

## CONCEPTUAL STRUCTURE AND THE STRUCTURE OF CONCEPTS: A DISTRIBUTED ACCOUNT OF CATEGORY-SPECIFIC DEFICITS

L.K. Tyler<sup>1</sup> H.E. Moss<sup>1</sup> M. R. Durrant-Peatfield<sup>2</sup> & J.P. Levy<sup>2</sup>

<sup>1</sup>Centre for Speech and Language, Department of Experimental Psychology, University of Cambridge

<sup>2</sup>Department of Psychology, Birkbeck College, University of London

### ABSTRACT

We present a new account of the fine-grained structure of semantic categories derived from neuropsychological, behavioral and developmental data. The account places theoretical emphasis on the functions of the referents of concepts. We claim (i) the perceptual features representing specific functional mechanisms are strongly correlated with their function; (ii) the distinctiveness of functional features correlated with perceptual features varies across semantic domains; (iii) Category structure emerges from the complex interaction of these variables. The representational assumptions that follow from these claims make strong predictions about what types of semantic information are preserved in patients showing category-specific deficits following brain damage. These claims are illustrated with a connectionist simulation which, when damaged, shows patterns of preservation of distinctive and shared functional and perceptual information which varies across semantic domains. The data model both dissociations between knowledge for artefacts and for living things and recent neuropsychological evidence concerning the robustness of functional information in the representation of concepts.

### INTRODUCTION

Neuropsychological studies of patients with semantic impairments provide important data concerning the structure and content of semantic memory. Detailed investigations of patterns of impaired and preserved semantic knowledge demonstrated by such patients provide invaluable insights into the way in which semantic information is stored and accessed. One especially important source of data comes from patients with what appear to be 'category-specific deficits'; selective impairments restricted to specific domains or categories of semantic knowledge (e.g., de Renzi & Lucchelli, 1994; Farah, McMullen & Meyer, 1991; Hart & Gordon, 1992; Hillis & Caramazza, 1991; Moss, Tyler & Jennings, 1997; Sacchett & Humphreys, 1992; Sartori & Job, 1988; Warrington & McCarthy, 1983; 1987; Warrington & Shallice, 1984; for a review see Saffran & Schwartz, 1992). The most frequently observed pattern is that knowledge of living things is impaired with knowledge of artefacts relatively preserved, although the reverse pattern has occasionally been reported (Hillis & Caramazza, 1991; Moss & Tyler, 1997; In Press; Sacchett & Humphreys, 1992; Warrington & McCarthy, 1983; 1987). Semantic deficits force us to confront the nature of semantic memory and its organization. How is semantic knowledge represented and structured such that brain damage can lead to specific patterns of preserved and impaired knowledge? Do these deficits indicate the discrete localization of different types of semantic information, or can they be accounted for by more diffuse and distributed forms of mental representation?

There have been reports of patients with selective deficits for a variety of categories, such as fruits, vegetables and animals (Caramazza & Shelton, 1998; Hart & Gordon, 1992) and domains of knowledge (living vs. non-living things; abstract vs. concrete words). Such selective deficits have been interpreted as evidence for two kinds of semantic organization. In one, conceptual knowledge is conceived of as being organized into distinct domains or categories such as animals, plants and tools (Caramazza & Shelton, 1998). Selective deficits have also been interpreted as evidence for the organization of conceptual

knowledge in terms of different types of critical information, such as perceptual and functional features (Warrington & Shallice, 1984; Warrington & McCarthy, 1983; 1987). This second account was motivated by the finding that, for some patients, the pattern of intact and impaired categories did not conform straightforwardly to a living/non-living dissociation. For example, one patient, JBR, showed a living things deficit but also had problems with some categories within the non-living domain -- such as musical instruments and gemstones (Warrington & Shallice, 1984), while a second, YOT, had an artefact deficit but also had difficulty with body parts (Warrington & McCarthy, 1987). Warrington and colleagues argued that musical instruments and gemstones are similar to living things in that they are primarily distinguishable in terms of perceptual properties, whereas artefacts and body parts are categories of knowledge for which function is most salient. Thus, if brain damage selectively impairs visual semantic knowledge, this will have a greater effect on living things than artefacts (with the exception of sub-categories such as musical instruments), giving rise to the typical category-specific pattern.

The reformulation of category structure according to differential weighting of perceptual and functional properties gained support from both behavioral and computational studies. On the one hand, many patients with category-specific deficits for living things have been shown to have specific problems with the visual properties of the impaired category (e.g., de Renzi & Lucchelli, 1994; Farah, McMullen & Meyer, 1991; Hart & Gordon, 1992; Moss, Tyler & Jennings, 1997; Sartori & Job, 1988; but see Caramazza & Shelton, 1998, for a different view). On the other, computational simulations have shown how a system in which living things have a higher proportion of visual semantic properties to functional properties than do artefacts, exhibits a greater deficit for living things when the visual semantic properties are 'lesioned' (Farah & McClelland, 1991; Small, Hart, Nguyen & Gordon, 1995). However, there have been a number of recent reports of patients who have living things deficits without accompanying selective deficits for perceptual properties (Laiacona, Capitani & Barbarotto, 1997; Lambon Ralph, Howard, Nightingale & Ellis, 1998; Moss,

Tyler, Durrant-Peatfield & Bunn, 1998), as well as cases of patients who have poor knowledge of visual information without an accompanying disproportionate deficit for living over non-living things (Lambon Ralph et al. 1998), thus questioning the behavioral evidence for distinctions between different types of semantic properties as the appropriate account of category specific deficits. Moreover, the empirical support for the assumption that living things have a greater proportion of visual than functional properties, that is essential for Farah & McClelland's computational demonstration of the dissociation, has also been questioned (Caramazza & Shelton, 1998; see Tyler & Moss, 1997 for a related argument).

Recasting the living/non-living distinction in terms of the relative contributions of perceptual and functional properties does not change the basic claim that category-specific deficits are associated with damage to distinct stores of knowledge. The basis for differentiation has simply changed from category membership to information type, such that functional and perceptual information are stored separately and can be independently impaired by focal brain damage. In fact, several accounts suggest, at least implicitly, that localization is based on both category membership and information type; for example, that there is a separate subsystem within semantic memory dedicated to the visual properties of living things (Farah et al., 1989; Gainotti, Silveri, Daniele & Giustolisi, 1995; Warrington & McCarthy, 1987; see Moss, Tyler & Jennings, 1997, for a discussion).

Recently, a new approach to the study of category-specific effects has started to be developed, suggesting that these deficits can emerge as a result of differences in the content and structure of concepts in different semantic categories, rather than from broad divisions of semantic memory in independent stores. In this framework, category-specific deficits are not necessarily the result of selective damage to specific stores of one or other type of semantic information; they can also result from damage throughout a unitary semantic system, such as that described by Caramazza, Hillis, Rapp & Romani (1990).

The possibility that category-specific effects could emerge from non-selective damage to a unitary system depends on the assumption that concepts are represented as patterns of activation distributed over multiple units corresponding to different semantic properties or 'micro-features'. This has been modeled in a number of connectionist simulations of normal semantic memory (e.g., Masson, 1995; McRae, de Sa & Seidenberg, 1997). There are several reasons why this promises to be a fruitful approach. First, this system of representation makes generalization to new concepts and pattern completion simple to perform (Rumelhart, Hinton & McClelland, 1986). In addition, the relations among similar concepts are readily captured as overlapping patterns of activation over common sets of micro-features (McRae et al., 1997; Masson, 1995) and the possibility of partial activation of representations allows variations in conceptual representations over different contexts of use and the differential time course of activation of semantic properties to be modeled (e.g., Kawamoto, 1993).

Second, distributed connectionist models show graceful degradation when damage is simulated by removing units or connections between units (e.g., Hinton & Sejnowski, 1986; Plaut & Shallice, 1993; Devlin, Gonnerman, Andersen, & Seidenberg, 1998). This characteristic is important, since patients who have sustained brain damage rarely show an all-or-none deficit for one specific process or knowledge type, but rather show varying degrees of impairment for different functions. Thus, category-specific deficits are not truly specific, since performance is rarely perfect in the 'preserved' category; there is usually some degree of impairment, albeit milder than for the affected category. Distributed connectionist systems are well suited to model such graded impairments.

These considerations have motivated recent studies in which category-specific deficits are modeled by lesioning distributed connectionist models of semantic memory (Devlin et al., 1998; see also Gonnerman et al., 1997). Devlin et al. (1998) were specifically concerned to show how such a model could account for the emergence of category-specific deficits for living things as a result of the widespread, patchy brain damage caused by Alzheimer's Disease (DAT). Like the earlier model of Farah & McClelland (1991), Devlin et al. (1998) represented concepts as vectors distributed over two types of semantic feature -- perceptual and functional -- and they incorporated the finding that living things have a higher proportion of perceptual to functional properties than do non-living things (as determined by property

generation norms). However, additional characteristics of the vector representations used enabled the model to produce category-specific deficits, not only as a result of selective damage to either the functional or the perceptual properties (as was the case in Farah & McClelland's study) but also as a result of random, non-selective damage to a certain proportion of all properties, including both functional and perceptual types. The additional characteristics of the Devlin et al. (1998) model were:

1) *Distinctiveness of features:*

The features of a semantic representation were more or less informative (or distinctive) on a probabilistic basis; that is, they varied in how reliably they picked out one category member from others within the same category. Living things generally had more shared semantic features and fewer distinctive features than artefacts.

2) *Correlations:*

Features which regularly occurred together in semantic representations reinforced each other with mutual activation. There were more of these strongly correlated features for living things than for artefacts, as has been demonstrated in norming studies with unimpaired subjects by McRae et al. (1997) and Devlin et al. (1998).

Mild non-selective damage (simulated by removing a proportion of the connections among nodes in the semantic layer of the network) impaired performance on artefacts more than living things because there were fewer correlated features for artefact items to provide compensatory activation when individual distinctive features were lost. However, when lesions were more severe, sets of correlated features no longer accrued sufficient activation to support each other, and so large numbers of living things were impaired together, resulting in a deficit for living things. The Devlin et al. model predicts, then, a progression from a mild artefact deficit to a living things deficit with increasing severity of semantic impairment, a pattern for which Gonnerman et al. (1997) found some evidence in their group of DAT patients (but see Garrard et al. (1998) for a counter-argument).

The Devlin et al. (1998) model was specifically designed to address the occurrence of category-specific deficits in patients with DAT. Random damage throughout a distributed semantic system would seem to be a more plausible approximation to the widespread, patchy brain damage associated with DAT, than does selective damage to a discrete store of a particular kind of information. It could be argued, however, that the distributed account is less convincing when applied to other etiologies typically associated with category-specific deficits, most notably, patients who have suffered from Herpes Simplex Encephalitis (HSE). HSE patients have relatively focal damage in the sense that it tends to be confined to specific brain regions, especially the medial temporal lobes bilaterally including the hippocampi and limbic association cortex. The fact that such patients frequently have a pronounced deficit for living things has led to the hypothesis that they have sustained damage to the specific semantic stores for living things and/or visual semantic information Gainotti, Silveri, Danielli & Giustolisi, 1995). However, given the role that these neural structures play in receiving and transmitting information between different brain regions, it is likely that they form part of an extensive distributed neural system (Damasio, Grabowski, Tranel, Hichwa & Damasio, 1996).

Of the few patients reported in the literature to have specific deficits for artefacts relative to living things, there has been some suggestion of an association with damage to the fronto-parietal area, which again may seem to implicate a discrete, localized semantic store. However, this is based on a small number of cases, and detailed neuroanatomical data is not available (e.g. Sacchett & Humphreys, 1992; Warrington & McCarthy, 1983) and there are several patients who do not unambiguously show this association. We have recently reported the case of a patient with generalized cerebral atrophy who developed a disproportionate deficit for artefacts over living things as her semantic impairment became more severe, indicating that this kind of deficit can arise in the absence of any obvious focal lesion (Moss & Tyler, 1997; submitted). We have also seen a similar pattern for a second progressive aphasic patient, AA, who has a semantic disorder in the context of generalized cerebral atrophy (Moss, Tyler & Devlin, 1999). Lambon-Ralph et al. (1998) have also reported a patient, IW, who shows a significantly greater deficit for artefacts than living things, and who has the pattern of left temporal lobe damage typical of semantic dementia. Hillis & Caramazza (1991)'s artefact deficit patient, JJ, had damage to the left temporal lobe as a result of thromboembolic stroke<sup>1</sup>.

These cases suggest that artefact deficits do not necessarily arise from fronto-parietal lesions.

Thus, although the distinction between diffuse and focal brain damage is difficult to define, there does not appear to be strong evidence that categories of semantic knowledge are necessarily localized to discrete regions of the brain; therefore a model which explores the emergence of selective deficits in the context of random, widespread damage throughout a distributed system of semantic representation may be a more fruitful approach.

In this paper we present a new model of conceptual representation in semantic memory, that can account for category-specific deficits, without assuming that there are distinct subsystems for the storage of living and non-living things, or of perceptual and functional features. Our model is similar in many respects to that of Devlin et al. (1998), in that distinctiveness and correlatedness of properties are major determinants of conceptual structure, but we advance on the earlier study in an important way; we place considerable theoretical emphasis on the role of functional information in our account of conceptual representation and the cognitive system's resistance to damage.

#### Functional Information:

Whereas perceptual information can be argued to be derived straightforwardly from the senses, functional information concerns the manner in which an object interacts with the environment and with the observer. Perceptual information is what an object looks like, feels like, tastes like and so on. Functional information is what gives the object a meaning in a dynamic environment characterized by cause and effect. The semantic representation of *telephone*, for example, must include what a telephone looks and feels like. The function of a telephone is also recorded in such a manner that the set of perceptual features representing a telephone is sufficient to make readily available its functional significance. A person 'knows' about a telephone (i.e. what significance the object has in the environment) only when the person understands what it does. Likewise, the semantic representation of animals may be associated from an early age with information that allows one to anticipate the many non-contingent activities performed by living things, as Mandler (1992) has claimed.

Central to our model is the claim that functional semantic information is a particularly important aspect of conceptual representations and more resilient to brain damage than other kinds of information (see Tyler & Moss, 1997; Moss, Tyler, Patterson & Hodges, 1995). For an artefact, it seems clear that its purpose or function lies at the heart of its conceptual representation (e.g. Barton & Komatsu, 1989; Keil, 1986), and that there is a strong relationship between this specific function and the physical form of the object (e.g. de Renzi & Lucchelli, 1994; Wierzbicka, 1985; Moss, Tyler & Jennings, 1997). For example, the different shapes of artefacts such as a knife, a fork and a spoon depend on their different functions (to cut, spear and scoop food respectively). Each artefact tends to have a unique function, which is associated with distinctive perceptual features. For example, the pointed prongs on the end of a fork distinguish a fork from other items in the category of cutlery. We are simplifying here by suggesting that the function of a man-made object is completely distinctive and always differentiates it from all other items in the category. Some objects do have the same or very similar functions (e.g. a cup and a mug are both used to drink hot liquids from). However, a detailed analysis often reveals highly specific functional and perceptual properties that are truly distinctive (e.g. the size of the vessel and the position of the handle in relation to the situations in which cups and mugs are typically used; Wierzbicka, 1985). Distinctiveness of features is a matter of degree; our claim here is that the functional properties of artefacts are relatively distinctive.<sup>2</sup>

In the neuropsychological literature, functional information is generally considered to be relatively unimportant for the representations of living things. However, we take a different view, claiming that one type of functional information - which we term *biological function* - is important (Tyler & Moss, 1997). This concerns the range of activities that living things engage in, such as moving, flying, eating, drinking, seeing, hearing, or sleeping. This claim is largely motivated by developmental research showing that from an early age, infants distinguish between living and non-living things on the basis of the ways in which they interact with the environment (Mandler, 1992; Mandler & McDonough, 1993). Mandler proposed that functional information is ascribed to living things and artefacts as a result of the

same cognitive processes; infants observe the events in which living and non-living things take part and then base their interpretation of what kind of thing an object is on a very general analysis of the movements and spatial relations that characterize the event. The movements of living things are generally complex (i.e., the beating of wings, the motion of legs) and their trajectories through space are unpredictable, whereas an artefact's movement through space tends to be smooth and predictable. Most importantly, the perceptual analysis of movement is combined with the notion of agency and contingency. Things whose movements are erratic and whose trajectory is generally unpredictable, initiate their own movement. Moreover, these things interact with the environment (and the infant) in a non-contingent manner. In contrast, things whose movements are smoother and predictable generally move at the prompting of an agent. The net result is that from an early age, our conceptual system distinguishes between living things and non-living things on the basis of different modes of interaction with the environment (i.e., acts upon, is acted upon). The primitive notion of *mode of interaction with the environment* is the precursor of the more sophisticated property of function. We adopt this idea of function as a specific manner of interacting with the environment<sup>3</sup>.

The inter-dependencies between functional and perceptual properties are a central aspect of this theoretical framework. For example, wings become associated with flying, legs with walking, eyes with seeing. Shared perceptual properties become associated with shared biological functions; that is, those properties which are true of most or all members of a category. Artefacts, in contrast, have distinctive perceptual features, which become strongly associated with an equally distinctive mode of interaction with the environment (i.e., the artefact's function). This is not necessarily the case for the shared perceptual features, which are not associated with a specific function -- rather they implicate the common characteristics of events or scenarios in which the category of artefacts are used.

In addition to the biological functions of living things and the specific uses of artefacts, there are, of course, many other non-perceptual properties associated with our semantic knowledge of objects, for example that lions generally live in Africa and can be called "The king of the jungle" or that it is illegal to sell knives to children under sixteen. We suggest that these properties should be distinguished from genuine functional properties, although in the literature they have often been run together under the general heading of functional-associative or non-perceptual properties. Such "encyclopedic" or "associative" information is beyond the scope of our current theory, since it is not clear to what extent these properties enter into systematic correlations with other semantic properties. It may be that there is no absolute cut-off in kind between functional and encyclopedic information, but we err on the side of caution in limiting our current model to the clearest examples of biological functions and specific uses of artefacts.

In summary, artefacts tend to have distinguishing perceptual features, which are clearly related to the object's precise function. In contrast, the distinctive perceptual features of living things are not usually closely associated with a specific mode of interaction with the environment (i.e., *tiger-stripes*, *lion-mane*). Rather, it is the perceptual features, which are shared by many similar objects, that become closely associated with an event in which those features interact with the environment - such as eating, moving or breathing. Our claim is that the shared perceptual features of living things are generally closely associated with activities that are characteristic of living things i.e., their biological functions.

#### Correlations

A feature of distributed learning systems is that they discover that the presence of one set of features is correlated with the presence of another. Thus, a system which embodies our assumptions should learn that, for artefacts, distinctive functional information co-occurs with distinctive perceptual features, while for living things biological functional information - which we claim is shared by many members of a category - co-occurs with shared perceptual features. This pattern of co-occurrence will be reflected in the structure of the resulting semantic representations as different patterns of correlated features for concepts in different categories.

These different correlations<sup>4</sup> of semantic features for artefacts and living things have implications for the pattern of loss and preservation of information we expect to see when the system is damaged, on the basis that sets of strongly correlated features tend to be more robust to damage than weakly correlated features. This is because, if one of a pair of features which regularly co-occurs in the environment is lost from the semantic representation, activation contributed by the undamaged feature is often sufficient to reinstate the lost feature. Perceptual and functional features that regularly co-occur are therefore expected to be relatively robust to damage. In our account, functional information varies in distinctiveness across the domains of artefacts and living things, and therefore the types of preserved perceptual feature supported by the presence of functional information will likewise differ. Specifically, distinctive perceptual features for artefacts and shared perceptual features for biological kinds will be correlated with functional information and will be the most resistant to damage.

The theoretical assumptions of our account make a number of predictions concerning the pattern of deficits that we will observe in patients when the semantic system is damaged. However, the predictions are complicated by the fact that patients' performance should be affected by two major factors: severity of damage and task demands.

First, overall, we predict that deficits for living things will frequently occur because their distinctive properties are weakly correlated and so vulnerable to loss. This will lead to poor performance on tasks such as naming and word-picture matching, where distinctive perceptual properties are essential to distinguish between e.g. a lion and a tiger. However, where distinctive information is not crucial for the task (e.g. sorting into categories), patients may well not show a clear disadvantage for living things compared to artefacts since the preservation of shared information will tend to favor living things which have a higher number of more densely inter-correlated shared properties (see Moss & Tyler, 1997; submitted).

Second, we do not predict a uniform loss of visual semantic knowledge for living things. We would expect *shared* visual properties of living things (such as having legs, eyes, mouth etc) to be well preserved due to their mutual correlations and intercorrelations. In this respect our prediction is clearly different to that of the perceptual/functional account, in which living things deficits arise from damage to regions of the brain that are thought to be involved in storing perceptual semantic attributes.

Third, we predict that for artefacts there will not be such an advantage for shared over distinctive properties, since it is distinctive properties that have strong correlations with each other. We have recently reported data for an HSE patient with a living things deficit, which is consistent with both of these claims (Moss et al. 1998).

Fourth, we predict that deficits for artefacts relative to living things should arise when damage to the semantic system is very severe. Our claim is that, in general, patients with severe damage will still have shared features for animals (but not distinctive) because they are strongly correlated, while for artefacts both shared and distinctive features will have been severely damaged. The degree to which this will result in an advantage for living things will depend on the exact *task* demands. For example, in a sorting task it will mean that the patient should do well on living things because they have the shared properties they need for placing something in the right category/domain. In a property knowledge task they should do better for living things to the extent that questions are about shared rather than highly distinctive properties. So, over a whole property verification task where degree of sharedness is not manipulated as a factor (i.e. most tests of this type in the literature) a patient would do better on average on living things. In word-picture matching, having intact shared information would reduce the range of foils that would be incorrectly accepted as the target for living things. Items from a different category or domain would not be chosen because the shared information would be enough to reject them. Only within-category foils would be a problem, because distinctive information would be needed to distinguish them. For artefacts, if not even shared information is available, then even cross-domain foils could not be rejected. [This is something of an idealised account, but the general pattern should hold. The exact balance will depend on the nature of the foils in a given word-picture matching task.] Picture naming requires access to distinctive information for good performance. However, if only shared information is available, then it can get the

patient into the right general semantic space for a response from the correct category. This means that the system may well settle into the pattern for the correct target on some proportion of trials, enabling the correct name to be retrieved. This will give a better response rate for living things than for artefacts where there is no shared information even to get into the correct semantic space. In addition to more correct responses, we would also expect more within-category or superordinate label errors for living things than for artefacts, where errors are less constrained to the category (as we explain in the discussion). We discuss these issues in greater detail in the General Discussion.

In the next section we present a simple simulation to illustrate the principles which, we claim, underpin the structure of semantic representations for living things and artefacts. We then globally 'lesion' the model to explore the resulting patterns of preservation and loss of information of different kinds. We simulate global damage by randomly removing connections among features without restricting the scope of the damage to either living or non-living concepts or to perceptual or functional features.

### THE COMPUTATIONAL MODEL

In the previous section we motivated a number of claims concerning the nature of semantic representations of living and non-living things. We have implemented these claims in a connectionist model in order to investigate the behavior of the model when "lesioned" to simulate non-selective brain damage. An advantage to using a computational modeling approach is that it enables us to directly evaluate our claims by examining the internal semantic representations on a feature by feature basis. With patients we can only infer this information from their performance across a range of tasks.

Specifically, we designed the computational model to incorporate the following representational assumptions:

- 1) We assume that concepts in semantic memory are represented as distributed patterns of activation over a set of units. Artefacts and living things are represented as sets of perceptual and functional features (e.g., Warrington & Shallice, 1984; Farah & McClelland, 1991; Devlin et al., 1998).
- 2) Functional information is always associated/correlated with perceptual information but not necessarily vice-versa.
- 3) The distinctiveness of functional information varies across semantic domains. For living things, biological functions (e.g. *eating, moving*) are shared, whereas artefacts tend to have distinctive functions (*sawing, cutting*). This is supported by an analysis of the Devlin et al. (1998) training set, and by data from our property norms (Tyler, Devlin, Moss & Bunn, submitted; and see Table 1 below).
- 4) As a consequence of assumption (3), the perceptual information that is correlated with functional information will also vary in distinctiveness across semantic domains; for living things, shared perceptual properties are correlated with shared biological functions, whereas for artefacts it is the distinctive perceptual properties that are correlated with distinctive functional properties. The distinctive perceptual properties of living things and the shared perceptual properties of artefacts, in contrast, are not strongly correlated with functional information.
- 5) Properties that are correlated, in that they reliably co-occur in the input to the model, will support each other with mutual activation.
- 6) Living things have more functional properties than artefacts. This is simply because living things tend to engage in a wider range of activities than do artefacts. This assumption is supported by evidence from our property generation norms in which subjects list a greater number of functional properties for each living thing (mean = 7.5) than for each artefact (mean = 5.85;  $t[89]=3.07$ ,  $p<.005$ ). Categories within the living things domain (e.g. *animals, fruits*) tend to fall into well-defined taxonomies. There are two aspects to this claim; one is that many properties are inherited from the superordinate concept, increasing within-category similarity and the other is that there is little overlap among categories within a domain, making categories such as fruit and animals relatively distinct from each other. On the other hand, categories within the artefact domain are more "open-textured" with fewer shared properties and less clearly marked boundaries among categories; For example, there is room for debate as to whether objects belong to one category or another - whether a sledge is a toy or a vehicle, whether a knife is a tool or a weapon and so on (e.g. Atran, 1989).

These theoretical assumptions are supported by the distributional statistics calculated over a set of property generation norms we have recently collected for 93 living and non-living things. Forty five young adult subjects were asked to list all the properties that they believed to be true for a large set of animals, fruit, tools and vehicles. They were given unlimited time to do so. The responses were cleaned according to the criteria used by McRae et al (1997) and Devlin et al (1998). First, quantifiers (e.g., *generally, usually*) were removed. Second, certain key words were used to code the properties; for example: *isa tool* would indicate a category label; *is brown* would indicate an adjective. Third, adjective-noun features, such as *has four wheels* were divided into *has wheels* and *has 4 wheels* on the assumption that the subject has given two pieces of information. Fourth, disjunctive features (e.g., *is green and red*) were also divided up into *is green* and *is red*. Fifth, verb-noun features, such as *eats grass*, were divided into *eats* and *eats grass*. Sixth, some features were nested (e.g., *is used for beef*) and thus were coded both as the specific property (*used for meat-beef*) and the more general property *used for meat*. This was done to capture properties that are shared across items. For example, *used for meat* might be shared among cows, pigs, sheep and chickens, but *used for meat-beef* would be specific to cows.

Each property was coded to identify it as either a perceptual, functional, encyclopedic or category label. A property was coded as perceptual if it was something that could be apprehended by the senses (e.g., what something looks, tastes, smells or feels like). Functional properties were those that denoted how an item interacted with the environment, what it does in the world. For example, *eats* and *walks* are functional properties of animals because they denote what animals do in the world. *Cuts* and *sews* are functional properties of the artefacts *knife* and *needle*. Encyclopedic properties are those that refer to general facts about an item and are neither perceptual nor functional, such as *cows have BSE* and *tigers live in India*. Of the total set of features given by subjects, only 1.8% were encyclopedic properties. These were removed from the analyses. Finally, we included production frequencies for each property, showing how many subjects (out of a possible 45) gave that particular property. We deleted all of those properties which were given by fewer than 5 out of 45 subjects, on the assumption that they were not reliably part of the representation of a given item.

As can be seen from Table 1, our main theoretical assumptions were supported<sup>3</sup>. First, subjects gave significantly more properties overall for living things than artefacts. Second, the properties of artefacts were, overall, more distinctive than those for living things. Third, there were more shared properties given for living things compared to artefacts (whether calculated as raw numbers or as a proportion of the total number of properties). Fourth, the perceptual/functional correlated property pairs (CPPs) for artefacts were more distinctive than those for living things. Fifth, there were more perceptual features given for living things and more functional features given for artefacts. Sixth, as can be seen from Table 2, items within the living things categories were more similar to other members of the same category than were items within the artefact categories, as measured by Hamming distances among the feature vectors for a familiarity-matched subset of 40 items (10 in each of the following categories; animals, fruit, tools, vehicles). In addition, the distance between items *across* categories within the living domain (i.e. between animals and fruit) was greater than across categories within the artefact domain (between vehicles and tools). These two measures support the claim that categories within the living domain, such as animals and fruit, form more stable, distinct hierarchies than do artefact categories like vehicles and tools. These data are described in more detail elsewhere (Moss, Tyler & Devlin, submitted) but for current purposes they illustrate the general agreement of our theoretical assumptions with empirically derived property listings.

	iving	Non-living
Mean number of properties per concept	17.5	11.4 **
Mean distinctiveness of properties <sup>a</sup> 0.029		0.073 **
Mean number shared properties/concept	15.3	8.5 **
Proportion of shared properties/total number of properties <sup>c</sup>	.87	.73 **
Mean distinctiveness of perceptual/functional CPPs <sup>b</sup>	.124	.300 **
Mean number functional properties/concept	7.5	5.8 **
Mean number of perceptual properties/concept	10.3	5.5 **

CPP: Correlated property pairs

\*\* Living/non-living difference significant at p<.005

<sup>a</sup> Distinctiveness calculated as 1/number of concepts for which the property is given.

<sup>b</sup> A significant correlation is defined as one where r>.20, p<.05.

<sup>c</sup> Shared properties are all those that have a distinctiveness value of less than one.

**Table 1 Distributional statistics for property norms for a set of 93 living and non-living things**

	Living	Non-living
Within-category	13.95	16.8
Across –category within domain	40.4	29.2
Across-domain	37.6	37.6

**Table 2 Mean Hamming distance between feature vectors for familiarity matched subset of 40 items from property norm study**

Our theoretical assumptions generate the following set of predictions concerning the behavior of the network when it is lesioned to simulate brain damage:

- 1) Correlated properties will be relatively robust to lesioning because of mutual activation compensating for degraded properties.
- 2) Functional information for both living things and artefacts will tend to be preserved following damage because it is always correlated with perceptual information. In addition, biological functional information is expected to be more robust than the functional information for artefacts because biological functional features are themselves inter-correlated (for example, *moving, eating* and *walking* all occur together for many items).
- 3) The preservation of perceptual properties will vary depending upon whether they are inter-correlated with each other and/or correlated with functional properties. The distinctive perceptual properties of living things are neither inter-correlated nor correlated with functional information and so should be highly susceptible to damage. The distinctive perceptual properties of artefacts are correlated with distinctive functional features and so should be relatively resistant to damage. The shared perceptual properties of living things should also be very robust, since they are correlated with functional properties and are also inter-correlated with each other (e.g. *has a mouth, has legs, has eyes* all frequently co-occur).

### The representations

To investigate our predictions within a connectionist framework we developed a set of vectors encoding the representational assumptions listed above. The vectors were designed to represent members from the domains of artefacts and living things. Our aim was to investigate whether the principles used to design these representations would generate domain-specific deficits (i.e. a selective deficit for either living things or artefacts) when the computational model was randomly damaged<sup>6</sup>. In addition, since our account makes specific claims about the internal structure of concepts, the model should also correctly demonstrate the relative preservation of different types of semantic information within the domains of artefacts and living things as predicted above. The representations had the following characteristics:

- 1) Part of each representation encoded perceptual features and another encoded functional features. Distinctive perceptual features were represented by allowing one feature to be active per category. Shared features were represented by allowing a feature to be active over all items in the category.
- 2) Functional properties always co-occurred with perceptual properties (but not vice versa).
- 3) The distinctive functional and perceptual properties of artefacts were equally distinctive, occurring once per category.
- 4) The shared functional and perceptual properties of living things were shared by all items within a category.
- 5) There were a larger number of shared perceptual features than distinctive perceptual features for both artefacts and living things
- 6) There was greater similarity between items within living things categories (2 bits different) than within the artefact categories (4 bits different). But across categories within domains, the distance was greater for living things (between 8 and 10 bits different) than for artefacts (between 4 and 8 bits different).<sup>7</sup>

Sixteen vectors were created to represent members of two categories of artefacts and two categories of living things. Each vector is intended to represent one category member. For purposes of illustration, the artefact categories can be thought of as tools and vehicles, while the living things categories could be animals and birds, with the individual vectors corresponding to basic level concepts within those categories such as *knife*, *car*, *dog* and *robin*. The vectors capture the notion that living things differ in terms of distinctive perceptual features, whereas artefacts differ in terms of distinctive perceptual and functional information. This was implemented by manipulating the vectors as follows, and as shown in Figure 1.

**Perceptual features:** For both living and non-living things, each category member shared two perceptual features with the three other members of the category (shared perceptual) and had one perceptual feature which was unique (distinctive perceptual).

**Functional features:** For living things, the functional features were shared by all four members of a category, whereas for artefacts, each functional feature occurred only for one vector (i.e. it was distinctive). There were two functional features for each living thing and one for each artefact.

The correlations between functional and perceptual features were represented as follows. For artefacts, a distinctive perceptual feature (i.e. one that occurred once per category) always co-occurred with a distinctive functional feature. For living things, the shared perceptual properties for a category (i.e. those that occurred for all category members) always co-occurred with the same shared functional features. Thus, in both domains, functional features consistently co-occur with perceptual features.

### ARTIFACTS

<i>Distinctive Perceptual</i>	<i>Shared Perceptual</i>	<i>Functional</i>
1 0 0 0	0 0 0 0	1 0 0 0
0 1 0 0	0 0 0 0	0 1 0 0
0 0 1 0	0 0 0 0	0 0 1 0
0 0 0 1	0 0 0 0	0 0 0 1
0 0 0 1	0 1 0 1	0 0 0 1
0 0 1 0	0 1 0 1	0 0 1 0
0 1 0 0	0 1 0 1	0 1 0 0
0 1 0 0	0 1 0 1	0 1 0 0
1 0 0 0	0 1 0 1	1 0 0 0

### LIVING THINGS

<i>Distinctive Perceptual</i>	<i>Shared Perceptual</i>	<i>Functional</i>
0 0 0 0	1 0 0 0	0 0 0 0
0 0 0 0	0 1 0 0	0 0 0 0
0 0 0 0	0 0 1 0	0 0 0 0
0 0 0 0	0 0 0 1	0 0 0 0
0 0 0 0	0 0 0 1	0 0 0 1
0 0 0 0	0 0 1 0	0 0 0 1
0 0 0 0	0 0 1 0	0 0 0 1
0 0 0 0	0 0 1 0	0 0 0 1
0 0 0 0	0 0 1 0	0 0 0 1

Figure 1 : The sixteen vectors used in the simulation representing two categories within each domain.

The crucial variable in our model is the contrast in the distinctiveness of correlated information for artefacts and living things. We quantify this difference by using the Devlin et al. (1998) measure of distinctiveness (1/number of times a feature is active) on our vectors. This confirms that artefact distinctive perceptual features and associated functional information are twice as distinctive (0.5) as the shared perceptual features for living things and the associated biological functional information (0.25). Although the features in our model are not based on real semantic properties, we can illustrate the distinctions in a more concrete way using hypothetical examples of features of each kind. For example, the feature *mouth* is shared by all animals and it is correlated with its specific function of *eating* ( $r = 1$ ), which is also shared by all members of the category. Thus, the perceptual feature *mouth* is on 4 times out of 16, as is the correlated functional feature *eating*, giving a value of 0.25. In contrast, for artefacts, distinctive perceptual and functional features are fully correlated ( $r = 1$ ). For example, the distinctive perceptual feature *serrated edge* is correlated with the distinctive function of *sawing*; each feature only occurs once per category. These features are on twice and off the remaining 14 times, giving a value of 0.5. Thus, the distinctive features are twice as informative (in terms of reliably distinguishing between items within a category) as the shared features. Distinctive perceptual features for living things are similarly distinctive (0.5) but do not consistently co-occur with other features.<sup>3</sup> In summary, the two vector sets can be distinguished in terms of the number and distinctiveness of strongly correlated features. The artefact vector sets have four strongly correlated and highly distinctive features and a further four strongly correlated shared features. The vector representations of living things have eight strongly correlated features, which are shared across a category.

As will be apparent from the above description, we made a number of simplifying assumptions in order to keep the model simple and tractable:

1. We have not taken into account other variables that have been implicated in category-specific deficits, such as concept familiarity (e.g. Funnell & Sheridan, 1992). In the current model, all vectors are encountered an equal number of times during training, simulating a familiarity-matched set of living and non-living items.
2. The two domains were represented with two sets of mutually exclusive features. We decided to implement the vectors in this way, since most features associated with one domain do not occur for the other. For example, the biological functional properties of living things (*eating*, *seeing*, *breeding*) do not occur for artefacts, while most properties of artefacts are not applicable to living things (e.g. *made of metal*, *used for cutting*, *sawing*, *drilling*). Nevertheless, we acknowledge that this is a simplification and

that in reality there are properties that occur over both domains (e.g. *moves, has legs, is black*) although it is not clear that these are really identical semantic properties over the two domains, in spite of the same words being used to describe them (e.g. the legs of a chair or a table may not be represented by the same property as the legs of animals). Moreover, on the basis of our property norms, we can confirm that such properties represent only a very small proportion of the total properties that subjects provide; only 5% of the properties that subjects produced consisted of properties which were true of concepts within the domains of both living things and artefacts.

3. Our vectors do not include any semantic properties that are shared across a whole domain, such as across all living things or all artefacts. In this respect it captures real world domain structure, where there are very few properties that are true of all concepts within a domain. Consider, for example, the two living things categories of animals and fruits/vegetables, where very few properties are true of the concepts within both categories. Shared information in our model refers to information that is common to members of a category within a domain, such as birds, animals, or tools. The main reason for excluding domain-general properties was to avoid increasing the size of the model, but also because such properties are very abstract (e.g. a domain-general property of all living things might be *is organic* or *has a genetic structure*) and were never given in our property norms. Such high-level information may be better characterized as part of our background theoretical knowledge of domains, rather than as featural information than is represented within individual concepts (e.g. Keil, 1994).
4. It is an intrinsic aspect of our model that the distinctiveness of a feature is inversely related to its frequency of occurrence. For example, a shared property that occurs for all four members of a category will occur four times as often in the training set as a distinctive property. This will lead to higher baseline activation for shared than distinctive properties. It would be possible to vary the effect of frequency and distinctiveness independently in a larger-scale model, by manipulating the frequency of presentation of individual vectors in the training set, but in the current simulations we hold vector frequency constant. We do not believe that the frequency/distinctiveness confound is a problem for our claims about the interaction of distinctiveness and intercorrelation, since the greater frequency of occurrence of shared properties emerges as an interpretable aspect of the model's performance as a function of increasing lesion-severity, as will be discussed further in the following sections.
5. The most obvious simplification is the small size of our computational model, both in terms of the number of vectors and the number of features within each vector.

The current simulation is intended as a straightforward, idealized illustration of the way in which the major theoretical assumptions of our account would interact to produce patterns of preserved and impaired performance when "lesioned". Our simulation clearly does not reflect the complexities of semantic representations, but is designed to capture some of the basic distinctions that we believe are important aspects of the structure of concepts in the living and non-living domains. Even though it is a simplification, it clearly captures some of the essential properties of domain differences, as is shown by our property norm data. The aim is to determine whether the interplay of factors such as correlational strength and distinctiveness operates in line with our basic predictions, before introducing the myriad of additional factors and variability that would be inherent in more realistic semantic representations. The empirical issue of how such factors will be affected when the computational model is scaled up remains open to future investigation, but we believe that the advantages of tractability and transparency of a small-scale simulation have their own value in generating and testing predictions, as is illustrated, for example, by Dell's interactive activation model of speech production (Dell, 1986; Martin, Dell, Saffran & Schwartz, 1994).

In spite of these points, we believe that the representational assumptions we have implemented in our semantic vectors are essentially correct, as they are supported by the distributional statistics calculated over the relatively large set of property norms we have collected.

#### Architecture:

We used the network architecture shown in Figure 2 which consisted of three layers; an input layer consisting of 24 units fully connected to a hidden layer of 20 units which was in turn fully connected to an output layer of 24 units. Input to the model was a pattern of activation corresponding to one of the sixteen vectors.

#### Training and testing:

The model was trained using the back-propagation learning algorithm (Rumelhart, Hinton & McClelland, 1986). The network was simply required to reproduce the input on the output layer. Three hundred networks were trained with different initial random weights with a learning rate of 0.25 and momentum 0.9. Training was stopped after a mean of 984 presentations of the complete target set when the summed squared error for each vector was about 0.01.

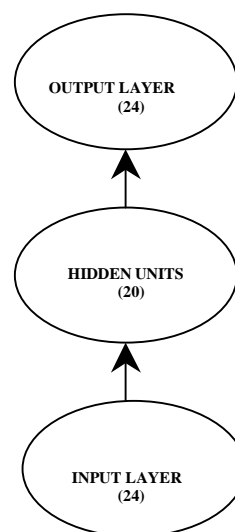


Figure 2: Architecture of the network

Following training we damaged the network to simulate the effects of brain damage in the model. The training set was then presented to the network's input layer and the pattern of activations on the output layer was examined. Because our account makes specific predictions concerning the effect of damage on particular types of features, we first evaluated the networks' performance at a featural level. A second analysis examined the model's performance at a more global level to determine whether category specific deficits resulted from the featural impairments.

- 1) *Featural analysis.* We have claimed that features which receive collateral support from correlated properties will be more resistant to damage than those which do not participate in correlations. Specifically, the more correlations a feature is involved in, the more robust it will be. Thus, we predict that the distinctive perceptual properties of living things (which do not participate in any correlations) will initially be the most susceptible to damage. Next, the distinctive perceptual and functional features of artefacts, i.e., their form-function correlates, will be affected by damage. Ultimately, the shared properties of living things (both perceptual and functional) will be the last to be impaired by a random, progressive damage. To evaluate these predictions an error score was calculated per feature as the difference between its target and output values. We examined each type of feature (distinctive perceptual, distinctive functional, shared perceptual, and shared functional) in both domains (living and non-living) and compared the extent to which each was preserved or impaired after damage.

- 2) *Overall analysis.* To examine the model's overall performance on an item, the pattern of activity over the output units was compared to all sixteen training patterns. The closest match, by a Euclidean distance measure, was considered the model's response. Before damage, the closest match was always the correct response but after damage errors began to occur. Each response was classified as either correct or as one of three types of error: a within-category error, a within-domain error, or a cross-domain error. Thus this task roughly simulates patient performance on a forced-choice task such as word-picture matching. Consequently, the model's behavior could be evaluated for evidence of category specific deficits over the damage progression.

## LESIONING

In this section we see how the model performs when it is globally damaged, thus testing the strong claim that category-specific deficits can arise from random damage to a distributed system. We randomly lesioned the network connections and then examined the patterns of preserved information of each vector type to determine the relative robustness of each type of feature to global damage. We lesioned the model by randomly setting a proportion (initially 10%) of all connection weights to zero. After each lesioning, the network's performance was analyzed. The proportion of damaged connections was increased by increments of 10% until all inter-layer connections were set to 0.

We then determined the extent to which each type of semantic information was affected by lesioning by examining in turn the units corresponding to distinctive perceptual, shared perceptual and functional features. We used the error (absolute difference between target and output unit activation) as a measure of the robustness of that feature to damage. Large error values for a feature unit compared with other units indicate a greater susceptibility to damage. Over many trials, consistent patterns of error highlight units that are particularly susceptible to damage.

Each of the three hundred simulations were lesioned at eight levels of lesioning severity. 10% to 80% connections randomly removed from both sets of connections -- those between the input and hidden units and those between the hidden and output units. Note that unit biases were not affected by this damage. The absolute error for each output unit was recorded following presentation of a vector from the training set. The errors were then summed for each level of lesioning severity and an average error calculated for each set of four feature units representing distinctive perceptual information, shared perceptual information and functional information.

## RESULTS

### 1. Feature analysis:

One of the strengths of our approach is that we are able to look at the micro-structure of conceptual knowledge by examining patterns of preserved and impaired features as a function of feature type and severity of damage. In this respect our modelling represents an advance over more traditional approaches to semantic deficits which tend to focus on more global properties of the damaged semantic system

#### 1.1 The preservation of correlated properties

We predicted that correlated features should be more robust to global damage than features not supported by correlations, since the system is able to determine the correct activation value for the damaged feature on the basis of evidence contributed by the other feature. Various aspects of the results support this prediction.

First, we examined the relative preservation of distinctive properties, comparing those that were correlated (perceptual distinctive properties of artefacts) with those that were not (perceptual distinctive properties of living things). A repeated measures analysis of variance was performed on the absolute feature error for each type of feature for each network at each level of lesioning severity. We found more errors associated with distinctive perceptual features of living things than artefacts [ $F(1,299)=24.76, p<0.001$ ], and a significant interaction between type of feature and severity ( $F[8,2392]=70.31, p<0.001$ ). Figure 3 shows that until 60% lesioning severity, distinctive perceptual

features of artefacts have a lower error and thus are better preserved than those of living things. After 60% lesioning, there is no difference between the two types of feature because both are too damaged to contribute reliable evidence to the network, canceling out the advantage that would otherwise hold for correlated features.

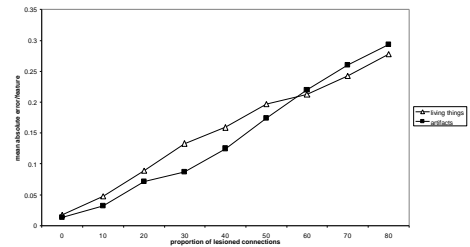


Figure 3 : Mean absolute error for distinctive perceptual features of artefacts and living things at different levels of lesioning severity over 300 simulations.

Second, we examined the perceptual properties of living things, comparing distinctive (non-correlated) with shared properties (correlated), and found an advantage for shared properties up until 50% of connections had been lesioned ( $F[1,299]=23.01, p<0.001$ ; see Fig 4). Thus, all of the comparisons show the same pattern -- correlated features are generally better preserved than non-correlated features<sup>9</sup> -- until quite severe levels of lesioning, where damage is so severe that differences are no longer significant (see footnotes 7 and 9).

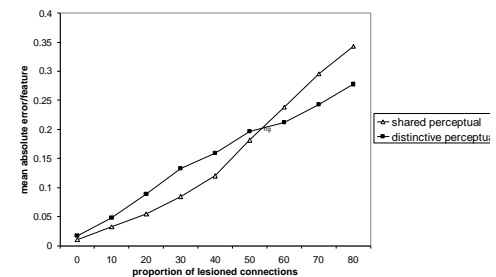


Figure 4: Mean absolute error for shared and distinctive perceptual features for living things at different levels of lesioning severity averaged over all 300 simulations.

#### 1.2. The preservation of perceptual properties

We predicted that the distinctive perceptual features of living things would be less well preserved than shared perceptual properties, because distinctive properties are not correlated. This was supported by the data (see Figure 4). In addition, we expected better preservation of the shared perceptual features of living things compared to artefacts because living things have more correlated features than artefacts. As Figure 5 shows, this was indeed the case. The shared properties of living things were better preserved than those of artefacts ( $F(1,299)=934.09, p<0.001$ ).

In the network this pattern occurs because artefact shared perceptual features are inter-correlated with other artefact shared perceptual features but not with functional features, whereas shared perceptual features for living things are correlated with other shared perceptual features and also with functional features. The more inter-correlations there are, the less evidence the network needs to determine the correct activation values for a unit; thus the shared features of living things should be better preserved.



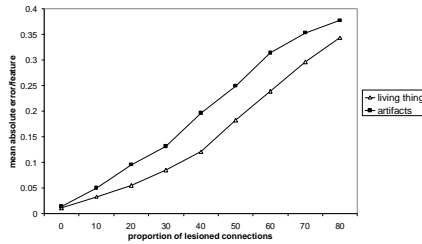


Figure 5: Mean absolute error for shared perceptual features for artefacts and living things at different levels of lesioning severity averaged over all 300 simulations.

### 1.3. The preservation of functional properties

A second important prediction of our account is that functional features should be well preserved because they are always correlated with perceptual features, although the reverse is not the case. To evaluate this claim we compared the preservation of functional and perceptual features after lesioning. Functional features vary across domains in one important respect; the functional features of living things are always shared whereas the functional features of artefacts are always distinctive. Therefore, to determine the relative robustness of functional compared to perceptual features, while holding the distinctiveness of those features constant, we compared the functional properties of artefacts (which are distinctive) with the distinctive perceptual properties of living things, and found a higher error rate for distinctive perceptual features than distinctive functional features [F(1,299)=24.76,  $p < 0.001$ ]. This is the same analysis as in Figure 3, since the distinctive perceptual and functional properties of artefacts are correlated one to one

We also looked at how damage to the network affects functional features according to their degree of distinctiveness, by comparing the functional properties of living things and artefacts (which are completely shared and completely distinctive respectively in our network). Functional properties in both domains should be relatively robust to damage because, as we have claimed, functional features are consistently correlated with the presence of perceptual features. The data shown in Figure 6 reveal an interaction between functional property distinctiveness and lesioning severity (F [8,2392]=387,  $p < 0.001$ ). The functional features of artefacts are initially more robust than those of living things, but at more severe levels of lesioning it is the shared functional information for living things that is consistently and significantly better preserved.

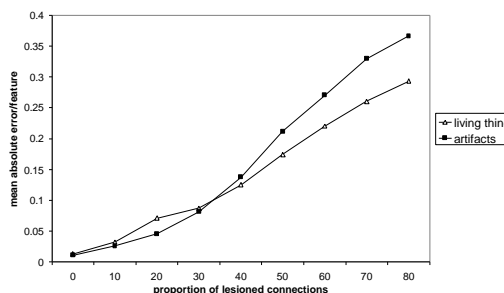


Figure 6: Mean absolute error for functional feature units for artefacts and living things at different levels of lesioning severity averaged over all 300 simulations.

## 2. Identity mapping

We now turn to the question of how the relative loss and preservation of different feature types (as charted in the analyses above) will be reflected in the global performance of the network in identifying items in the living and non-living domains. As discussed earlier, this is more akin to the kind of data that are available from neuropsychological studies of patients completing semantic memory tests than is the detailed

feature by feature analysis, although it is not a simulation of specific tasks since this would involve additional assumptions.. We examined the ways in which the pattern of preserved feature information affected the damaged networks' ability to correctly reproduce the input on the output layer. We call this *identity mapping*. It can be thought of as an approximation to a word-picture matching experiment for a patient; the input is presented (e.g. a word) and the patient has to select the correct output from an array, which includes both within and between category distracters as well as the correct picture. We examined the pattern on the output layer to see whether it was most similar to the target or to some other item, as measured by Euclidean distance, to measure the network's accuracy in this task. The network's pattern of activity can be one of three types:

- (i) a correct match between input and output layer; i.e. it correctly identifies an object.
- (ii) a within-category error, where the best match is a member of the same category (e.g. misidentifying a *dog* for a *horse*).
- (iii) a between-category error, with the best match being an item from a different category but the same domain (e.g. misidentifying a *banana* as a *horse*).
- (iv) a between-domain error, with the best match being an item from a different domain (e.g. misidentifying a *fork* as a *horse*).

The pattern of loss and preservation of features that were revealed in our analyses reported above should give rise to the following pattern of errors in the network.

- (1) The damaged network should perform badly at identifying living things because this requires discrimination between items on the basis of distinctive perceptual information (e.g. *stripes* vs. *spots* distinguishes between a tiger and a leopard). This will be difficult for the net since only shared biological functional properties with their correlated shared perceptual attributes are preserved for living things, and not the individual perceptual features that differentiate between category members.
- (2) The network should perform better at identifying artefacts, in which distinctive form and function are most strongly correlated, and so will be preserved well enough to support discrimination among category members.
- (3) Although the network should perform badly at identifying individual living things, knowledge of category membership should be relatively well-preserved because shared properties (typical of most members of a category) will be preserved given their correlations with biological functional information. Therefore we expect that the network will make a high proportion of within-category errors for living things, where the pattern on the output layer closely matches a member of the same category as the target item (i.e. mistaking *dog* for *horse*).
- (4) In contrast, knowledge of artefact categories should be less well preserved since, for artefacts, shared properties are not supported by correlations with functional information and thus shared information should be more vulnerable to damage. This suggests that the network will make between-category errors in addition to within-category errors.

### a) Correct identity mappings:

To test these predictions, in each of the three hundred simulations, the network was lesioned in the same incremental proportions as described above and presented with the training set. The lesioned network's output for each item was compared on a vector by vector basis with the complete target set to establish the closest neighbor. The number of correct mappings was recorded. A correct mapping in this context means that the lesioned network reproduced a pattern of activation whose nearest neighbor was the target.

Figure 7 shows how lesioning affects the model's mean performance over 300 simulations. As we predicted, more correct identifications were made for artefacts than for living things [F(1,299)=35,  $p < 0.001$ ] until about 60% of the connections had been lesioned. At more severe levels of damage, performance on living things was more accurate. Post-hoc Scheffe tests showed that correct performance for artefacts and living things differed

significantly ( $p < .001$ ) at all levels of lesioning bar 10%. This models the double dissociation between knowledge for artefacts and living things reported widely in the neuropsychological literature (Warrington & McCarthy, 1983; 1987; Warrington & Shallice, 1984; Saffran & Schwartz, 1992).

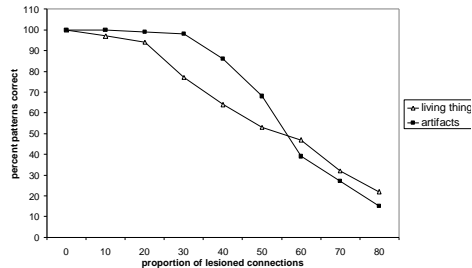


Figure 7: Correct identity mapping for living things and artefacts averaged over all 300 simulations.

In our network, significant numbers of artefact errors are only made when there is severe damage. This contrasts with the Devlin et al. (1998) in which 76% of the simulations in the model showed worse performance on artefacts with mild damage that crossed over into a significant impairment for living things as the damage increased. The initial artefact deficit occurs in the Devlin et al. model because there were fewer inter-correlated artefact features (compared with the number of inter-correlated features for living things) to provide compensatory activation when individual distinctive features were lost. Our simulations suggest an alternative account where artefact deficits occur when damage is severe enough to destroy artefact shared perceptual features, which are not correlated with specific functional information. This results in between-category errors becoming increasingly frequent. A deficit for artefacts corresponds to the complete loss of semantic information for artefacts (both perceptual and functional) whereas shared features (shared perceptual features and functional features) for living things may be, by comparison, relatively intact.

#### b) Types of errors:

The mapping errors made by the damaged networks averaged over 300 simulations were categorized as within-category or between-category errors. Figure 8 plots the type of error made against lesioning severity. The network consistently makes more within-category errors for living things than for artefacts ( $F[1,299]=447.92, p < .001$ ). This reflects the difference in robustness of distinctive perceptual features for artefacts (which are correlated) and living things (which are not correlated). In contrast, between-category errors for artefacts occur with mild levels of lesioning, earlier than the first between-category errors for living things. This is expected since the shared perceptual features for living things which provide category information are supported by correlations with functional features and are relatively robust to damage. In summary, mapping errors for living things tend to be members of the same category until the most severe levels of damage because shared category information (i.e. shared perceptual and functional features) is so robust to damage. Artefacts generate between-category errors as well because the shared category-defining features are lost first.

In addition to the patterns of correlation over features, the semantic neighbourhood of concepts in the different domains also plays a part in determining the probability of errors of different types. For living things, within-category errors are likely because concepts within these categories are close together; nearest neighbours are only two bits different in our model. Although this does not cause any difficulty for the intact model (with 100% correct mappings, see Table 7), it is predicted to increase the probability of within-category errors when the vulnerable non-correlated distinctive properties have been lost. For artefacts, on the other hand, the probability of making between-category errors is increased by the fact that some concepts in different artefact categories are as close to each other as those within categories (four bits different). Again, this does not lead to difficulty in discriminating among artefact concepts in the intact model (100% correct mappings, see Table 7), but it is predicted to have an effect

when damage has eroded the shared properties in the ways reported in the previous sections. Thus, although the distance between within and between-category neighbours differs across the living and non-living domains (in a theoretically motivated way) these differences only have an impact on the performance of the model when it has undergone the pattern of damage predicted as a function of feature distinctiveness and intercorrelation.

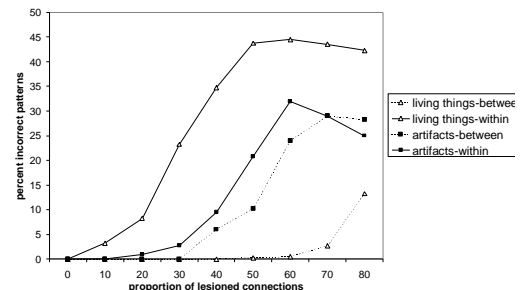


Figure 8: Between and within-category errors for living things and artefacts averaged over all 300 simulations.

## DISCUSSION

This research has shown that category specific impairments can result from damage to a conceptual system in which there is no explicit category structure. Moreover, we see that the patterns of impairment result from a complex interaction of type of semantic feature, correlations between features and the extent to which features are shared or distinctive. Thus, the modelling data provide a demonstration that category-specific impairments can emerge from a damaged learning system, which has encoded the functional significance of everything it encounters. Global damage, where both perceptual and functional feature connections were randomly lesioned, produced an initial impairment for living things followed, with more severe lesioning, by an impairment for artefacts. At mild to moderate levels of damage, the patterns of preservation and loss in the computational model predict a behavioral deficit characterized by difficulty in discriminating between living things within a category (e.g. *horse*, *cow*) but relatively intact knowledge about shared information - both perceptual and functional. Artefact functional knowledge should also be relatively preserved, enabling artefacts to be distinguished from each other and their function retrieved, but knowledge of artefact category membership (e.g. *tools*, *vehicles*) should be relatively impaired (e.g. sorting artefact category members into categories on the basis of shared features). A patient presenting with this pattern of deficit would be able to correctly identify the specific function of a *kettle* on the basis of its perceptual features, yet be unable to place it in the category of *kitchen utensils* or appreciate its relationship to other members of the same category such as *saucapans* and *toasters*. Conversely, though unable to recognize a specific living thing and distinguish it from other members of the same category, the patient would be able to retrieve relevant shared perceptual and functional information that place it in the correct category (e.g. *animals* or *fruit*).

How do these predictions correspond to the actual behavior of patients with category-specific deficits? Unfortunately it is impossible to assess this issue for most patients in the literature, since the crucial comparisons between knowledge of distinctive and shared information for different categories have not been reported; an overall paucity of knowledge about items in one or other category, or even a discrepancy between perceptual and functional information cannot be evaluated against our model, since the critical variables of distinctiveness and intercorrelation of properties are not manipulated.

However, we have recently been able to test our predictions with an HSE patient, RC (Moss et al, 1996; Tyler & Moss, 1997; Moss, Tyler, Durrant-Peatfield & Bunn, 1998), showing the typical pattern of a marked category-specific deficit for living things. Across a range of tests designed to probe semantic knowledge, RC had difficulties in discriminating between living things on the basis of their distinctive properties, but was good at grouping them together on the basis of their similarities. In contrast, he had no difficulty in discriminating between artefacts. This was demonstrated, for example, in a picture sorting task

where RC was able to sort pictures of living things according to category (i.e., identify their shared properties). In fact, he scored better for living things than artefacts on this test. In contrast he had considerable difficulty when asked to sort the pictures of living things according to more distinctive properties of category members (e.g. fierce vs. non-fierce animals), with significantly poorer scores than for artefacts (e.g. electrical vs. non-electrical appliances). Similarly, in a property verification task, RC was unable to verify the distinctive perceptual properties of living things but had little difficulty with shared properties. For artefacts, he performed as well for distinctive as for shared properties. Like a number of other patients with a deficit for living things, RC did not show a disproportionate loss of visual compared to functional information, when the distinctiveness of properties was held constant, suggesting that damage to a neuroanatomical region specialised for the representation of visual/perceptual information is not an adequate account.

A second novel prediction of our account is that it predicts an interaction between the degree of damage to the semantic system and the direction of category-specific impairment. This was confirmed in our computational simulation, which showed a selective deficit for living things until approximately 60% of the connections were lesioned. At more severe levels of lesioning the network produced an artefact deficit, thus reproducing the double dissociation reported in the literature (Warrington & McCarthy, 1983, 1987). This pattern implies that we should only observe artefact deficits in patients with profound semantic impairments. This seems to have been the case for the two classic patients reported by Warrington & McCarthy. However, it is difficult to evaluate overall severity and extent of brain lesion in patients recovering from HSE or CVA since comparable tests and data are rarely reported. Several reports have suggested that there may be category-specific impairments in patients with DAT (e.g. Silveri et al. 1991) although data from group studies are mixed, and no clear interaction of the direction of category-specific deficits with severity of semantic disorder has emerged (Garrard et al., 1998; Gonnerman et al., 1997). Again, this may be due to the difficulty of identifying the appropriate criteria of severity, particularly for patients with more wide-ranging cognitive impairments.

A more promising approach is to carry out longitudinal studies of individual patients with progressive semantic disorders, which enables us to track the nature of category-specific effects over the course of the disorder. Longitudinal studies of two DAT patients are reported by Gonnerman et al., (1997). One patient showed a consistent deficit for living things, while the other showed a deficit for artefacts, although this was not significant at most of the testing sessions. It is possible that the former patients would have progressed to show the predicted artefact deficit, but it was not possible to carry out further testing, presumably due to the more widespread cognitive deficits associated with DAT. The pattern for the latter patient is unclear.

In a recent study we were able to investigate the nature of semantic impairments during the progression of the disorder for a patient with a generalized cerebral atrophy, but whose cognitive functions in other domains were less compromised than those of DAT patients at comparable levels of semantic impairment. ES showed a marked deficit for artefacts only at a late stage of the disorder, as predicted by our model. Earlier in the disease she either showed no difference between living things and artefacts, or in some tasks, a deficit for living things (Moss & Tyler, 1997; Moss & Tyler, In Press). A similar pattern has also begun to emerge for a second progressive aphasic patient, AA, with a similar progressive decline in semantic knowledge (Moss, Tyler & Devlin, 1999). Neither patient showed a significantly greater overall deficit for functional information compared to visual information, questioning the claim of the perceptual/functional weighting hypothesis that deficits for artefacts arise as a result of damage to regions of the brain involved in representation of functional information.

According to our account, the discrimination of artefacts is relatively spared because their distinctive perceptual features are supported by correlations with functional features. The discrimination of living things is impaired because their distinctive perceptual features are *not* supported by correlations with other features. Functional information for both domains is relatively resistant to damage because it is supported by correlations with perceptual features. The data therefore model not only the well-documented dissociation between knowledge for artefacts and for living things, but also more recent neuropsychological evidence concerning the robustness of functional

information (Moss, Tyler, Hodges & Patterson, 1995; Tyler & Moss, 1997; 1998). For example, Tyler & Moss (1998) report a patient who has impaired perceptual knowledge about living things yet relatively well preserved functional information. AM, a patient with fluent progressive aphasia, was tested on a semantic priming task three times over a 2-year period. AM initially showed priming only for functional and visual properties, then only for functional properties, finally showing no priming at all. As AM's semantic representations gradually deteriorated, the functional properties of concepts remained accessible well after other aspects of meaning were no longer available<sup>10</sup>.

A final prediction of our account, which was confirmed by the performance of the computational model, is the occurrence of between-category errors for artefacts but not for living things. This type of error would be, for example, identifying a tool as a vehicle. Our claim is not that cross-category errors are necessarily very common for artefacts, but rather that they will be more likely than for living things, for which between-category errors should hardly ever occur, even in the most severe deficits. There is some evidence for this pattern in longitudinal studies of picture naming performance. Hodges, Graham & Patterson (1995) report a detailed breakdown of the naming errors of a semantic dementia patient, JL, over a period of 18 months. For living things, JL makes progressively more category-coordinate and superordinate errors, but he never produces a name that crosses categories (e.g. *animal/fruit*) or domains. However, for artefacts, occasional cross-category (e.g. *paintbrush – piece of vehicle*) and even cross-domain (*telephone-animal*) errors emerge at the later time slices. We have carried out a similar analysis of the naming errors for one of the progressive aphasic patients mentioned above (Moss, Tyler & Devlin, 1999). AA was presented with a set of color pictures of living and non-living things, matched over domains for visual complexity and familiarity (see Bunn, Tyler & Moss, 1998, for details of the picture set) on four occasions over a period of almost two years. AA never produced a single cross-category or cross-domain error for living things until the last test session when there were two such responses out of 69 items (*rabbit-the boys; cauliflower-camel*). For artefacts, in contrast, occasional cross-category errors (e.g. *roller-skate-shoe*) and cross-domain errors (e.g. *skittles-orange*) were produced even at the first test session when she was already significantly impaired, and by the last two sessions there were eight and six such errors respectively, including such striking examples as *pram-banana, table-umbrella and book-candle*.<sup>11</sup> Although the absolute numbers of cross-category and cross-domain errors are small, there appears to be a consistent tendency for such errors to be produced in greater numbers for artefacts than living things, and to increase over time, as predicted by our account.

## CONCLUSIONS

This research has addressed the question of how conceptual knowledge is represented and structured in such a way that brain damage can lead to specific patterns of preserved and impaired knowledge. Do these deficits indicate the discrete localization of different types of semantic information, or can they be accounted for by more diffuse and distributed forms of neural representation? Our key claim was that concepts in different semantic categories differ in structure in such a way that damage to the system will affect them in different ways, producing apparently selective deficits. The modelling data we have presented support this claim. Moreover, there is a close correspondence between patient and simulation data -- we have shown how a damaged learning system trained on representations which encode our assumptions can mimic category-specific deficits following damage to the brain.

These converging lines of evidence allow us to speculate that the mental representation of objects in the world are indeed organized along similar lines as in the model. The function of an item in the world is closely correlated with its perceptual attributes. Our account therefore considers the human cognitive apparatus as a system adapted to encoding the functional significance and statistical regularity of everything it perceives. Statistical properties of the perceptual environment produce different patterns of distinctiveness and relatedness of features in the system's internal representations. These produce category specific deficit behavior when damaged. Although the model described here represents a simplification of our assumptions underlying category structure because of its small scale, in future research we intend to use more psychologically realistic representations in a scaled up model. This will enable more extensive investigations of the basic findings reported here.

<sup>1</sup> Although JJ is often referred to as having a deficit for artefacts, in fact the pattern was more complicated, with a selective preservation of animals and vehicles relative to other categories. Thus, the impairment crosses the living and non-living domain and is difficult to accommodate in terms of selective-damage to a category-specific semantic store.

<sup>2</sup> Although similar in some ways, our claim about the correlation between the form and function of artefacts is not reducible to the Gibsonian notion of affordances. On our model, it is not simply the case that inferences about the likely function of an object are constrained to some degree by its perceptual attributes. We claim that correspondences between perceptual and functional properties are learned and stored as part of the conceptual representation of that object, with those correspondences instantiated by links which support activation not only from form to function, but also in the reverse direction.

<sup>3</sup> When Mandler (1992) notes that the property of being acted on is the primitive precursor of the more sophisticated notion of function she is referring to artefacts. Our account does not distinguish between being acted on and acting on, and applies to both artefacts and living things.

<sup>4</sup> We use the term "correlated" to refer to correlations between different types of semantic features (i.e. functional and perceptual) and "intercorrelated" to indicate correlations between the same type of semantic features (e.g. shared perceptual features).

<sup>5</sup> A frequency-matched subset of 20 living things and 20 artefact concepts showed essentially the same pattern of results.

<sup>6</sup> Up to this point we have referred to selective impairments for either living things or artefacts as category-specific deficits, because this is how they are usually described in the literature. However, from this point we will refer to artefacts and living things as semantic domains and reserve the term category to refer to superordinate categories within those domains (e.g. mammals and birds in the living domain, vehicles and clothes in the artefact domain). Strictly speaking, then, our model is an account of domain-specific rather than category-specific deficits.

<sup>7</sup> For artefacts, this means that the minimum Hamming distance for two items in different categories was the same as two items within the same category. However, it was never the case that there was an item in the other category that was actually *closer* than one in the same category.

<sup>8</sup> Distinctive features for artefacts and living things are both weakly correlated with the shared or category-specific information ( $r = 0.218$ ).

<sup>9</sup> The picture is not quite as straightforward as this since correlation interacts with other variables such as the frequency of the correlation. We illustrate this point by comparing the distinctive properties of artefacts, which are uniquely correlated, with shared perceptual properties, which are inter-correlated with themselves. Distinctive correlated properties are significantly better preserved than shared properties ( $F [1,299]=1084.8, p < .001$ ). This is because the damaged network tends to over-generalize and becomes increasingly unable to accurately determine whether a feature should be active or not. It compromises by keeping the units similarly activated, where the activation of a unit is proportional to how many times the unit was on during training. When a feature should not be active, the difference in level of activation between a feature that should be on and one that should be off diminishes. The activations on the output layer therefore become increasingly undifferentiated as damage to the network accumulates, with the result that discrimination between shared features becomes increasingly error prone. The problem of differentiating between distinctive features is less severe because distinctive features were active fewer times during the training phase. This has the effect of making distinctive correlated information more robust than shared correlated information.

<sup>10</sup> Our account predicts that shared perceptual features for living things will also be preserved because of correlational support from functional properties. However, the materials used to test for visual properties included colors, textures, shape and size. These are distinctive perceptual attributes, which we suggested earlier are not necessarily associated with biological functional features, and should therefore not be robust to lesioning.

<sup>11</sup> We have not been able to analyze the data for ES in the same way as her naming errors were almost always circumlocutions or no responses, rather than an incorrect item name.

## REFERENCES

- Atran, S. (1989). Basic conceptual domains. *Mind and Language*, 4, 7-16.
- Barton, M.E. & Komatsu, L. Defining features of natural kinds and artefacts. *Journal of Psycholinguistic Research*, 1989, 18 (5), 433-447.
- Basso, A., Capitani, E. & Laiacona, M. Progressive language impairment without dementia: a case with isolated category-specific semantic deficit. *Journal Of Neurology, Neurosurgery & Neuropsychiatry*, 1988, 51, 1201-1207.
- Bunn, E.M., Tyler, L.K., & Moss, H.E. Naming pictures: The role of familiarity and property type re-examined. *Neuropsychology*, 1998, 12, 367-379.
- Caramazza, A., Hillis, A. E., Rapp, B. C. & Romani, C. The multiple semantics hypothesis: multiple confusions? *Cognitive Neuropsychology*, 1990, 7, 161-189.
- Caramazza, A. & Shelton, R. Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, 1998, 10, 1-34.
- Damasio, H., Grabowski, T.J. , Tranel, D, Hichwa, R.D. & Damasio, A.R. A neural basis for lexical retrieval. *Nature*, 1996, 380, 499-505.
- Dell, G. A spreading activation theory of retrieval in sentence production. *Psychological Review*, 1986, 93, 283-321.
- 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, 77-94.
- De Renzi, E. & Lucchelli, F. Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex*, 1994, 30, 3-25.
- Durrant-Peatfield, M., Tyler, L. K., Moss, H. E & Levy, J. The distinctiveness of form and function in category structure: A connectionist account. In: *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, 1997, Mahwah NJ: Erlbaum,
- Farah, M. J. & McClelland, J. L. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 1991, 120(4), 339-357.
- Farah, M.J, Hammond, K.M., Mehta, Z., & Ratcliff, G. (1989). Category-specificity and modality-specificity in semantic memory. *Neuropsychologia*, 27, 193-200
- Farah, M. J., McMullen, P.A. & Meyer, M. M. Can recognition of living things be selectively impaired? *Neuropsychologia*, 1991, 29, 185-193.
- Gainotti, G. Silveri, M. C., Daniele, A. Giustolisi, L. How Luria and Vinogradova's theory on the dynamics of the semantic systems could explain an unusual case of lexical comprehension disorders. Special Issue for A. R. Luria. *Aphasiology*, 1995, 9(2), 153-161.
- Garrard, P., Patterson, K., Watson, P.C. & Hodges, J.R. Category-specific semantic loss in dementia of the Alzheimer's type. *Brain*, 1998, 121, 633-646.
- Gonnerman, L. M., Andersen, A. S., Devlin, J. T., Kempler, D. & Seidenberg, M. Double dissociation of semantic categories in Alzheimer's disease. *Brain and Language*, 1997, 57, 254-279
- Hart, J. & Gordon, B. Neural subsystems for object knowledge. *Nature*, 1992, 359, 60-64.
- Hillis, A. E. & Caramazza, A. Category-specific naming and comprehension impairment: A double dissociation. *Brain*, 1991, 114, 2081-2094.

- Hinton, G., & Sejnowski, T. Learning and relearning in Boltzman machines. In D. Rumlehart & J. McClelland (Eds.), *Parallel Distributed Processing*, 1986, (pp 282-317). Cambridge, MA: MIT Press.
- Hodges, J., Graham, N. & Patterson, K. Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory*, 1995, 3, 463-495.
- Hodges, J., Patterson, K., Oxbury, S & Funnell, E. Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. *Brain*, 1992, 115, 1783-1806.
- Kawamoto, A. Time-course of processing in feed-forward connectionist networks: An analysis of the exclusive-or problem. *Journal of Mathematical Psychology*, 1993, 37(4) 556-574.
- Keil, F. Conceptual domains and the acquisition of metaphor. *Cognitive Development*. 1986, 1(1) 73-96.
- Keil, F. Explanation, association and the acquisition of word meaning. *Lingua*, 1994, 92, 169-196.
- Laiacona, M., Capitani, E. & Barbarotto, R. Semantic category dissociations: A longitudinal study of two cases. *Cortex*, 1997, 33, 441-461.
- Lambon Ralph, M.A., Graham, K.S., Ellis, A.W. & Hodges, J.R. Naming in semantic dementia –what matters? *Neuropsychologia*, 1998, 36, 775-784.
- Lambon Ralph, M.A., Howard, D., Nightingale, G. & Ellis, A.W. 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, 311-338.
- McRae, K., de Sa, V., & Seidenberg, M. S. On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 1997, 126, 99-130.
- McRae, K., de Sa, V., & Seidenberg, M. S. Semantic priming and the structure of semantic memory. *Journal of Clinical and Experimental Neuropsychology*, 1993, 15, 385-386.
- Mandler, J M. How to build a baby II: Conceptual primitives. *Psychological Review*, 1992, 99, 587-604.
- Mandler, J. M. & McDonough, L. Concept formation in infancy. *Cognitive Development*, 1993, 8, 291-318.
- Mandler, J. Precursors of linguistic knowledge. *Philosophical transactions of the Royal Society*, series B, 1994, 346, 1315..
- Mandler, J. M. & McDonough, L. Drinking and driving don't mix: Inductive generalization in infancy. *Cognition*, 1996, 59, 307-335.
- Martin, N., Dell, G., Saffran, E. & Schwartz, M. Origins of paraphasias in deep dyslexia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, 1994, 47, 609-660.
- Masson, M. A distributed memory model of semantic priming. *Journal of Experimental Psychology Learning, Memory, and Cognition*, 1995, 21(1), 3-23.
- Moss, H. E. & Tyler, L.K. A category-specific semantic deficit for non-living things in a case of progressive aphasia. *Brain & Language*, 1997, 60: 55-58.
- Moss, H.E. & Tyler, L.K. (In Press). A progressive category-specific deficit for non-living things. *Neuropsychologia*.
- Moss, H.E., Tyler, L.K. & Devlin, J. Concepts, categories and domains of knowledge: Conceptual structure in a distributed system (submitted for publication)

- Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., Levy, J. & J. Morris. Conceptual structure and category-specific deficits: Drawing distinctions and identifying similarities. *Second International Congress on Memory, Padova, Italy, 1996*.
- Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., & Bunn, E. ( 1998). "Two eyes of a see-through"; Impaired and intact semantic knowledge in a case of a selective deficit for living things. *Neurocase*, 4, 291-310.
- Moss, H. E., Tyler, L. K., Hodges, J. Patterson, K. Exploring the loss of semantic memory in semantic dementia: Evidence from a primed monitoring study. *Neuropsychology*, 1995, 9, 1, 16-26.
- Moss, H., Tyler, L.K. & Jennings, F. When leopards lose their spots: Knowledge of visual properties in category-specific deficits for living things. *Cognitive Neuropsychology*, 1997, 14, 901-50.
- Moss, H.E., Tyler, L.K. & Devlin, J.T. Modeling progressive impairments of semantic memory: Interactions between severity and domain of knowledge. *Paper presented to the British Neuropsychological Society, London, 1999*.
- Plaut, D. C. & Shallice, T. Perseverative and semantic influences on visual object naming errors in optic aphasia: a connectionist account. *Journal of Cognitive Neuroscience*, 1993, 5:1, 89-117.
- Rumelhart, D. E., Hinton, G. E. & McClelland, J. L. A general framework for parallel distributed processing. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* , 1986, (pp. 45-76). Cambridge, MA: MIT Press.
- Sacchett, C. & Humphreys, G. W. Calling a squirrel a squirrel but a canoe a wigwam: A category-specific deficit for artefactual objects and body parts. *Cognitive Neuropsychology*, 1992, 9 (1), 73-86.
- Saffran, E. J. & Schwartz, M. F. Of cabbages and things: Semantic memory from a neuropsychological view---A tutorial review. In *Attention and Performance, XV*, 1992.
- Sartori, G. & Job, R. The oyster with four legs: A neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology*, 1988, 5, 105-132.
- Silveri, M. C., Daniele, A. Giustolisi, L. Gainotti, G. Dissociation between knowledge of living and nonliving things in dementia of the Alzheimer type. *Neurology*, 1991, 41(4) 545-546.
- Small, S. L., Hart, J., Nguyen, T. & Gordon, B. Distributed representations of semantic knowledge in the brain. *Brain*. 1995, 118(2), 441-453.
- Snowden, J., Goulding, P. & Neary, D. Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioral Neurology*, 1989, 2, 167-182.
- Tyler, L.K. & Moss, H. Functional properties of concepts: Studies of normal and brain-damaged patients . *Cognitive Neuropsychology*. 1997, 14, 511-545.
- Tyler, L.K., & Moss, H.E. Going, going, gone?... Implicit and explicit tests of conceptual knowledge in a longitudinal study of semantic dementia. *Neuropsychologia*, 1998, 36, 1313-1323.
- Warrington, E. K. & McCarthy, R. Category specific access dysphasia. *Brain*, 1983, 106, 859-878.
- Warrington, E. K. & McCarthy, R. Categories of knowledge: Further fractionation and an attempted integration. *Brain*, 1987, 110, 1273-1296.
- Warrington, E. K. & Shallice, T. Category specific semantic impairments. *Brain*, 1984, 107, 829-853.
- Wierzbicka, A. *Lexicography and conceptual analysis*. Ann Arbor: Karoma Publishers, 1985.

**ACKNOWLEDGEMENTS**

This research was supported by an MRC (UK) programme grant and Senior Research Fellowship to Lorraine K. Tyler, and a Wellcome Trust Fellowship to Helen E. Moss.

We are especially grateful to Joe Devlin for many important contributions and for his help with many aspects of the modelling reported in this paper. We also thank Jay McClelland for his criticisms on an earlier draft of this paper, and J. Kate Voice, Gareth Gaskell and Matt Davis for helpful discussions on this research.