doi:10.1093/brain/awm025 *Brain* (2007), **130**, 1127–1137

Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model

Matthew A. Lambon Ralph,[1] Christine Lowe[1] and Timothy T. Rogers[2]

[1]School of Psychological Sciences, University of Manchester, Manchester, UK and [2]Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

Correspondence to: Prof. M. A. Lambon Ralph, School of Psychological Sciences, University of Manchester, Oxford Road, Manchester, MI3 9PL, UK and Dr T. T. Rogers, Department of Psychology, University of Wisconsin-Madison, Madison, Wisconsin, USA
E-mail: matt.lambon-ralph@manchester.ac.uk

Studies of patients with semantic impairments following brain damage offer key insights into the cognitive and neural organization of semantic memory. Especially important in this regard are studies of category-specific semantic impairment. We report a direct comparison of semantic deficits in two groups suffering from different diseases: semantic dementia (SD) and herpes simplex virus encephalitis (HSVE). Although pathology in both disorders is centred on the anterior temporal lobes bilaterally, category-specific semantic impairment is rarely observed in SD yet commonly found in HSVE. Using a combination of neuropsychology and computational neuroscience, we tested the possibility that category-specific deficits for living things depend not solely upon the location of damage within the cortical semantic network but also critically upon the type of impairment. When the semantic representations within the model are degraded or 'dimmed' then a generalized, global semantic impairment results (as found in SD) but when the representations are distorted then a category-specific pattern emerges (as per HSVE). Three novel predictions from this model were tested and confirmed, thereby adding weight to the hypothesis that both type and distribution of pathology can be critical in producing neuropsychological phenomena.

Keywords: semantic memory disorders; semantic dementia; herpes simplex virus encephalitis; PDP neural network models; category-specific disorders

Abbreviations: CVA = cerebral vascular accident; HSVE = herpes simplex virus encephalitis; PDP = parallel distributed processing; SD = semantic dementia

Received September 26, 2006. Revised January 28, 2007. Accepted January 29, 2007

Introduction

The study of semantically impaired patients has yielded important insights into the neural processes and representations underlying conceptual knowledge. One of the most striking and theoretically important types of neuropsychological deficit is category-specific semantic impairment, in which patients present with differential loss of knowledge for one semantic category over another. For instance, Warrington and McCarthy (1983) described a patient who had relatively better knowledge of animate categories than inanimate objects; and the reverse pattern was documented by Warrington and Shallice (1984) in four HSVE patients. These clinical observations have since led to a large number of case reports documenting similar dissociations between these domains (at least 125 cases since 1984: Capitani *et al.*, 2003). Such findings are provocative in part because they provide insight into the organizational principles underpinning the semantic system. Category-specific impairments may indicate, for instance, that knowledge about living and non-living things is supported by independent and anatomically segregated modules formed through phylogenesis (Caramazza and Shelton, 1998) or rely differentially upon sensory and motor properties encoded in separate cortical regions (Warrington and Shallice, 1984; Farah and McClelland, 1991). Despite the quantity of data and a lively debate about the merit of these and other hypotheses, a consensus on the theoretical explanations of category-specific semantic impairment remains elusive.

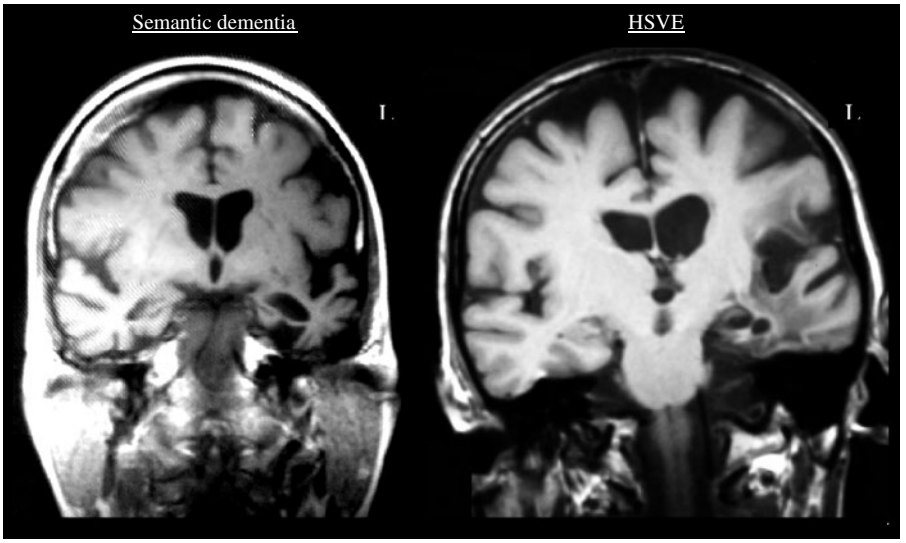

Fig. 1 Example coronal MRI for SD versus HSVE. These two example coronal slices show the typical nature and type of disease-specific damage shown on structural MRI. It should be noted, however, that the distribution of damage for both HSVE and SD varies across patients from bilateral damage through to strongly right or left lateralized (as shown here).

In this study, we adopted a novel approach that combines detailed neuropsychological investigation with computational neuroscience. Previously, we have found that potentially important theoretical insights can be gleaned through the direct, case-series comparison of two relevant but contrasting patient types (e.g. impairments of semantic cognition in semantic dementia versus aphasia: Jefferies and Lambon Ralph, 2006). In this study, therefore, we directly compared the semantic breakdown observed in semantic dementia (SD) and HSVE using the same battery of measures; and we sought to understand the observed differences between the groups using simulations with an extension of the computational model of semantic memory described by Rogers *et al.* (2004*a*, *b*).

The comparative neuropsychological investigation was based on two contrasting yet complementary neurological conditions. SD is the temporal variant of frontotemporal dementia in which anterior temporal lobe atrophy (Fig. 1) is associated with a very selective yet progressive degradation of conceptual knowledge (Hodges *et al.*, 1992). Despite their substantial semantic impairment, it is rare for SD patients to present with a category-specific pattern (for example, combining across studies, performance on different categories has been investigated, with all appropriate controls, in 30 SD patients and only one showed a living < manmade pattern: Lambon Ralph *et al.*, 2003), so the group provides an important neuropsychological and neurological baseline against which to compare patients with category-specific semantic impairment. Although category-specific deficits have been reported in association with a wide variety of cortical brain diseases (Lambon Ralph *et al.*, 1998), they are most commonly observed in semantically impaired HSVE patients. Nearly half of all reported living < non-living category-specific cases arise from HSVE (Capitani *et al.*, 2003). HSVE is a heterogeneous clinical disorder in which variable distributions and degrees of focal necrosis are associated with varying neuropsychological deficits. Semantic impairment is found in patients with more extensive temporal lobe damage affecting lateral as well as medial cortex (Kapur *et al.*, 1994; Noppeney *et al.*, 2007). A few studies have found moderate-to-high rates of category-specific semantic impairment in this group (Barbarotto *et al.*, 1996), which is certainly higher than the incidence in SD. The ideal design goal—contrasting two semantically impaired patient groups with and without a category-specific presentation—was met, therefore, by comparing seven HSVE patients with seven SD patients matched for the severity of their semantic deficits.

To understand the different patterns of impairment observed in the two groups, we used an implemented computational model of conceptual knowledge and its breakdown following brain damage (Fig. 2 and Rogers *et al.*, 2004*a*, *b*). The key proposal exemplified by the model is that modality-specific perceptual, linguistic and motor representations distributed throughout the cortex communicate with one another by means of representations encoded in the anterior temporal lobes, which act as a kind of cross-modal 'hub'. This proposal builds upon the notion that concepts are formed through the distillation of verbal and non-verbal information encoded within unimodal association areas (Eggert, 1977; Martin and Chao, 2001) and echoes previous suggestions that 'convergence zones' within temporal cortex provide an amodal indexing mechanism (Damasio *et al.*, 1996). Our model extends these ideas by addressing important neuroanatomical and computational considerations.

With regard to neuroanatomy, the model's architecture is motivated by the observation that the heteromodal cortex

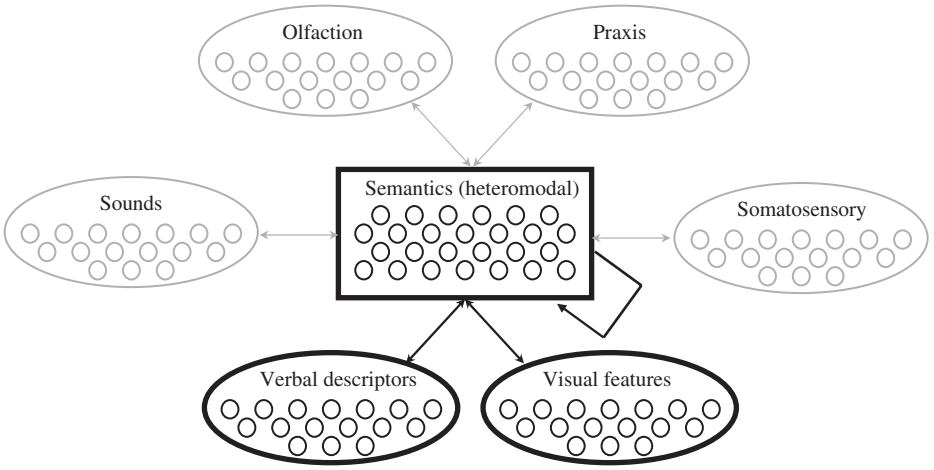

Fig. 2 PDP computational framework of conceptualization. Abstract conceptual representations are formed through the interactive activation of modality-specific, unimodal representations across a set of heteromodal units. The implemented elements of this full framework are shown in bold (see Rogers *et al*., 2004*a*, *b* for details). SD is simulated by removing a proportion of the connections to/from the semantic units while HSE is simulated by adding noise to the weight values on these same connections.

of the anterior temporal cortex forms the apex of both ventral visual and superior-temporal auditory processing streams (Wise, 2003), and is reciprocally interconnected with all secondary perceptual and motor cortices (Gloor, 1997). Although current neuroimaging and patient studies have not resolved exactly which subregions within this area are critical for forming semantic representations (e.g. rostral superior temporal sulcus and/or anterior perirhinal regions), these areas are ideally situated to compute cross-modal mappings due to their patterns of connectivity.

Computationally, the representations formed across the connections to and from the semantic (anterior temporal lobe) units function as more than a set of index cards. Instead, they capture the semantic similarity structure that exists amongst various concepts, providing the only basis on which (i) accurate transformations between surface representations can be generated and (ii) appropriate generalizations to new and old concepts can be computed (Rogers and McClelland, 2004; Rogers *et al*., 2004*a*, *b*). Importantly, the intermediating representations are understood to be amodal (individual units are not associated with 'visual', 'propositional' or 'functional' features, for example, but connect reciprocally with all unimodal regions) and homogeneous (all units participate in representing all concepts irrespective of semantic domain). In this framework, category-specific differences arise from graded factors that vary intrinsically across the domains.

The current simulation work was motivated by the observation that, as evident in Fig. 1, the semantic impairment in both SD and HSVE is associated with damage to the anterior temporal lobes, suggesting that this location of damage is critical in producing semantic deficits *per se*. A recent formal comparison of HSVE and SD using voxel-based morphometry found that the pathology in both diseases is concentrated upon very similar, overlapping bilateral anterior temporal structures, albeit with greater medial involvement in HSVE and somewhat more lateral involvement in SD (Noppeney *et al*., 2007). The authors of this study note that it is somewhat puzzling to find two different types of semantic impairment associated with highly overlapping regions of damage. Given the high degree of overlap in the affected regions found in SD and HSVE, we were motivated to investigate a novel hypothesis: might the different behavioural profiles be attributable to different forms of damage to largely the same neuroanatomical locus?

We addressed this question using a variant of a computational model of semantic memory that had previously proved successful at explaining several aspects of semantic impairment in SD (Rogers *et al*., 2004*a*, *b*). In that work, we simulated brain damage in SD by removing an increasing proportion of connection weights projecting into or out from the model's 'anterior temporal lobe'. Here we began by considering whether there were alternative ways of disrupting processing in the same region of the model that would yield a pattern of impaired behaviour similar to that observed in HSVE (see later). Having established the face validity of the proposal with these simulations, we then tested and confirmed three novel predictions from the model in SD and HSVE patients. In what follows, we first present the case-series contrast of patient groups; we then discuss the modelling effort to understand these differences *post hoc*; and finally show the model predictions and corresponding patient data. In the discussion, we consider what the modelling work suggests about the nature of the functional and neurophysiological dysfunction in the two diseases.

Material and methods

Participants

Seven of the eight SD patients were identified through the Memory and Cognitive Disorders Clinic at Addenbrooke's

Hospital, Cambridge, UK, where they were seen by a senior neurologist, a senior psychiatrist and a clinical neuropsychologist. An additional patient was recruited through the Research Institute for the Care of the Elderly (RICE) in Bath, UK. In addition to clinical assessment, all patients were given a number of standard psychiatric rating scales to exclude major functional disorders such as depression and schizophrenia, and underwent MRI scanning. All patients fulfilled the criteria for SD (Hodges *et al.*, 1992). The patients were selected such that their mild to moderate levels of semantic impairment and anomia were commensurate with the levels of semantic impairment in the HSVE patients (see later for details). Background neuropsychological tests confirmed the selective nature of semantic impairment that is a cardinal feature of this disorder. The patients had excellent non-verbal problem-solving skills (as measured on the Raven's Coloured Progressive Matrices: Raven, 1962), visual and spatial perception (VOSP: Warrington and James, 1991), excellent memory for current events though, like other SD cases, were down on tests of word and unfamiliar face recognition memory (Warrington, 1996). Their recall of complex, abstract figures (Osterrieth, 1944) was normal immediately and as within the normal range after a 45-min delay in all but two patients (CS and MA).

The seven HSVE patients were recruited from two sources. Five of the patients were identified through the Cerebral Function Unit at Hope Hospital, Salford. These patients were seen by a senior neurologist and clinical neuropsychologist. Two other patients were recruited from the Encephalitis Support Group. All patients were given a confirmed diagnosis of HSVE via virology testing. All patients were tested at least 18 months following their discharge from hospital. Background neuropsychological testing demonstrated that like the SD patients, the HSVE cases had normal perceptual and spatial functioning. Similarly, on the tests of recognition memory all but one of the patients (JD) were impaired on word and face recognition tests. On other measures of cognitive function there is a great contrast between the SD and HSVE patients. All but one of the seven HSVE patients performed poorly on the non-verbal problem-solving test and all patients performed at, or almost at, floor on the immediate and delayed conditions of the complex figure recall task. Four of the patients were also impaired in the copy phase of the test, which in light of their normal performance on the visual perception battery, suggests an executive/planning deficit rather than a low-level visuospatial impairment.

Experimental neuropsychology

(i) Short matched naming: Tests for category differences over and above the influence of familiarity and other potentially confounding factors. The assessment comprises 16 living–manmade pairs closely matched for familiarity and word frequency taken from Garrard *et al.* (2002). Control participants score at ceiling on this test.

(ii) Extended matched naming: A longer naming battery comprising 30 animal–manmade pairs that are closely matched for psycholinguistic factors including familiarity, frequency, length and imageability (Lambon Ralph *et al.*, 1998). Control participants score at ceiling on this test.

(iii) Word–picture matching: The 64-item comprehension test used extensively by the Cambridge group, in which participants listen to a spoken name and must point to the matching item in an array of 10 line drawings (Bozeat *et al.*, 2000). Control participants score between 62 and 64 on this test.

(iv) Specific (subordinate) naming: A 28-item, simple picture naming task. The stimuli are colour photographs of real objects, half animals and half manmade objects. Participants are asked, and where necessary prompted, to provide the specific name for the object (e.g. 'Rolls Royce' rather than 'car', 'kingfisher' rather than 'bird', etc.). All items were pre-selected to ensure that at least 90% of the age-matched controls could name each item, and rated familiarity was matched across categories.

Computational PDP model

The simulation was based closely on our previous model of conceptual knowledge and its breakdown in semantic dementia (Rogers *et al.*, 2004*a*, *b*). Further details about the model and its behaviour under damage can be found in our previous paper, however, key elements are noted here. The model contained three layers of units: one coding the visual similarity structure across a set of 'visual' units; another 'verbal' layer coded the structure of verbal propositions about objects as well as their individual names. The third 'semantic' layer was made up of a set of hidden units that provide the basis for transmission of information within and between the visual and verbal layers. All units employed a standard logistic transfer function with a gain of 1; and all units had an untrainable bias of −2 so that, in the absence of external input, a unit's activation state would settle at 0.12. Thus when a unit in the model failed to get external input, it did not strongly activate, and did not have a strong influence on the units to which it projected. The model was trained to associate visual representations, verbal descriptions and individual names so that, when trained, it could reproduce all the correct information for a concept when only one type of input was applied to the model (e.g. the name as input leading to the activation of visual features and verbal propositions).

The model was trained with 30 analogues of 'living things' and 30 analogues of 'manmade objects'. The 30 items in each domain were subdivided into six 'basic-level' categories consisting of five items each and the visual and verbal training patterns were constructed to capture two important aspects of similarity structure apparent in the majority of attribute-listing experiments. First, in both domains, items from the same basic-level category tended to have many visual and verbal properties in common (Tversky and Hemenway, 1984). Second, living things from different basic categories shared many more properties with one another than did manmade objects from different basic-level categories (Garrard *et al.*, 2001; McRae and Cree, 2002). Effectively, the manmade objects were organized in six tight clusters that were well-separated from one another (corresponding to basic categories such as 'car' and 'boat', which have little in common apart from shared function), whereas the animals were organized in six tight clusters which were not widely separated (corresponding to basic categories such as 'bird', 'dog' and 'fish' which, while easily discriminated, still share many properties in common). For all items, the model learned all possible mappings among individual names, verbal descriptions and visual representations. In contrast to our earlier work, the model learned subordinate-level names for some items in addition to learning the basic-level name for all items. Finally, all items were associated with a 'basic-level' name common to the five items in each basic category and with a superordinate name common to all 30 items in the same semantic domain. For half of the basic categories, three of the five items were also assigned a unique subordinate-level name. Thus the model provides a simple analogue to the real-world situation in which participants know that

there are several different breeds of dog but only know the subordinate names for some of the various breeds and further have knowledge of other items (e.g. cows) without much knowledge of their subordinates.

The model was trained with backpropagation over time for 10 000 epochs, using a learning rate of 0.005, at which point it was able to activate all visual and verbal units to within 0.2 of their target states for all patterns. The fully trained model was then subjected to two different forms of damage: (i) removal of an increasing proportion of randomly selected weights from those entering, leaving or intrinsic to the semantic layer (to simulate SD) or (ii) permanent positive or negative changes in the value of the weights for these connections—implemented by selecting the size of the weight change from a uniform distribution with a mean of zero and an increasing range to simulate increasingly severe impairment (in the HSVE simulations) (the maximum noise range investigated was −1.45 to 1.45. The trained weights had a mean of −0.176 and a standard deviation of 0.93). Each form of damage was administered 25 times at each level of severity and reported results were averaged across these runs to ensure that the interesting behaviour did not result from the chance disruption of a select set of weights.

Word–picture matching and confrontation naming were simulated as described in Rogers *et al.* (2004*a*, *b*) For each trial of word–picture matching, the model was provided with a single name as input and the resulting pattern of activation across semantic units was recorded. Subsequently the model was given visual inputs for eight different items in series, one of which matched the named item. The model was considered to have selected as a match whichever visual input produced an internal representation most similar to that generated by the word. This procedure was carried out for all 60 items in the training set. Confrontation naming was simulated by simply applying a visual input pattern and inspecting the activation of the model's output units. The model was considered to have produced the most specific name activated above a threshold of 0.5 (Levelt, 1989). If no unit exceeded this threshold, it was considered to have produced no response. Errors were classified as 'category-coordinate' if the model produced the name of a different item from the same broad domain and as 'omission/superordinate' if the model failed to activate any name other than a superordinate over threshold.

Performance on all naming tasks was assessed for levels of damage where the model's word–picture matching scores most closely matched each individual patient in the two groups. This means that the parameters governing overall semantic damage were adjusted only to fit the word–picture matching score. All other data from the model (naming on the various sets) reflect these same settings without further changes. Not only were the two patient groups matched on word–ndash;picture matching performance but the two versions of the damaged model (weight adjustments—HSVE versus weight removal—SD) were matched to the patients on this same measure, as well. Consequently, differential changes in overall performance that result from the two forms of damage were equated (removing a connection and adjusting its strength have different effects on a unit's performance—so it is preferable to use a behavioural measure to match both models and patients). In effect the models were 'severity' matched (like the patients) in terms of the behavioural outcome of the damage and thus qualitative differences in performance cannot reflect this global factor.

Results

Rate of category-specificity in HSVE and SD

Individual scores on the word–picture matching task are shown in Table 1. Performance on this measure of semantic memory was matched across the two groups: performance spanned roughly the same range and the group means did not differ [$t(12) = 0.57$, n.s.]. On the short matched naming test, a category-specific pattern favouring manmade objects was observed in six/seven HSVE patients, with the

Table 1 Case-series data on presence of category-specific naming differences in HSVE versus SD

HSVE	Domain	JD	PS	JF	MW	SS	YW	EB	Mean	SD
Word–picture matching		1.0	1.0	0.95	0.94	0.88	0.88	0.64	0.90	0.12
Short matched naming	Animal	0.81	0.75	0.88	0.75	0.75	0.81	**0.31***	**0.72***	0.25
	Artefact	0.81	1.00	0.94	1.00	0.94	0.94	**0.88***	**0.93***	0.26
Extended matched naming	Animal	**0.43***	**0.83***	**0.63***	**0.77***	0.66	**0.63***	**0.30***	**0.61***	0.18
	Artefact	**0.77***	**1.00***	**0.97***	**1.00***	0.56	**0.90***	**0.63***	**0.83***	0.18
Specific naming	Animal	0.29	0.14	0.50	0.29	0.29	0.14	0.29	0.28	0.12
	Artefact	0.43	0.21	0.50	0.36	0.43	0.36	0.07	0.34	0.15

SD	Domain	AN	LS	ATe	MA	EK	JCh	ATh	Mean	SD
Word–picture matching		1.0	0.98	0.91	0.89	0.83	0.72	0.70	0.86	0.13
Short matched naming	Animal	0.94	0.69	0.31	0.31	0.38	0.44	0.31	0.48	0.25
	Artefact	0.81	0.69	0.13	0.39	0.44	0.56	0.24	0.44	0.26
Extended matched naming	Animal	1.0	NT	0.20	*0.57**	0.03	NT	NT	0.45	0.43
	Artefact	0.93	NT	0.37	*0.23**	0.07	NT	NT	0.40	0.48
Specific naming	Animal	0.50	0.36	0.21	inc.	0.0	0.29	0.21	0.26	0.17
	Artefact	0.80	0.27	0.20	inc.	0.0	0.20	0.13	0.27	0.28

Statistically significantly worse performance for animals than artefact is denoted in bold type. A significant difference in the reverse direction is shown in italics. Patients are rank-ordered, left-to-right, in terms of overall severity of semantic impairment. NT, not tested; inc., incomplete results. *$p < 0.05$.

remaining patient showing equivalently poor performance in both domains. The group mean performance for living things was significantly worse than for manmade objects [$t(6) = 2.97$, $P = 0.03$]. In contrast, only two/six patients in the SD cohort showed numerically (but not statistically significantly) worse performance for living things, with three showing the reverse pattern and one showing equivalent performance for the two domains. Mean performance across the SD group was the same for the two domains [$t(6) = 0.40$, n.s.]. The differential category effect across groups was significant: a (group × domain) ANOVA confirmed a significant interaction between group and semantic domain [$F(1,12) = 7.56$, $P = 0.02$], reflecting the reliable category effect in HSVE and the absence of such an effect in SD.

To determine whether the category effect observed in the HSVE group was reliable in individual patients, the HSVE cohort was further tested on the much larger extended matched naming test. These data, also shown in Table 1, revealed significant category effects favouring artefacts in six/seven individual patients and was significant for the group as a whole [$t(6) = 3.80$, $P = 0.009$]. Four of the seven SD patients were also tested on the extended naming test and, consistent with previous studies (which have run this extended naming test in more than 20 SD cases, yielding the same null result) (Lambon Ralph *et al.*, 2003), none of them exhibited the same difference on this test [nor as a group: $t(3) = 0.46$, n.s.]. In sum the data show that, in two patient groups with comparably severe semantic deficits resulting from anterior temporal-lobe pathology, category-specific naming impairments were reliably observed for one aetiology (HSVE) but not the other (SD) (group × domain interaction was significant: $F(1,9) = 6.02$, $P = 0.04$].

Different types of damage produce differing levels of category-specificity

We investigated two forms of damage to the trained model. We first replicated our previous work (Rogers *et al.*, 2004*a*, *b*) in which the semantic impairment of SD was simulated by randomly removing connections to and from the semantic units. This manipulation had two functional consequences for the signals passing in and out of the semantic units: (i) they were partially distorted as important weights were removed and (ii) they were 'dimmed' as fewer and fewer connections were available to conduct information through the model, so that inputs became less and less effective at driving semantic unit activations. Under this form of damage, the model does not exhibit a category-specific impairment: naming is equally poor for both domains and the overall degree of naming impairment is comparable to that observed in the patients (Fig. 3B and C).

The second form of damage was motivated by considering how processing in the intermediating layer might be disrupted so as to produce the category-specific pattern typically observed in HSVE. Specifically, we damaged the model by disrupting the values of the weights projecting in or out of the semantic layer with increasing amounts of random noise. This manipulation distorts the signals passing between layers without attenuating them: with increasing damage, inputs can still strongly drive the semantic units, albeit in increasingly random directions. This in turn means that the model will tend to confuse items with similar internal representations. Since animals tend to have somewhat more similar internal representations than do artefacts, we reasoned that this form of disruption would tend to produce a category-specific impairment—and indeed, subject to this form of disruption, the model consistently produced a category-specific naming deficit of a magnitude comparable to that observed in the patients (Fig. 3B and C). In other words, different forms of damage applied to an identical locus in the same model produced no category effect in one case yet a strong category effect in the other.

What accounts for the difference? The explanation concerns the interaction of the form of damage with the similarity structure of the model's internal representations. Consider again what happens when signals are distorted but not dimmed: it becomes difficult for the model to differentiate items with similar representations and, since animals from different basic-level categories have somewhat similar representations, distortion severely affects basic-level animal naming. In contrast, basic-level clusters are dispersed for artefacts, so that small amounts of distortion are better tolerated in this domain. When signals are dimmed, however, shared structure in the animal domain becomes something of an asset: because the animals are all somewhat similar to one another, they form a broad basin of attraction which can guide the internal state toward the correct neighbourhood even when inputs are dimmed. As signals are also partially distorted, however, the model may still settle into an incorrect neighbouring attractor and activate the name of a semantic neighbour. In the domain of artefacts, basic-level clusters are widely spread so no broad 'artefact' attractor forms in this part of the space. Consequently, the model requires more robust input to get close enough to the correct basic-level attractor. When signals are dimmed, the representation state fails to get close to any attractor basin and thus fails to activate any name units. Thus, the SD-variant of the model makes comparable numbers of errors in both domains but for somewhat different reasons.

Testing three predictions from the PDP simulations

The preceding simulation establishes the face validity of our working hypothesis: in the context of the model, different forms of disruption to the same locus can produce either an SD-like or an HSVE-like pattern of semantic impairment. The particular form of disruption employed to

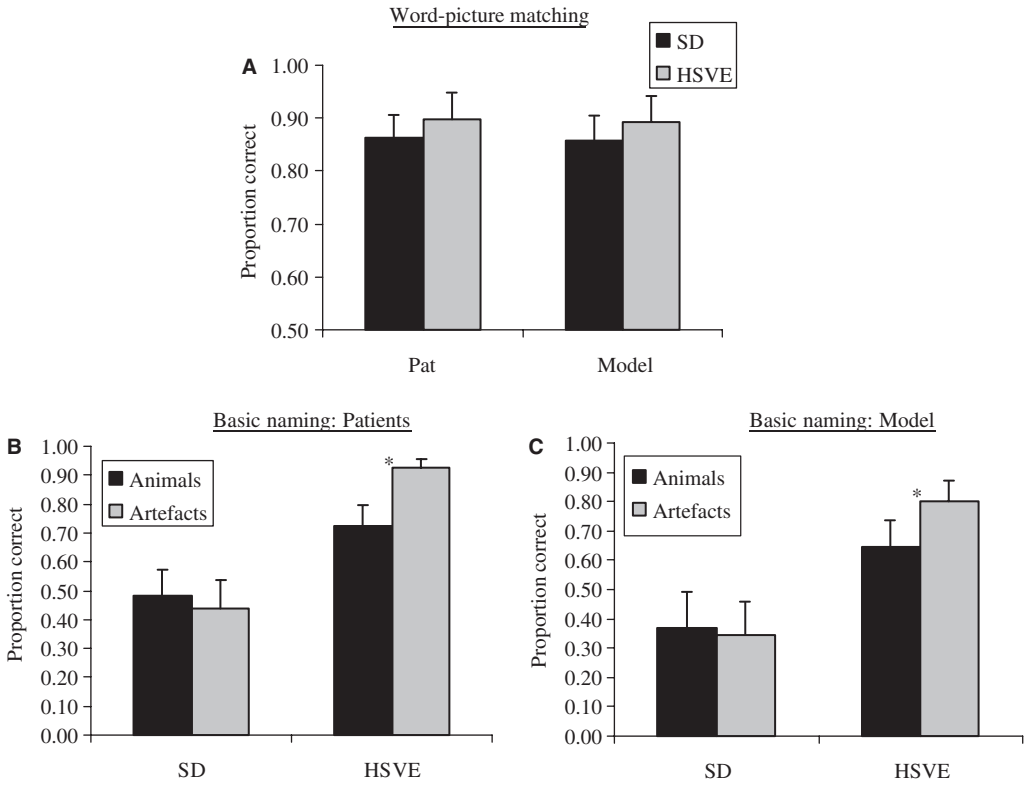

Fig. 3 Comprehension and basic naming performance for patients and model. Mean and standard errors of proportion correct for word–picture matching and basic-level naming of animals and artefacts, plotted for each patient group and for the SD- and HSVE-variants of the model. Model performance was assessed at seven points for each form of damage, at the points where word–picture matching scores most closely matched the scores obtained by the individual patients. The figure shows (**A**) that the two groups perform comparably on the word–picture matching task; (**B**) that significant category effects are observed in HSVE but not SD and despite matched performance in word–picture matching, the SD group is more anomic than the HSVE group and (**C**) all of these effects are replicated by the model. An asterisk denotes a significant category effect.

simulate HSVE was, however, selected *post hoc* precisely because, given an adequate understanding of how the model works, it seemed likely to produce an HSVE-like pattern of behaviour. To assess whether this manipulation provides a true explanation of the semantic impairment in HSVE, we tested three novel predictions offered by the model about differences in patient behaviour.

First, patients with SD should produce less information from any given stimulus than patients with HSVE. In the case of picture naming, for example, the reduction in activation along the visual semantic verbal pathway means that the SD-variant—like real SD patients (Lambon Ralph *et al.*, 2001)—becomes very anomic. In contrast, the HSVE-variant distorts the representations but does not dampen the activation in this pathway, so that it should produce comparatively more names as output. The discrepancy predicted by the model is already apparent in the data shown in Fig. 3 and Table 1: despite being matched for overall comprehension impairment, the SD patients are considerably more anomic than the HSVE group (group × task interaction: $F(1,12) = 8.27$, $P = 0.01$; naming in SD versus HSVE: $t(12) = 3.62$, $P = 0.004$).

The second prediction relates to the types of naming error made by the model and the patients. Under damage both model variants make a mixture of correct responses, category-coordinate (e.g. FOX→'dog') and other errors (omissions and superordinate errors—e.g. FOX→'animal'). From our explanation of the category-specific effect in HSVE, naming errors often occur through the confusion of items with similar representations—so this group should make a comparably large number of category-coordinate errors. In contrast, naming errors in SD can frequently occur when, as a consequence of dimmed representations, the model cannot get close enough to the correct attractor to activate any basic-level name—hence such patients should make comparably fewer category-coordinate errors and more superordinate and omission errors. This prediction was also confirmed by the patient data (Fig. 4A): the rate of coordinate semantic errors is higher in the HSVE than SD, and the rate of other errors (in particular omissions) is higher for SD than HSVE.

The third prediction stems from our hypothesis that category-effects in HSVE arise because animal representations inhabit a more densely packed conceptual

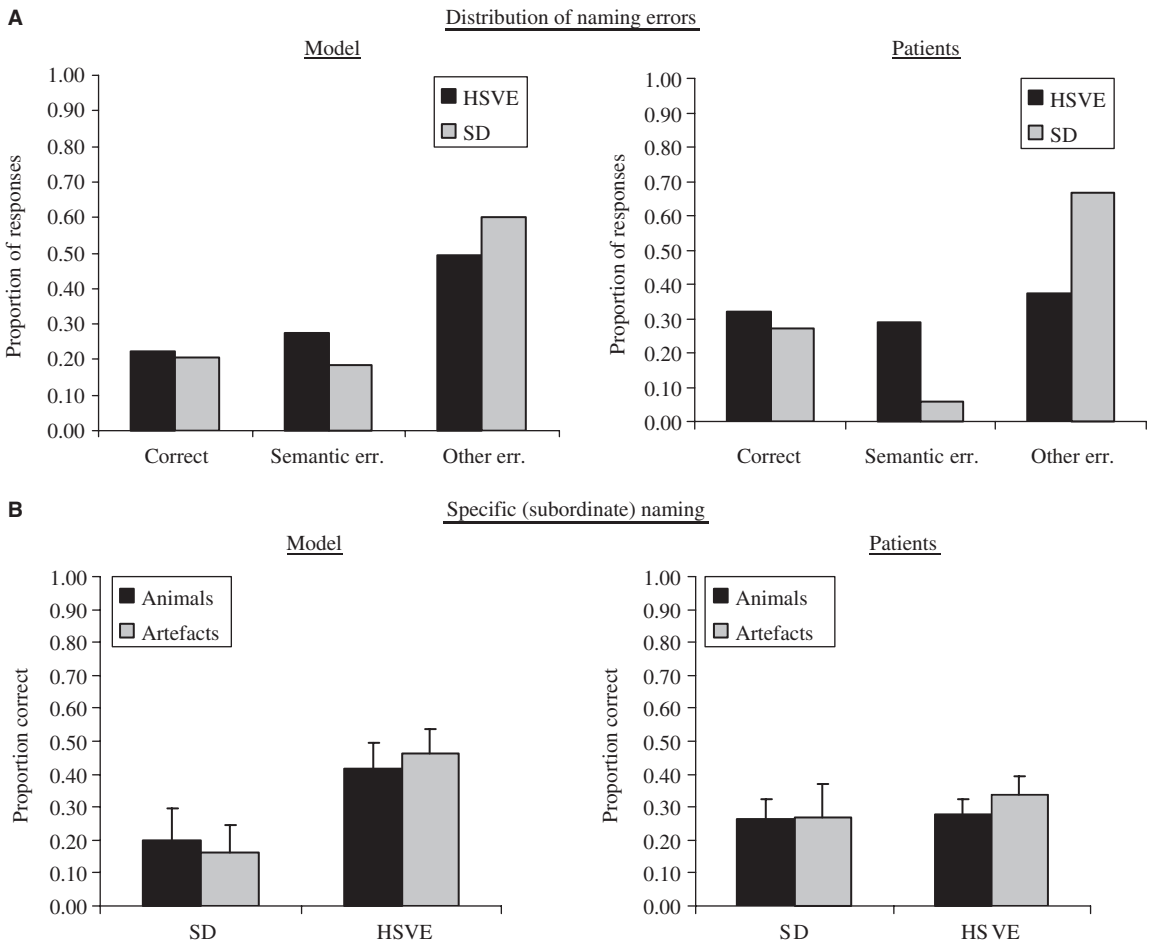

Fig. 4 Novel predictions from the model and corresponding patient results. (**A**) Predicted and observed proportion of responses that are (i) correct, (ii) category-coordinate errors or (iii) omission/other errors in the specific naming task, when overall accuracy is matched. Because SD patients are overall more anomic than HSVE, we selected the four best-performing SD cases and the four worst-performing HSVE cases in order to match for overall naming accuracy in this comparison. Though the two cohorts have similar overall accuracy, the HSVE group yields more category-coordinate errors and fewer omission/other errors than the SD group, as predicted by the model. (**B**) Accuracy on the specific naming task in the two groups predicted by the model and observed mean accuracy in the patient groups. The error bars indicate standard errors.

neighbourhood. This suggests that category-effects should be eliminated in HSVE if the conceptual neighbourhoods for the living and nonliving items to be named are comparably dense. Such is the case for subordinate-level naming, where both animal (e.g. different kinds of dog) and artefact (e.g. different makes of car) representations are tightly packed. When the model was tested on subordinate-level naming under the two forms of damage, (i) the category effect in the HSVE-variant was greatly attenuated; (ii) no category effect was observed in the SD-variant and (iii) overall performance was considerably worse than basic-level naming for both variants (Fig. 4B). All three of these characteristics were observed when the patients were tested on a specific naming task in which animals and manmade objects were matched for familiarity and name agreement. In the HSVE cohort, the category effect observed in basic-level naming was greatly attenuated, eliminated, and even (in one patient) reversed, so that no individual nor the group showed a reliable category effect [$t(6) = 1.12$, $P = 0.30$]. Note that this elimination of the category effect in the HSVE group was not attributable to floor performance—for instance, the highest-scoring individual patient with HSVE, JF, produced correct names for exactly half of both the animals and the artefacts. No effect of category was apparent in the SD cohort either and, as is apparent from Figs. 3 and 4A, subordinate naming was considerably worse than basic naming for both groups.

General discussion

This computational neuropsychological investigation found that, as is growingly suspected in the literature (Noppeney *et al.*, 2007), SD produces a generalized semantic impairment (Lambon Ralph *et al.*, 2003), while semantic deficits in the context of HSVE often have a category-specific flavour (with worse performance on living

than non-living things: Capitani *et al.*, 2003). We found that this pattern holds even when the patient groups are matched for overall severity of their semantic impairment.

Given the very overlapping anterior temporal focus to the damage observed in both SD and HSVE (Noppeney *et al.*, 2007), we investigated the possibility that the qualitative differences between the semantic impairment of HSVE versus SD might arise from different types of functional disruption to largely the same locus. Building upon our previous model of semantic impairment in SD (Rogers *et al.*, 2004*a*, *b*), we were able to demonstrate that category effects in naming are routinely observed when the intermediating (semantic) units in the model robustly propagate increasingly noisy signals but no such effects are observed when the same units become increasingly unable to propagate activation. As a simple mnemonic, category effects are observed when semantic representations are distorted (HSVE) but not when they are dimmed (SD). When representations are distorted (HSVE), items with similar representations are easily confused. Consequently, it is more difficult to retrieve names or other semantic information in domains with rich similarity structure (such as living things) than in domains with dispersed representations (artefacts). When representations are dimmed (SD), rich similarity structure can partially benefit processing because such structure produces broad basins of attraction that can help guide the system toward the correct semantic neighbourhood (for further details, see Rogers and McClelland, 2004; Rogers *et al.*, 2004*a*, *b*).

This account also led to three non-trivial predictions about naming in HSVE and SD that were confirmed by further neuropsychological assessment. Specifically, distorting (HSVE) rather than removing (SD) connections produces: (i) a lesser degree of anomia for any given level of comprehension impairment; (ii) a different distribution of naming errors with higher rates of category-coordinate errors and correspondingly lower numbers of omission/superordinate errors in HSVE than SD and (iii) a category-specific pattern when concepts are probed at the basic level but not at the subordinate level. This third prediction/finding poses a particular challenge to some other theories of category-specific impairment because it demonstrates that the very same patients may show robust category-specific effects under some test conditions yet no such pattern in other conditions. Put differently, patients in the HSVE group do not really have a category-specific impairment at all. Rather, they have severe difficulty discriminating amongst items with somewhat similar semantic representations. In some tasks, they appear to show category-specific effects because representations of living things tend to be more similar to one another overall than do representations of artefacts. Similar proposals have previously been put forward in terms of 'lumpy semantic space' (Hillis *et al.*, 1990) or, through formal analysis, in terms of greater numbers of shared attributes for living than non-living concepts (Lambon Ralph *et al.*, 1998; Tyler *et al.*, 2000; Garrard *et al.*, 2001). To our knowledge this is the first theory, however, to reconcile such 'crowding' accounts of category-specific impairment in HSVE with the absence of a category-specific pattern in SD.

Our account suggests that the flow of activation through the anterior temporal cortex is disrupted in different ways in HSVE and SD. There is not a direct relationship between units and connections in this type of computational model and microscopic features of brain tissue (neurons, synapses, axons, neurotransmitters, etc.) Clearly, therefore, there is further work to be done in linking these functional consequences of two types of damage in the model to the actual neuropathology observed in the two diseases. Though a comprehensive review of these issues is beyond the scope of this article, it is worth briefly mentioning three working hypotheses about neuropathological differences that could result in distorted or dimmed representations.

The first hypothesis is a direct neural proxy of the simulation—that is to say, damage occurs in the same location but is of a different type: both diseases are centred on the anterior temporal lobe but involve knife-edge atrophy in the case of SD (resulting from neuronal loss and corresponding thinning of the white matter) and full-thickness necrosis of the cortex and underlying white-matter in HSVE. Future studies utilizing pathological and neuroimaging analyses are needed to test how these different types of damage affect the transmission of neural activity to and from the damaged anterior temporal structures in much greater detail.

Second, it is possible that distorting effects may arise from damage to other cortical systems that interact with the anterior temporal lobe semantic system. For instance, HSVE frequently produces extensive damage in frontal as well as temporal-lobe regions, potentially disrupting regions known to be involved in semantic 'control'/'selection' (Thompson-Schill *et al.*, 1997; Jefferies and Lambon Ralph, 2006). In this regard, it may be important that the semantically impaired HSVE patients also had disinhibited or apathetic behaviour and poor problem-solving skills (though performance on classical executive tests is also compromised by their behavioural changes). The possibility arises, therefore, that their executive dysfunction might be the source of the semantic distortion. We think this latter possibility is unlikely to be the correct explanation for the following reason. We recently compared impaired semantic cognition in SD versus stroke aphasia (Berthier, 2001; Jefferies and Lambon Ralph, 2006). The semantic impairment in the aphasic group seems to be primarily related to poor control of semantic information for task/context appropriate behaviour and is consistent with the frontoparietal lesions of these patients (Berthier, 2001; Jefferies and Lambon Ralph, 2006) and with the proposed control role of ventrolateral, prefrontal cortex in semantic processing (Thompson-Schill *et al.*, 1997). Despite this, we found that the patients' executive dysfunction/semantic impairment did not produce a category-specific impairment in any case. Although preliminary, these observations suggest

that the 'deregulation' of semantic cognition (Jefferies and Lambon Ralph, 2006) arising from damage to frontoparietal structures does not produce an HSVE-like (category-specific) semantic syndrome.

Third, there are critical differences in the time-course of the two diseases that may be important in understanding the functional consequences of the neuropathology in each case. Specifically, SD is a chronic and progressive illness in which frontotemporal cortex gradually disintegrates. HSVE, by contrast, is a very rapidly progressing illness that is typically halted by treatment with anti-viral medications—so that the characteristic time course is an acute injury followed by a long (often years-long) period of recovery before testing. The behavioural profile in HSVE thus typically reflects the performance of a system that has been seriously damaged but has undergone some degree of relearning/reorganization. Correspondingly, the robust transmission of distorted neural signals that explains the behavioural profile of HSVE in our theory may capture the behaviour of a damaged semantic network that has undergone a substantial degree of relearning (for a computational instantiation of this idea, see Welbourne and Lambon Ralph, 2005).

Finally, the current study was designed to help explain systematic differences in two semantic syndromes, both resulting from anterior temporal-lobe pathology. It is important to note that we have not set out to explain the full variety of category-specific patterns reported in other single-case studies. Nevertheless we may inquire what our theory has to say about other (non-HSVE) patients with category specificity and, in particular, about patients with the opposite dissociation (animals > artefacts)? We suggest that case-series comparison of different patient groups, each exhibiting relatively consistent patterns of behaviour, will be critical in addressing such phenomena. In this spirit, it appears that there may be two other coherent patient groups that could form the basis of further comparative, case-series studies. The first are patients with category-specific agnosia for living things following posterior temporal lobe infarction (Humphreys and Forde, 2001) who have a breakdown in visual perception or the transmission of visual information into the semantic system. A considerable body of work now suggests that the factor critical to apparent category-specific patterns in this group is the degree of visual and semantic overlap amongst the test items (Tranel *et al.*, 1997; Humphreys and Forde, 2001). The second group consists of patients with poor knowledge about manmade objects. On the few occasions that this has been explored in a case-series rather than single-case design, impaired knowledge about objects was associated with impoverished information about how to manipulate objects in the context of inferior parietal lesions (Buxbaum and Saffran, 2002). It is reasonable to suppose, therefore, that disruption of the mapping between semantic representations and representations of praxis and/or function will produce the reverse category-specific pattern (artefacts < animals). Exploration of these factors in these two patient groups may well offer the best next steps in future research on category-specific impairments.

Acknowledgements

This work was supported by a grant from Research into Ageing and Dunhill Medical Trust and a National Institute of Mental Health IBSC grant (MH64445). We would like to thank all the patients and the Encephalitis Society (UK) for their continued support of our research. We are particularly grateful to John Hodges, Karalyn Patterson, David Neary, Julie Snowden and Roy Jones for referring some of the patients to us.

References

Barbarotto R, Capitani E, Laiacona M. Naming deficit in herpes-simplex encephalitis. Acta Neurol Scand 1996; 93: 272–80.

Berthier ML. Unexpected brain-language relationships in aphasia: evidence from transcortical sensory aphasia associated with frontal lobe lesions. Aphasiology 2001; 15: 99–130.

Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. Non-verbal semantic impairment in semantic dementia. Neuropsychologia 2000; 38: 1207–15.

Buxbaum LJ, Saffran EM. Knowledge of object manipulation and object function: dissociations in apraxic and nonapraxic subjects. Brain Lang 2002; 82: 179–99.

Capitani E, Laiacona M, Mahon B, Caramazza A. What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. Cogn Neuropsychol 2003; 20: 213–61.

Caramazza A, Shelton JR. Domain-specific knowledge systems in the brain: the animate-inanimate distinction. J Cogn Neurosci 1998; 10: 1–34.

Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR. A neural basis for lexical retrieval. Nature 1996; 380: 499–505.

Eggert GH. Wernicke's works on aphasia: a source-book and review. The Hague: Mouton; 1977.

Farah MJ, McClelland JL. A computational model of semantic memory impairment: modality specificity and emergent category specificity. J Exp Psychol Gen 1991; 120: 339–57.

Garrard P, Lambon Ralph MA, Hodges JR, Patterson K. Prototypicality, distinctiveness and intercorrelation: analyses of the semantic attributes of living and nonliving concepts. Cogn Neuropsychol 2001; 18: 125–74.

Garrard P, Lambon Ralph MA, Hodges JR. Semantic dementia: a category-specific paradox. In: Forde EME, Humphreys GW, editors. Category-specificity in mind and brain. Hove: Psychology Press; 2002.

Gloor P. The temporal lobe and the limbic system. Oxford: Oxford University Press; 1997.

Hillis AE, Rapp B, Romani C, Caramazza A. Selective impairment of semantics in lexical processing. Cogn Neuropsychol 1990; 7: 191–243.

Hodges JR, Patterson K, Oxbury S, Funnell E. Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. Brain 1992; 115: 1783–806.

Humphreys GW, Forde EME. Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. Brain Behav Sci 2001; 24: 453–96.

Jefferies E, Lambon Ralph MA. Semantic impairment in stroke aphasia vs. semantic dementia: a case-series comparison. Brain 2006; 129: 2132–47.

Kapur N, Barker S, Burrows EH, Ellison D, Brice J, Illis LS, et al. Herpes-simplex encephalitis: long-term magnetic-resonance-imaging and neuropsychological profile. J Neurol Neurosurg Psychiatry 1994; 57: 1334–42.

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.

Lambon Ralph MA, McClelland JL, Patterson K, Galton CJ, Hodges JR. No right to speak? The relationship between object naming and semantic impairment: neuropsychological evidence and a computational model. J Cogn Neurosci 2001; 13: 341–56.

Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. Semantic dementia with category specificity: a comparative case-series study. Cogn Neuropsychol 2003; 20: 307–26.

Levelt WJM. Speaking: from intention to articulation. Cambridge, MA: MIT Press; 1989.

Martin A, Chao LL. Semantic memory in the brain: structure and processes. Curr Opin Neurobiol 2001; 11.

McRae K, Cree GS. Factors underlying category-specific semantic deficits. In: Forde EME, Humphreys GW, editors. Category-specificity in brain and mind. Hove: Psychology Press; 2002.

Noppeney U, Patterson K, Tyler LK, Moss H, Stamatakis EA, Bright P, et al. Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. Brain 2007; In press.

Osterrieth P. Le test de copie d'une figure complexe. Arch Psychol 1944; 30: 205–550.

Raven JC. Coloured progressive matrices: sets A, AB, B. London: HK Lewis; 1962.

Rogers TT, Lambon Ralph MA, Garrard P, Bozeat S, McClelland JL, Hodges JR, et al. The structure and deterioration of semantic memory: a neuropsychological and computational investigation. Psychol Rev 2004; 111: 205–35.

Rogers TT, McClelland JL. Semantic cognition: a parallel distributed processing approach. Cambridge, MA: MIT Press; 2004.

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

Tranel D, Logan CG, Frank RJ, Damasio AR. Explaining category related effects in the retrieval of conceptual and lexical knowledge for concrete entities: operationalization and analysis of factors. Neuropsychologia 1997; 35: 1329–39.

Tversky B, Hemenway K. Objects, parts and categories. J Exp Psychol Gen 1984; 113: 169–91.

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

Warrington EK. Short recognition memory test. Hove: Psychology Press; 1996.

Warrington EK, James M. The visual object and space perception battery. Bury St Edmunds: Thames Valley Test Company; 1991.

Warrington EK, McCarthy R. Category specific access dysphasia. Brain 1983; 106: 859–78.

Warrington EK, Shallice T. Category specific semantic impairments. Brain 1984; 107: 829–54.

Welbourne SR, Lambon Ralph MA. Subtracting subtractivity? A connectionist account of recovery in single word reading following brain damage. Cogn Affect Behav Neurosci 2005; 5: 77–92.

Wise R. Language systems in normal and aphasic human subjects: functional imaging studies and inferences from animal studies. Br Med Bull 2003; 65: 95–119.