterior temporal cortex as being important for semantic processing.

A conjunction of the contrasts comparing natural kinds to artifacts from the two experiments (A + F-T-V, NK-Art masked with the individual contrasts) revealed *no* voxels in common even at an uncorrected $P<0.001$ level. The same was true for the conjunction of the contrasts comparing artifacts to natural kinds—no voxels were present even at an uncorrected level. Because these contrasts were independent, this meant that there were no voxels activated at a 0.032 uncorrected level in either domain common to both experiments. In other words, even at a very lenient criterion, there was no evidence of domain-specific activations common to both experiments.

3.2.3. Summary of results

The first two experiments produced fairly consistent results; namely, they identified a network of left hemisphere fronto-temporal regions, as well as a single region in the right anterio-medial temporal pole, commonly activated in tasks requiring semantic processing of written words. There were no domain differences at a corrected statistical threshold and although both experiments identified a number of candidate areas for domain-specific processing at an uncorrected level most were inconsistent with the existing literature. The two regions which had been reported previously, namely in the left anterio-medial temporal pole for natural kinds (Exp. 1) and the left posterior middle temporal gyrus for man-made items (Exp. 2), were small effects in our data and were not consistently present across experiments.

4. Experiment 3: Domains of knowledge in fMRI

In a third study, we essentially replicated the second PET study in an fMRI experiment, with only minor adjustments in the paradigm which were necessary by virtue of data acquisition differences (see the Methods section below). The main motivation for the fMRI study was to address a subset of the issues relating to category specificity while taking advantage of the better spatial resolution in fMRI compared to PET. In particular, we can ask whether the left posterior middle temporal region—which has been shown to be selectively activated in response to tools in earlier studies [44]—is preferentially activated for man-made concepts. If category-specific neural responses are small, the spatial resolution in PET may be insufficient to

differentiate functional differences among anatomically proximal areas (e.g. [12]) and our PET studies may simply have lacked sufficient statistical power to detect them. Thus, by using a technique which gives greater spatial resolution and statistical power, we can maximise the opportunities for detecting small regional differences in activation if they exist.

Although many studies have shown good replicability between PET and fMRI [13,15,41], very few have tested replicability with language paradigms. Indeed, there is evidence of important differences between the modalities which are particularly relevant to language tasks. These include the effects of serial data acquisition with transient stimuli [60,76] and the loss of BOLD signal near air-tissue interfaces at high magnetic field strengths [18,40]. In an attempt to minimise these problems we collected data from multiple points in the peri-stimulus interval per condition and defined a priori regions-of-interest (ROI) based on the results of the second PET experiment to enhance statistical signal detection in regions of macroscopic susceptibility artifacts.

4.1. Method

4.1.1. Participants

Eight right-handed, healthy volunteers aged 22–64 (mean 34), all of whom spoke British English as their first language, participated. There were 3 women and 5 men. Each gave informed consent after the experimental methodology was explained. Volunteers were screened for magnetic resonance compatibility prior to entering the scanning room.

4.1.2. Stimuli and design

Because this was a replication of the previous experiment, the same categorisation tasks and stimuli were used. Each subject was tested in two separate sessions thus doubling the number of stimuli needed. So instead of 96 semantic trials and 48 letter trials, participants saw 192 semantic trials and 96 letter trials. These additional stimuli were constructed in the same way as the original set, and matched on the dimensions of word frequency, familiarity, and letter length. Natural kind trials had a mean ($\pm$SD) familiarity of 496 ($\pm$ 32.9) and a mean frequency of 15 ($\pm$14.2) while artifact trials had a mean familiarity of 503 ($\pm$47.1) and a mean frequency of 16 ($\pm$10.2). The items in the letter task were matched on letter length with those in the semantic task. In both conditions, stimuli ranged from 3–9 letters in length with a mean ($\pm$S.D.) of 5.4 ($\pm$0.9) letters.

4.1.3. Procedure

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 eight trials (at 3750 ms each) per block rather than 12, as in the PET experiment.

During each imaging session 180 images were collected. An additional four 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 peri-stimulus time, decreasing the potential contribution of artifactual biases in the signal estimation (cf. [60]).

All scans were carried out using the Varian-Siemens 3 Tesla MRI scanner at the Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB) in Oxford. A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage head radio-frequency 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 (3T), a manual shim was set up for each subject using eight terms (three linear and five quadratic) to reduce magnetic field inhomogeneities and a TE of 30 ms was used to jointly optimize BOLD contrast-to-noise and image signal-to-noise while minimizing intra-voxel de-phasing.

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 normalised to the EPI template transforming them onto the MNI mean brain. Finally, each image was smoothed with a 6 mm FWHM Gaussian filter.

Technical difficulties resulted in the loss of the behavioral data from three (out of sixteen) scanning sessions. Consequently the functional images for these sessions were not included in any of the analyses because we could not verify that the task was being adequately performed. The functional imaging data from 13 sessions (a total of 2340 scans) were analyzed using a within-subject analysis. The data were temporally smoothed and the estimated response was modeled using a box-car function convolved with a canonical HRF [80]. Temporal derivatives were also included to better model regional timing deviations from the canonical HRF.

4.2. Results and discussion

4.2.1. Behavioral data

Subjects performed the task comparably to those who participated in the previous experiment. Their

Table 10
Experiment 3: The results of the fMRI semantic categorisation experiment showing areas commonly activated by natural kinds and artifacts[a]

Description	*x*	*y*	*z*	SPM{*t*}	Extent	*n*/N
Activations significant at a corrected level						
Temporal lobes						
L. superior temporal sulcus	−58	−38	−2	5.81	166	5/13
	−64	−44	−2	4.28		
Frontal lobes						
L. inferior frontal gyrus	−46	30	−8	11.86	3023	9/13
	−48	36	4	11.26		
	−48	24	24	10.17		
R. inferior frontal gyrus	48	40	12	9.71	901	8/13
	48	46	−8	6.98		
	46	30	28	5.23		
R. inferior frontal gyrus	34	22	−8	6.12	239	7/13
L. medial superior frontal gyrus	−2	40	40	11.22	2000	9/13
	−6	18	50	10.19		
	−4	26	46	9.20		
L. superior frontal gyrus	−20	26	46	4.95	30	5/13
L. middle frontal gyrus	−44	40	26	5.61	6	4/13
Parietal lobes						
L. T-O-P junction	−38	−62	40	8.07	703	9/13
	−44	−74	40	6.68		
	−46	−62	24	5.01		
R. T-O-P junction	38	−62	40	5.43	111	4/13
Other areas						
L. precuneus	−6	−56	22	6.79	119	5/13
Bilateral cerebellum	−6	−82	−28	7.82	767	6/13
	8	−82	−28	7.58		
	6	−78	−20	5.14		
L. cerebellum	−36	−44	−32	5.68	396	4/13
	−38	−50	−24	5.54		
	−52	−66	−20	4.64		
L. thalamus	−4	−22	8	6.49	461	6/13
	−18	−6	14	4.78		

[a] The extents are shown for a height threshold of $P<0.0001$. All activations were reliable at the voxel level ($T>4.5$). Activations are reported as in previous tables except that '*n*' refers to the number of sessions (rather than subjects) which showed activation at $P<0.01$ (uncorrected) to reflect the greater sensitivity of the fMRI data.

mean ($\pm$SD) reaction time and error rate in the semantic categorization task were 803 ($\pm$165.4) ms and 8 ($\pm$14.3)%, respectively. In the letter categorization task the mean RT and error rate were 678 ($\pm$71.0) ms and 1 ($\pm$3.6)%, respectively. Thus, as in the PET experiment, the participants were reliably faster in the letter task relative to the semantic task ($F_2(1, 286) = 120.3$, $P < 0.0001$) and they made few errors. An analysis of the two semantic conditions revealed no reliable differences between domains (796 $\pm$ 165.4 vs. 811 $\pm$ 162.7 ms, $F_2(1, 190) = 0.9$, ns) and the error rates for the two conditions were essentially identical (8 $\pm$ 13.9 vs. 8 $\pm$ 14.8%). These results indicate that the subjects were performing the task adequately and comparably to the subjects in Experiment 2.

4.2.2. Imaging data.

4.2.2.1. Common semantic activations. The analysis of common semantic activations was computed as in the previous experiment. Namely, a main effect of semantic categorisation relative to letter categorisation (i.e. [N]atural [K]inds + [Art]ifacts − [B]aseline) was calculated and masked with the individual contrasts (NK-B, Art-B). The results are displayed in Fig. 1d and shown in Table 10. After correcting for multiple comparisons, there were reliable activations bilaterally in the inferior frontal gyri, the medial surface of the superior frontal gyri, the temporal-parietal-occipital (TOP) junction, and the posterior-medial aspects of the cerebellum. In addition, there was reliable left hemi-

Table 11
Experiment 3: The results of the domain specific analyses[a]

Description	x	y	z	SPM$\{t\}$	Extent	n/N
Activations at uncorrected level of significance						
Natural kinds relative to man-made items						
L. precuneus	−2	−58	32	4.11	56	1/13
L. lingual gyrus	−12	−102	−10	3.24	3	1/13
Man-made items relative to natural kinds						
L. hippocampus	−28	−20	−14	3.59	14	3/13
R. anterior cingulate	16	18	28	3.37	12	1/13

[a] The table shows areas of activation for natural kinds relative to artifacts and vice-versa. All activations were present at the uncorrected $P<0.001$ level and none reached statistical significance ($T>4.5$, extent ≥ 70). Activations are reported as in the previous table.

sphere activation in the posterior superior temporal sulcus, the middle frontal gyrus, the superior frontal gyrus, the precuneus, and the thalamus. There were no other activations at a reliable level in the right hemisphere.

The areas that were absent in the fMRI relative to the PET maps were the temporal poles and the lateral posterior and inferior aspects of the temporal lobes. These two areas, however, have been shown to suffer from susceptibility-induced BOLD signal loss caused by sharp changes in local field gradients adjacent to air-tissue interfaces at the sinuses [40]. In a series of analyses of our fMRI data, we found a marked reduction in BOLD signal in the regions of the temporal lobes that were active in the PET experiments but not in the fMRI experiment (see [18] for details). Thus these temporal lobe activation differences were most likely due to the current technical limitations of fMRI, and not to task differences, which were kept to a minimum. In an attempt to enhance the sensitivity of the analysis in these areas, we used an a priori defined ROI based on the results of our second PET experiment (see [18], for more details)[7]. Using [82] small volume correction calculation, we determined the corrected t-threshold for this volume to be $t>3.15$. We then masked the SPM$\{t\}$ map from the fMRI data with the ROI and looked for voxels with SPM$\{t\}$ values greater than 3.15. The result was a single region of reliable activation in the left anterior fusiform of 7 voxels with a peak at $(-30\ \ -34\ \ -26)$ and an SPM$\{t\}$ value (at the peak) of 4.12. There was no significant activation in the left anterio-medial temporal cortex. Thus, using the small volume correction was sufficient to overcome some, but not all, of the signal reduction in the temporal lobe.

4.2.2.2. Domain-specific activations. Domain differences were identified by contrasting the natural kinds and artifacts conditions and masking the contrast with natural kinds relative to baseline (or vice versa). The results are shown in Table 11. As in Experiments 1 and 2, there were no reliable differences between natural kinds and artifacts at a corrected statistical level; the only differences were at an uncorrected threshold ($P<0.001$). Relative to artifacts, natural kinds induced small regions of activation in the left precuneus and the left lingual gyrus. The Artifacts—Natural Kinds contrast demonstrated activity in the left hippocampus and the right anterior cingulate. None of these findings were consistent with either the existing literature nor the previous two experiments. Although other studies have reported left lingual gyrus activation for natural kinds relative to artifacts [44,56], these were more anterior and superior than the activation we observed. Interestingly, the left pre-cuneus (activated by natural kinds in this experiment) was observed to be more active for *artifacts* in [57]. As both of these areas are visual processing regions, this inconsistency may reflect specific differences in the visual characteristics of the stimuli rather than semantic differences between concepts.

In the previous experiment there was an activation for artifacts in the left posterior middle temporal gyrus which was consistent with many other studies in the literature. The activation, however, consisted of only one voxel at an uncorrected significance level of $P<0.001$. This may have been a result of the artifact condition containing both tools and other man-made objects given that this region is most often observed for tools. To test this, we re-examined our fMRI data with an event-related analysis to specifically look for effects only from tools. Each stimulus was defined as an event and classified into one of four conditions: tools, other artifacts, natural kinds, or baseline. Tools were defined as manipulable objects (e.g. hammer, rope, spade, fork: [44]) To further enhance the sensitivity of the analysis, we defined an anatomical region-of-interest based on those studies that have

[7] Note that the results presented here differ slightly from those reported in [18]. This is due to the fact that here we reported a standard fixed-effect fMRI analysis whereas the previous study used a simplified analysis to make the comparison of the PET and fMRI data more similar.

previously reported activity for tools in the left posterior middle temporal gyrus (see Table 1). To be precise, we calculated the mean co-ordinates for the peak tool-activity ($x = -50$, $y = -56$, $z = -2$) and defined a sphere with a diameter of 10 mm around this peak as our ROI. Interestingly we did observe activation for tools relative to natural kinds and other man-made objects within the ROI ($x = -50, y = -56, z = 8$, SPM$\{t\} = 2.53$) but this did not reach a corrected significance level (p(corrected) $= 0.14$). Although this finding is consistent with both the previous experiment and a majority of the literature, it is worth noting that this is a very small effect and was not reliable even within a highly restricted search space.

Finally, we assessed the statistical power of this experiment as described earlier. First, for the semantics versus baseline contrast there was an estimated mixing ratio (λ) of 0.092 and an effect size (δ) of 4.8. Thus the mixing ratio and the effect size were larger than in Experiments 1 and 2. Based on the observed mean effect size and an uncorrected $P < 0.001$ α-threshold, the power $(1 - \beta)$ was calculated as 97%. The same analysis was repeated for the contrast between natural kinds and artifacts and once again the algorithm failed to find a mixture of Gaussians that fit the histogram suggesting no difference between the two domains.

4.2.3. Summary of results

The three experiments produced consistent results in that each identified a network of brain regions commonly activated in tasks requiring semantic processing. In the two PET experiments this included regions in the left inferior frontal lobe, left posterior temporal cortex, and the anterior temporal poles bilaterally. In the fMRI experiment, the same inferior frontal and posterior temporal regions were active whereas the anterior temporal poles were missing due to a loss of BOLD signal. Additional areas identified by the fMRI experiment included bilateral activations in the medial superior frontal gyri, in the TOP junction, and in the medial cerebellum. There were also left hemisphere activations in the middle temporal, the middle frontal, and superior frontal gyri, the precuneus, and the thalamus. In contrast to the robust findings for semantic representation/processing, there was no consistent evidence for specialisation for natural kinds or artifacts, or for any specific category in these domains.

5. General discussion

We have presented data from three functional neuroimaging experiments investigating the neural basis for category specific semantic deficits. The data reliably identified a distributed neural system involved in semantic processing in tasks such as lexical decision and categorisation, but activation was common to both natural kinds and artifacts. We found no evidence of functional segregation at the level of semantic domain or category which was consistent over experiments. On the other hand, like previous studies we found activation in the left anterio-medial temporal pole which was specific to animals and activation in the left posterior middle temporal gyrus which was specific to tools, although only in a subset of experiments and at a lowered statistical threshold. These results are relevant to both methodological and theoretical issues.

5.1. Methodical issues

The findings of our current experiments may seem, at first, to be at odds with reports of domain and/or category specific activations in earlier studies. However, as outlined in the introduction, there are a number of reasons why it is important to interpret the earlier results with some caution. First, as the data in this paper demonstrate, adopting a statistical threshold without correcting for multiple comparisons can generate many false positives, which will differ in location across experiments. In PET studies, small effect sizes and small quantities of data may make such liberal statistical thresholds necessary to maintain an acceptable level of sensitivity [4]. The cost, however, is the intrusion of false positives, which make it difficult to determine whether domain or category-specific effects in imaging experiments are genuine. Only those effects that replicate across studies are likely to be robust findings. fMRI offers a potential solution to this problem in that it provides roughly an order of magnitude more data per subject and thus correspondingly greater sensitivity—even at a corrected statistical threshold. On the other hand, the loss of BOLD signal due to macroscopic magnetic susceptibility artefacts is prominent in areas of the temporal lobes which often contribute to category-specific semantic impairments. Specifically, most patients with a category-specific semantic impairment for natural kinds have bilateral damage to anterior and medial temporal lobe structures subsequent to herpes simplex encephalitis [29,30]. This area, however, has been shown to demonstrate profound signal loss in fMRI and thus can not be investigated in typical gradient echo EPI whole brain imaging.

Second, as also discussed in the introduction, stimuli which have not been controlled along dimensions such as familiarity, frequency, and visual complexity may also contribute to the variable findings in the literature. These so-called ''nuisance variables'' can have a large impact on category-specific effects. For instance, Funnell and Sheridan [27] showed that JBR's natural kinds deficit was greatly exaggerated when the testing stimuli were not controlled for familiarity (but note that the effect did not disappear entirely, and see Ref. [8] for further evidence

that JBR's category-specific deficit is a genuine one). Similarly, Gaffan and Heywood [28] demonstrated a spurious impairment for natural kinds *in monkeys* based solely on visual complexity. Several neuroimaging studies have also shown that visual complexity can account for category-specific differences in humans [32,34,47]. By carefully matching the stimuli in our current studies and demonstrating identical reaction time responses across categories, this work adds weight to the conclusion that category-specific brain activations may not exist when confounding stimuli variables are controlled for.

5.2. Theoretical issues

The results of this study also speak to the theoretical issue of semantic organisation at an anatomical level. Previous studies [51,75] have shown that semantic processing tasks with either pictures or written words activate a network of regions in the left inferior frontal lobe, the left anterior temporal lobe, the left middle and inferior temporal gyri, the left TOP junction, and the right cerebellum. In a similar study using verbally presented words, Binder et al. [5] demonstrated a similar network of active regions which further included activation on the medial surface of the left superior frontal gyrus. Although the functional roles of these areas have not been fully elucidated as yet, these regions are consistently implicated in semantic processing. Each of the experiments presented here activated a subset of these regions (see Table 12). The differences between experiments were presumably due to both tasks differences between lexical decision and semantic categorisation and imaging modality differences between PET and fMRI.

In addition, each experiment identified differences between natural kinds and man-made items, although these differences were neither robust across studies nor statistically reliable. Even so, two findings were consistent with the existing category specificity literature. Mummery et al. [50] reported anterior temporal lobe activation when subjects generated examples of natural kinds relative to man-made objects and Moore and Price [47] replicated this finding in picture naming and word-picture matching tasks—although not when coloured pictures were used. The same region was present in our lexical decision task for animals relative to all other categories, although not in the semantic categorisation task. Similarly, the left posterior middle temporal region activated by tools relative to natural kinds has been shown in many imaging studies [9,12,16,44,47,50,51,57], including the two categorisation experiments reported here (although only at an uncorrected level of significance). These findings, then, could be interpreted as evidence—albeit weak—for a double dissociation of category. Certainly they are consistent with the fact that most patients with category specific impairments for natural kinds have lesions to their anterior temporal lobes subsequent to herpes simplex encephalitis [29] although there are exceptions (e.g. [38]).

There are, however, a number of problems with this account. Patients with the opposite impairment, namely a preferential impairment for man-made items, typically have left fronto-parietal lesions which do not include the posterior middle temporal region. Thus these patients are difficult to interpret in terms of the activation patterns observed in this study. In addition, it is not clear why these results were only present in a subset of experiments. Why were tools only weakly activated in our studies and why did they not activate the posterior middle temporal cortex in the lexical decision experiment? And why did the categorisation experiments—which presumably placed greater semantic demands than lexical decision—fail to find activation in the anterior temporal poles for natural kinds? These difficulties, coupled with the fact that the imaging evidence they are based on is neither consistent nor statistically reliable, lead us to favour an alternate explanation. Instead, we suggest these data may be most consistent with the claim that semantic memory is a unified but anatomically distributed system, in which there is no neural specificity as a function of category/domain of concept or type of semantic property [19,71].

5.3. Category-specific deficits revisited

If the brain is not organised according to semantic domain or feature type, then what is the genesis of category-specific semantic deficits? Such deficits have been reliably documented for many cases, even when potentially confounding variables such as familiarity and visual complexity are ruled out. As discussed in the introduction, one class of cognitive account in the

Table 12
A summary of the common semantic regions identified in the three experiments

Region	Lexical decision (PET)	Semantic categorisation (PET)	Semantic categorisation (fMRI)
L. inferior frontal cortex	×	×	×
L. medial superior frontal gyrus		×	×
Anterior temporal poles	×	×	
L. middle/inferior temporal gyri	×	×	×
R. cerebellum		×	×

literature explicitly claims that the existence of category-specific behavioural deficits does *not* necessarily arise from category organisation at the neural level. The "conceptual structure account" [19,71] claims that concepts are represented as patterns of activation over many "nodes" in a distributed connectionist learning system. Similar concepts are represented by overlapping patterns of activation, but there is no a priori distinction at either the functional or neural level, between different semantic categories or domains. There are two main ways in which category/domain effects can arise as the result of widespread damage to this kind of system. First, although the patterns of activation for concepts are highly distributed, the overlap among similar concepts may lead to clusters in semantic space, such that a group of related concepts may all suffer disproportionately if damage happens to affect a certain combination of nodes or connections among them (cf. [10]). Second, we have argued that the structure of concepts differs systematically across categories and domains, and that this leads to different outcomes when the system is damaged. For example, natural kinds have many properties that are true of all or most members of the category, and these occur frequently together (e.g. ⟨*has legs*⟩, ⟨*has eyes*⟩, ⟨*can walk*⟩, ⟨*can see*⟩). The strong correlations among these features mean that they can support each other with mutual activation and are relatively robust to loss of individual nodes or connections. The *distinctive* properties of natural kinds, on the other hand, are more weakly correlated and so more vulnerable to damage. This leads to the frequently observed pattern of category-specific deficit for natural kinds in which patients know the broad category to which an item belongs (e.g. *it is an animal*) and can generate shared information (*it's got legs*, *it breathes*) but are unable to identify the distinctive properties that identify it as a specific animal within the category (does it have spots or stripes, a mane or a trunk?: [49]). Man-made objects—and this applies most clearly to the category of *tools*—have fewer shared properties, but tend to have strong form-function correlations among their more distinctive properties (⟨*used for cutting*⟩-⟨*has a sharp edge*⟩/⟨*used for raking*⟩–⟨*has long tines*⟩) that make them relatively robust. This predicts a pattern in which knowledge of specific artifacts is well-preserved, until the overall damage to the semantic system is very severe, at which point all that remains intact are the sets of shared intercorrelated properties of natural kinds. Such patients will have very low scores on all semantic tasks, but have some preserved knowledge of natural kinds, hence revealing the reverse pattern of a category-specific deficit for artifacts [48]. It is worth noting at this point, that the conceptual structure account predicts that differences between domains and categories of knowledge will be relative rather than all-or-none, and this is consistent with the vast majority of patients reported in the literature. To recap, the conceptual structure account claims that conceptual knowledge is represented in a distributed system. The overlap of similar concepts and the different correlational structure of information within domains of knowledge interacts with damage to the system in such a way that category-specific deficits can emerge. The exact pattern of performance for patients will vary as a function of the kinds of semantic property tested (distinctive or shared, correlated or not correlated) as well as the task (does the task require identification of specific items or is shared category information sufficient).

The conceptual structure account is a theory about the functional organisation of the conceptual system. It is concerned with the microstructure of concepts and the relations among them. Nevertheless, it does have testable implications for the neural substrate of conceptual knowledge; i.e. that there will be no segregation of different brain areas according to category of concept or type of semantic feature. This prediction clearly differentiates the conceptual structure account from other current theories of category-specific deficits which imply regional specialisation as a function of either category of concept [11] or type of semantic property [79]. The conceptual structure account implies that the representational substrate for *all* domains of conceptual knowledge is distributed in a single semantic system (although in this paper we have only tested living and non-living objects, the account applies to other domains, such as abstract nouns, verbs and adjectives—[70]. The three experiments reported here support this view, reliably demonstrating common areas of widespread temporal activation for both natural and man-made items in two different tasks.

In conclusion, the data presented here suggest that that conceptual knowledge is represented in a unitary, distributed system and argue that a distributed neural system undifferentiated by categories of knowledge underlies semantic processing. This claim is compatible with the existing neuroimaging literature which has found little evidence of consistent specialisation for either natural kinds or artifacts. Clearly future neuroimaging studies will play an important role in resolving these questions but for this to happen certain methodological issues need to be taken into account. These include controlling the stimuli for so-called "nuisance variables" and designing experiments to maximise sensitivity while minimising false positives. In PET studies, this may involve scanning many more subjects than is typically done, while in fMRI it will most likely involve new data acquisition techniques which increase the signal from regions of macroscopic susceptibility.

Acknowledgements

We thank the members of the WBIC in Cambridge and the FMRIB Centre in Oxford for their assistance in collecting these data. In addition, we thank our volunteers for their participation. JTD and RPR were supported by an MRC programme grant to LK Tyler and WD Marslen-Wilson. MJF and MHD were supported by a McDonnell-Pew grant to LKT. HEM was supported by a Wellcome Trust Fellowship. The Oxford FMRIB Centre is supported by an MRC programme grant. R. Russell is at Department of Cognition and Brain Sciences, Massachusetts Institute of Technology; M. Davis is at Medical Research Council Cognition and Brain Sciences Unit, Cambridge, and J. Fadili is at GREYC laboratory, ISMRA-ENSI, Caen.

References

[1] Ashburner J, Friston K. Multimodal image coregistration and partitioning—a unified framework. NeuroImage 1997;6:209–17.
[2] Ashburner J, Neelin P, Collins DL, Evans AC, Friston KJ. Incorporating prior knowledge into image registration. NeuroImage 1997;6:344–52.
[3] Baayen RH, Pipenbrook R. The Celex Lexical Database. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania, 1995.
[4] Bailey DL, Jones T, Spinks TJ, Gilardi MC, Townsend DW. Noise equivalent count measurements in a neuro-PET scanner with retractable septa. IEEE Transactions in Medical Imaging 1991;10:256–60.
[5] Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. Human brain language areas identified by functional magnetic resonance imaging. Journal of Neuroscience 1997;17(1):353–62.
[6] Buckner R, Koustall W, Schacter D, Rosen B. Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. Brain 2000;123:620–40.
[7] Bunn EM, Tyler LK, Moss HE. Patient JBR: The role of familiarity and property type in a selective deficit for living things. Brain and Language 1997;60(1):10–2.
[8] Bunn EM, Tyler LK, Moss HE. Category-specific semantic deficits: The role of familiarity and property type re-examined. Neuropsychology 1998;12(3):367–79.
[9] Cappa SF, Perani D, Schnur T, Tettamanti M, Fazio F. The effects of semantic category and knowledge type on lexical-semantic access: A PET study. NeuroImage 1998;8(4):350–9.
[10] Caramazza A, Hillis A, Rapp B, Romani C. The multiple semantic hypothesis: Multiple confusions? Cognitive Neuropsychology 1990;7:161–89.
[11] Caramazza A, Shelton J. Domain-specific knowledge systems in the brain: The animate-inanimate distinction. Journal of Cognitive Neuroscience 1998;10(1):1–34.
[12] Chao L, Haxby JV, Martin A. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nature Neuroscience 1999;2(10):913–9.
[13] Clark VP, Keil K, Maisog JM, Courtney S, Ungerleider LG, Haxby JV. Functional magnetic resonance imaging of human visual cortex during face matching: A comparison with positron emission tomography. NeuroImage 1996;4:1–15.
[14] Coltheart M. The MRC Psycholinguistics database. Quarterly Journal of Experimental Psychology 1981;33A:497–505.
[15] Coull JT, Nobre AC. 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. Journal of Neuroscience 1998;18(18):7426–35.
[16] Damasio H, Grabowski T, Tranel D, Hichwa R, Damasio A. A neural basis for lexical retrieval. Nature 1996;380:499–505.
[17] Devlin J, Gonnerman L, Andersen E, Seidenberg M. Category specific semantic deficits in focal and widespread brain damage-A computational account. Journal of Cognitive Neuroscience 1998;10(1):77–94.
[18] Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Matthews PM, Moss HE, Tyler LK. Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. NeuroImage 2000;11(6):589–600.
[19] Durrant-Peatfield M, Tyler L, Moss H, Levy J. The distinctiveness of form and function in category structure: A connectionist model. In: Proceedings of the 19th Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates, 1997:193–8.
[20] Forde EM, Francis D, Riddoch MJ, Rumiati RI, Humphreys GW. On the links between visual knowledge and naming: A single case study of a patient with category-specific impairment for living things. Cognitive Neuropsychology 1997;14:403–58.
[21] Forde EME, Humphreys GW. Category-specific recognition impairments: a review of important case studies and influential theories. Aphasiology 1999;13(3):169–93.
[22] Forster KI, Forster JC. The DMASTR display system for mental chronometry. Tucson, Arizona: University of Arizona, 1990.
[23] Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images. Human Brain Mapping 1995;2:165–89.
[24] Friston KJ, Holmes AP, Worsley KJ, Poline J-B, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 1995;2:189–210.
[25] Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. Human Brain Mapping 1994;1:214–20.
[26] Funnell E, de Mornay-Davies P. JBR: A reassessment of concept familiarity and a category-specific disorder for living things. Neurocase 1996;2:461–74.
[27] Funnell E, Sheridan J. Categories of knowledge? Unfamiliar aspects of living and nonliving things. Cognitive Neuropsychology 1992;9(2):135–53.
[28] Gaffan D, Heywood C. A spurious category-specific visual agnosia for living things in normal human and nonhuman primates. Journal of Cognitive Neuroscience 1993;5:118–28.
[29] Gainotti G. What the locus of brain lesion tells us about the nature of the cognitive deficit underlying category-specific disorders: a review. Cerebral Cortex 2000;36:539–59.
[30] Gainotti G, Silveri MS, Daniele A, Giustolisi L. Neuroanatomic correlates of category specific semantic disorders: A critical survey. Memory 1995;3:247–64.
[31] Genovese CR, Noll DC, Eddy WF. Estimating test-retest reliability in functional magnetic resonance imaging I: Statistical methodology. Magnetic Resonance in Medicine 1997;38:497–507.
[32] Gerlach C, Law I, Gade A, Paulson OB. Perceptual differentiation and category effects in normal object recognition: A PET study. Brain 1999;122:2159–70.
[33] Gonnerman L, Andersen E, Devlin J, Seidenberg M, Kempler D. Double dissociation of semantic categories in Alzheimer's disease. Brain and Language 1997:254–79.
[34] Gorno-tempini ML, Cipolotti L, Price CJ. Which level of object processing generates category specific differences in brain activation? Proceedings of the Royal Society, London B 2000:1253–8.
[35] Grabowski TJ, Damasio H, Damasio AR. Premotor and prefrontal correlates of category-related lexical retrieval. NeuroImage 1998;7:232–43.

[36] Hart J, Berndt R, Caramazza A. Category specific naming deficit following cerebral infarction. Nature 1985;316:439–40.
[37] Herbster AN, Mintun MA, Nebes RD, Becker JT. Regional cerebral blood flow during word and nonword reading. Human Brain Mapping 1997;5(2):84–92.
[38] Hillis A, Caramazza A. Category specific naming and comprehension impairment: A double dissociation. Brain 1991;114:2081–94.
[39] Humphreys GW, Riddoch MJ, Quinlan PT. Cascade processes in picture identification. Cognitive Neuropsychology 1988;5:67–103.
[40] Jezzard P, Clare S. Sources of distortion in functional MRI data. Human Brain Mapping 1999;8(2–3):80–5.
[41] Kinahan PE, Noll DC. A direct comparison between whole-brain PET and BOLD fMRI measurements of single-subject activation response. NeuroImage 1999;9:430–8.
[42] Kinahan PE, Rogers JG. Analytic 3D image reconstruction using all detected events. IEEE Transaction on Nuclear Science 1989;35(1):680–4.
[43] Lambon Ralph MA, Howard D, Nightingale G, Ellis AW. Are living and non-living category specific deficits causally linked to impaired perceptual or associative knowledge? Evidence from a category-specific double dissociation. Neurocase 1998;4:311–38.
[44] Martin A, Wiggs C, Ungerleider L, Haxby J. Neural correlates of category-specific knowledge. Nature 1996;379:649–52.
[45] McRae K, de Sa V, Seidenberg M. On the nature and scope of featural representations of word meaning. Journal of Experimental Psychology: General 1997;126(2):99–130.
[46] Metter EJ, Kempler D, Jackson CA, Hanson WR, Riege WH, Camras LR, Mazziotta JC, Phelps ME. Cerebellar glucose metabolism in chronic aphasia. Neurology 1987;37:1599–606.
[47] Moore CJ, Price CJ. A functional neuroimaging study of the variables that generate category specific object processing differences. Brain 1999;122:943–62.
[48] Moss HE, Tyler LK. A progressive category-specific semantic deficit for non-living things. Neuropsychologia 2000;38(1):60–82.
[49] Moss HE, Tyler LK, Durrant-Peatfield M, Bunn EM. Two eyes of a see-through': Impaired and intact semantic knowledge in a case of selective deficit for living things. Neurocase 1998;4(4–5):291–310.
[50] Mummery CJ, Patterson K, Hodges J, Wise RJ. Generating 'tiger' as an animal name or a word beginning with T: Differences in brain activation. Proceedings of the Royal Society of London B Biological Sciences 1996;263:989–95.
[51] Mummery CJ, Patterson K, Hodges JR, Price CJ. Functional neuroanatomy of the semantic system: Divisible by what? Journal of Cognitive Neuroscience 1998;10(6):766–77.
[52] Mummery CJ, Shallice T, Price CJ. Dual-process model in semantic priming: A functional imaging perspective. NeuroImage 1999;9(5):516–25.
[53] Nickels L, Howard D. Aphasic naming: What matters? Neuropsychologia 1995;33:1281–303.
[54] Owen AM, Herrod NJ, Menon DK, Clark JC, Downey SP, Carpenter TA, Minhas PS, Turkheimer FE, Williams EJ. Redefining the functional organisation of working memory process within human lateral prefrontal cortex. European Journal of Neuroscience 1999;11:567–74.
[55] Parkin AJ, Stewart F. Category specific impairments? No—A critique. Quarterly Journal of Experimental Psychology 1993;46(3):505–9.
[56] Perani D, Cappa S, Bettinardi V, Bressi S, Gorno-tempini M-L, Matarrese M, Fazio F. Different neural systems for the recognition of animals and man-made tools. Neuroreport 1995;6:1637–41.
[57] Perani D, Schnur T, Tettamanti M, Gorno-Tempini M, Cappa SF, Fazio F. Word and picture matching: a PET study of semantic category effects. Neuropsychologia 1999;37:293–306.
[58] Price CJ, Friston KJ. Cognitive conjunction: A new approach to brain activation experiments. NeuroImage 1997;5(4 Pt1):261–70.
[59] Price C.J., Friston K.J. (in press), What has neuroimaging contributed to category-specificity? In: Humphreys G., Forde E. (Eds.), Category specificity in mind and brain.
[60] Price CJ, Veltman DJ, Ashburner J, Josephs O, Friston KJ. The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. NeuroImage 1999;10:36–44.
[61] 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(6):1255–69.
[62] Price CJ, Wise RJS, Frackowiak RSJ. Demonstrating the implicit processing of visually presented words and pseudowords. Cerebral Cortex 1996;6(1):62–70.
[63] Rumsey JM, Horwitz B, Donahue BC, Nace K, Maisog JM, Andreason P. Phonological and orthographic components of word recognition: a PET-rCBF study. Brain 1997;120(5):739–59.
[64] Saffran E, Schwartz M. Of cabbages and things: Semantic memory from a neuropsychological perspective-A tutorial review. In: Meyer D, Cornblum S, editors. International symposium on attention and performance XV. Cambridge, Mass.: MIT Press, 1994.
[65] Shelton JR, Caramazza A. Deficits in lexical and semantic processing: Implications for models of normal language. Psychological Bulletin Review 1999;6(1):5–27.
[66] Silbersweig DA, Stern E, Frith CD, Cahill C, Schnorr L, Grootoonk S, Spinks T, Clark J, Frackowiak R, Jones T. 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. Journal of Cerebral Blood Flow and Metabolism 1993;13:617–29.
[67] Spitzer M, Kischka U, Guckel F, Bellemann M, Kammer T, Seyyedi S, Weisbrod M, Schwartz A, Brix G. Functional magnetic resonance imaging of category-specific cortical activation: evidence for semantic maps. Cognitive Brain Research 1998;6:309–19.
[68] Stewart F, Parkin AJ, Hunkin NM. Naming impairments following recovery from herpes simplex encephalitis: Category specific? Quarterly Journal of Experimental Psychology 1992;44A:261–84.
[69] Thompson-Schill SL, Aguirre GK, D'Esposito M, Farah MJ. A neural basis for category and modality specificity of semantic knowledge. Neuropsychologia 1999;37:671–6.
[70] Tyler L.K., Moss H.E. (In press), Towards a distributed account of conceptual knowledge, Trends in Cognitive Science.
[71] Tyler LK, Moss HE, Durrant-Peatfield MR, Levy JP. Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. Brain and Language 2000;75(2):195–231.
[72] Tyler L.K., Russell R.P., Fadili J., Moss H.E. (submitted), The neural representations of nouns and verbs: PET studies.
[73] Ulrich R, Miller J. Effects of truncation on reaction time analyses. Journal of Experimental Psychology: General 1994;123(1):34–80.
[74] Van Horn JD, Ellmore TM, Esposito G, Berman KF. Mapping voxel-based statistical power on parametric images. NeuroImage 1998;7:97–107.
[75] Vandenberghe R, Price CJ, Wise R, Josephs O, Frackowiak RSJ. Functional anatomy of a common semantic system for words and pictures. Nature 1996;383:254–6.
[76] Veltman R, Friston KJ, Sanders G, Price CJ. Regionally specific sensitivity differences in fMRI and PET: Where do they come from? NeuroImage 2000;11(6):575–88.
[77] Warrington E, McCarthy R. Category specific access dysphasia. Brain 1983;106:859–78.

[78] Warrington E, McCarthy R. Categories of knowledge: Further fractionations and an attempted integration. Brain 1987;110:1273–96.
[79] Warrington E, Shallice T. Category specific semantic impairments. Brain 1984;107:829–53.
[80] Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited—Again. NeuroImage 1995;2:173–81.
[81] Worsley KJ, Marrett S, Neelin P, Evans AC. A three-dimensional statistical analysis for CBF activation studies in human brain. Journal of Cerebral Blood Flow and Metabolism 1992;12:900–18.
[82] Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping 1996;4:58–83.