The distinctiveness of form and function in category structure: A connectionist model.

M. R. Durrant-Peatfield (mark@pc29.lang.psyc.bbk.ac.uk)
L. K. Tyler (l.tyler@psyc.bbk.ac.uk)
H. E. Moss[1] (helen@psy.gla.ac.uk)
J. P. Levy[2] (j.levy@psyc.bbk.ac.uk)
C.S.L. Birkbeck College; Dept of Psychology, Glasgow University[1]; Dept of Psychology, Birkbeck College.[2]

Abstract

We present a new account of category structure derived from neuropsychological and developmental data. The account places theoretical emphasis on functional information. We claim i) the distinctiveness of functional features correlated with perceptual features varies across semantic domains. ii) the perceptual features representing specific functional mechanisms are strongly correlated with their function. The representational assumptions which follow from these claims make strong predictions about what types of semantic information is preserved in patients showing category-specific deficits following brain damage. We present a connectionist simulation which, when damaged, shows patterns of preservation of distinctive and shared functional and perceptual information varying across semantic domains. The data model both classic dissociations between knowledge for artefacts and for living things and recent neuropsychological evidence concerning the robustness of functional information.

Introduction

How is semantic knowledge represented and structured such that brain damage can lead to specific patterns of preserved and impaired knowledge, for example, deficits for living things versus artefacts (Warrington, 1975; Warrington & Shallice, 1984). Do these deficits indicate the discrete localisation of different types of semantic information, or can they be accounted for by more diffuse and distributed forms of neural representation?

In this paper we present a model of conceptual representations in semantic memory that can account for category-specific deficits without necessarily assuming neuroanatomically distinct subsystems for living and non-living things, or perceptual and functional features. Our model is similar in many respects to that of Devlin, Gonnerman, Andersen, & Seidenberg (in press), in that distinctiveness (how reliably a feature picks out one item from others which are similar) and correlatedness of properties (the extent to which properties regularly occur together) are major determinants of conceptual structure. However, we advance on this work in an important way. Our theoretical emphasis is on functional information and its role in the cognitive system's resistance to damage. We claim that the robustness of perceptual information following brain damage is determined by the extent to which it is associated with or entails functional information. Functionally significant perceptual information (e.g., the serrated edge of a saw) will be more robust to damage than perceptual information lacking functional significance (e.g., a lion's mane).

This approach amounts to a theory governing the construction and structure of semantic representations. Like other models, our account predicts overall differences between living and non-living things. Of greater theoretical importance is that we make explicit the contributions of different types of semantic information to category-specific deficits, generating detailed predictions about the kinds of information that are relatively well and poorly preserved within different categories following damage.

Functional Information

It has been suggested in the neuropsychological literature that functional information is not of primary importance for the representations of living things, and that it is not strongly correlated with perceptual properties (De Renzi & Lucchelli, 1994; Warrington & Shallice, 1984). Whereas artefacts have generally been designed with one function in mind and created to interact with the environment in a specific fashion, living things were not.

Nevertheless, recent evidence suggests (Tyler & Moss, in press) that the semantic representations of living things have at their core a type of functional information, which we refer to as *biological functional information*. This represents the many ways in which living things interact with the environment (e.g., being able to move, fly, eat, drink, see, hear etc).[1]

Of particular importance to our account is the proposal that the representations of living things become correlated with information derived from the same cognitive processes which ascribe functional information to artefacts (J. Mandler, 1992). Infants observe the events in which animals and vehicles take part and base their interpretation of what kind of thing something is on a general analysis of the movements and spatial relations that characterise the event. For example, the movement of living things is generally unpredictable and not contingent on an external agent. In contrast, movements associated with artefacts tend to be predictable and generally initiated by an agent.

Subsequent cognitive development permits a more sophisticated form of perceptual analysis of events so that the broad functional characteristics of each domain (e.g., animacy, inanimacy) become differentiated into more specific modes of interaction.

The perceptual features of artefacts correlate with a mode of interaction in the environment initially characterised as

[1]Living things also have non-biological functions as well (e.g., cats – are pets) but we do not believe that these are central to their representations

movement which is not self-initiated. Artefacts tend to be acted upon rather than act upon (e.g., tools). The shared features of, for example, hand-held tools (e.g., is used in a certain context, is small, shiny and hard) initially correlate with events where the artefact is held and moved by an animate agent. However, to fully understand the manner in which an artefact is used requires knowledge of the features which permit its *specific* function. The semantic representation of an artefact must therefore at some stage encode those perceptual features which distinguish one category member in a class of similar artefacts from others. Distinctive perceptual features therefore become strongly correlated with an equally distinctive mode of interaction with the environment (e.g., the artefact's function). This is not necessarily the case for the shared perceptual features which do not indicate, and are not correlated with, a specific function – rather they implicate the common characteristics of events in which hand-held artefacts are used.

Whereas the distinctive perceptual features of artefacts tend to be correlated with an activity or function specific to that artefact, the distinctive perceptual features of living things (e.g., a tiger having stripes or a male lion having a mane) tend not to be correlated with a psychologically salient function. Rather, it is the shared perceptual properties of living things which correlate with biological functional properties (e.g., having mouths, legs, eyes, ears). That is, shared perceptual features of living things correlate with common modes of interaction with the environment (e.g., eating, walking, seeing, hearing).

This predicts a tendency for artefact functional properties to be more distinctive than the functional properties of living things, which is borne out by analyses of property generation norms (Durrant-Peatfield, Tyler, Moss & Levy, 1997). Subjects provided three times as many distinctive functional properties for artefacts than for living things whereas the number of distinctive and shared perceptual properties were similar in both domains.

The model

We used a connectionist system to investigate our representational claims in relation to the patterns of preservation of semantic information following brain damage. A simple connectionist system learns statistical regularities in the environment. Through the repeated co-occurrence of different features it can learn that the appearance of one feature tends to accompany the presence of another. An interesting characteristic of a system which has learned correlations between features is that such features are more robust to damage than features not supported by correlations. The system tends to determine the correct activation value for a damaged feature on the basis of evidence contributed by the other feature. This has profoundly important implications for the patterns of preservation of perceptual and functional information following damage to a system which has recorded the functional significance of perceptual information. Specifically, distinctive perceptual features for artefacts and shared perceptual features for living things will be relatively robust to damage because of correlational support from functional information.

The assumptions motivating the model are:

- The distinctiveness of functional information (information correlated with a specific activity) varies across semantic domains. Artefact functional properties tend to be more distinctive than biological functional properties.
- Each functional feature is correlated with the presence of a perceptual feature. [2]
- Functional information is always correlated with the presence of perceptual information but not necessarily vice-versa (e.g., colors and textures have no specific functional importance).
- Correlated and inter-correlated features [3] will be relatively resistant to damage. The robustness of a feature is proportional to the number of features it is correlated or inter-correlated with.

Distinctive		**Shared**		**Functional**	
		Artefacts			
1000	0000	1010	0000	1000	0000
0100	0000	1010	0000	0100	0000
0010	0000	1010	0000	0010	0000
0001	0000	1010	0000	0001	0000
0001	0000	0101	0000	0001	0000
0010	0000	0101	0000	0010	0000
0100	0000	0101	0000	0100	0000
1000	0000	0101	0000	1000	0000
		Living things			
0000	1000	0000	1010	0000	1010
0000	0100	0000	1010	0000	1010
0000	0010	0000	1010	0000	1010
0000	0001	0000	1010	0000	1010
0000	0001	0000	0101	0000	0101
0000	0010	0000	0101	0000	0101
0000	0100	0000	0101	0000	0101
0000	1000	0000	0101	0000	0101

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

A set of 16 vectors representing two categories from the domains of living things and artefacts were constructed to encode our representational claims (see Figure 1). Each functional feature was correlated with the presence of a perceptual feature. Similar artefacts were distinguished in terms of function so that the corresponding perceptual information was as distinctive as the functional information (e.g., saw: serrated edge, sawing). Similar living things were distinguished in terms of perceptual information which did not correlate with

[2] Living things perform many biological functions (e.g., eating, walking, flying) and therefore have more perceptual features correlated with functional features than artefacts which are generally designed to have one function. Put simply, living things 'do' more than artefacts.

[3] We use the term correlated to indicate correlations between properties of a different type and inter-correlated to refer to correlations between properties of the same type

a function (e.g., tigers' stripes). Biological functional features which represent the shared activities common to similar living things were correlated with the presence of shared perceptual information.

A three layer feed-forward net was trained using back-propagation to reproduce the input on the output layer. Three hundred networks were trained for 500 epochs with different initial random weights with a learning rate of 0.25 and momentum 0.9 until individual pattern error was less than 0.01.

Lesioning

The system was lesioned to simulate global brain damage by randomly setting a proportion (initially 10%) of all connection weights to zero. The proportion was increased by increments of 10% until 80% of the inter-layer connections were set to zero. We then determined the extent to which each type of semantic information was affected by lesioning at each level, examining in turn, distinctive and shared perceptual and functional features.

Distinctive features

We predicted that the preservation of distinctive perceptual and functional properties will vary depending on whether or not they are inter-correlated or correlated with the presence of functional properties. Distinctive perceptual properties of artefacts should be more robust than the distinctive perceptual properties of living things because they are correlated with the presence of functional information. Thus, preserved distinctive perceptual information for artefacts (e.g., the serrated edge of a saw) should always accompany equally distinctive functional knowledge (e.g., used for sawing). Both types of artefact knowledge should be better preserved than distinctive perceptual knowledge for living things.

In behavioural terms, the importance of correlated artefact functional and perceptual features is that if sufficient evidence supports the existence of one feature, the other will tend to be reinstated. If the serrated edge of a saw can be recovered then so can its function, and vice-versa. That functional information can be derived from perceptual (and vice-versa) for artefact concepts is suggested by de Renzi & Lucchelli (1994) while Caramazza, Hillis, Rapp & Romani, (1990) argued that the functions of objects can frequently be inferred from their perceptual properties. In contrast, for living things a distinctive perceptual feature (e.g., tigers' stripes) does not consistently accompany an equally distinctive functional feature. There is no additional evidence from correlations learned by the network to reinstate a distinctive perceptual feature; thus it will be less robust to lesioning.

Examination of the simulation data support these predictions. The distinctive correlated features, which in our model represent artefact distinctive perceptual and functional properties, were contrasted with distinctive non-correlated features representing distinctive perceptual features for living things. Figure 2 shows how feature error (the absolute difference between the target and the output activation level) interacts with lesioning severity. Artefact distinctive perceptual and functional features are initially significantly more robust to damage than distinctive perceptual features for living things ($p < 0.001$). The difference becomes non-significant after 60% of the connections have been lesioned.

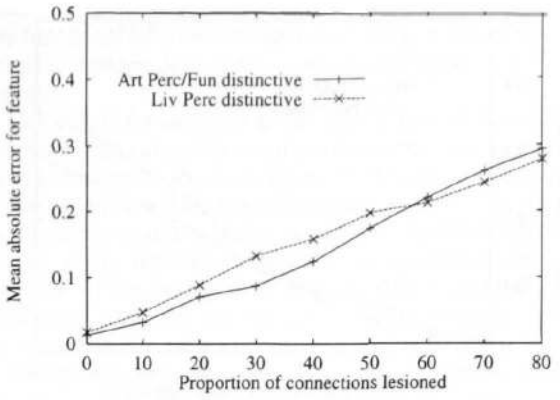

Figure 2: Mean absolute error for distinctive perceptual/functional feature units for artefacts and distinctive perceptual knowledge for living things at different levels of lesioning severity over 300 simulations.

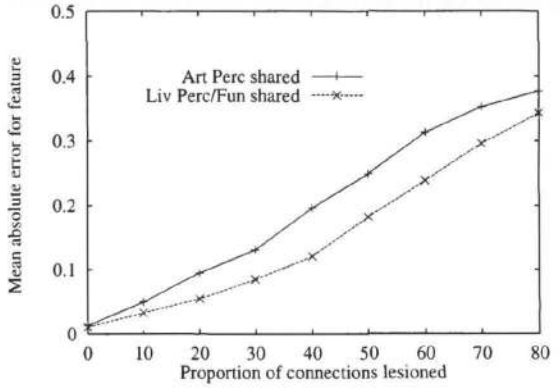

Figure 3: Mean absolute error for shared perceptual/functional feature units for living things and artefact shared perceptual features at different levels of lesioning severity averaged over all 300 simulations.

Shared features

Now we investigate how correlations between shared properties affects robustness to damage. We claim that artefact shared perceptual features are inter-correlated with other artefact shared perceptual features but not with functional information (e.g., being made of metal and having a handle do not imply a specific function). Shared perceptual features for living things are correlated with other shared perceptual features and with functional features. The robustness of a feature depends on the number of correlations it has with other features. In modelling terms, the effect of increasing the number of inter-correlations is that a damaged network needs less evidence to determine the correct activation values for a unit. Shared perceptual properties of living things should therefore be more robust to damage than shared perceptual properties of artefacts. We therefore predict that biological functional information (e.g., walking, seeing) and shared perceptual fea-

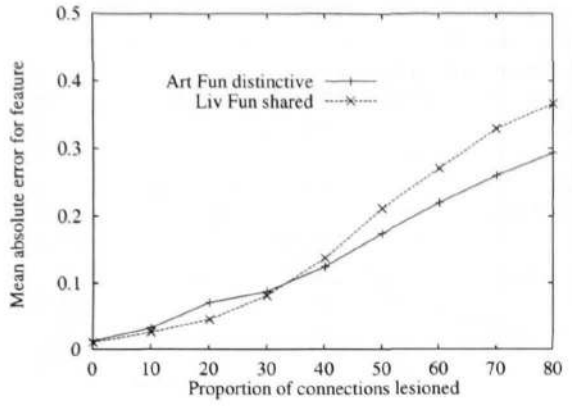

Figure 4: Mean absolute error for functional feature units for artefacts and living things at different levels of lesioning severity averaged over all 300 simulations.

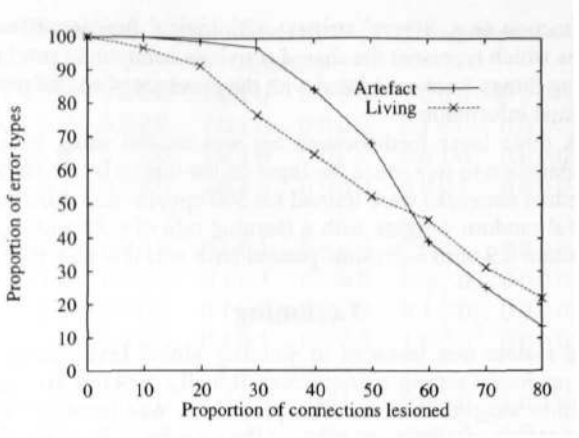

Figure 5: The model's identification performance averaged over all 300 simulations.

tures for living things (e.g., legs, eyes) will be better preserved than the shared perceptual features of artefacts (e.g., is made of metal).

An examination of the error associated with each type of shared feature in our simulation data supports this prediction. Figure 3 shows that shared perceptual and functional features for living things are consistently better preserved (lesioning severity in the range 10% to 80%) than shared artefact perceptual features ($p < 0.001$).

Distinctive vs Shared features

Now we contrast distinctive and shared correlated features and address the robustness of functional information. Functional properties in both domains will be relatively robust to damage because, as we have claimed, functional information always accompanies perceptual properties. That is, functional features are consistently correlated with the presence of perceptual features. However, biological functional properties are expected to be appreciably more robust than artefact functional properties because biological functional features are also inter-correlated. The simulation data in Figure 4 shows an interaction between functional property distinctiveness and lesioning severity which was not anticipated. Biological functional features are initially more robust to damage than artefact functional features as we predicted ($p < 0.001$) but at more severe levels of lesioning, artefact functional information is consistently and significantly better preserved ($p < 0.001$). As damage accumulates to the network and correlational information is lost, the output activations of all features tend to be greater than 0, regardless of the target. This baseline activation is proportional to the frequency with which a feature has been encountered. As shared features were encountered more frequently than distinctive, the absolute error for biological functional features was correspondingly higher than for artefact functional features at severe levels of lesioning.

Pattern classification performance

The previous section examined how distinctiveness and correlatedness interacted with lesioning severity to affect feature robustness. In the following section we how see how these variables affect performance on tasks whose behavioral equivalents are item identification and categorisation.

The type of within-category information that is lost or preserved following brain damage provides important information about the way in which categories are structured. Most accounts of semantic impairments predict that damage will non-selectively affect all types of featural information while preserving higher-level knowledge (category and super-ordinate), either because semantic information is organised hierarchically within a category (Warrington, 1975) or because super-ordinate and category information can be supported by partial semantic information in a distributed model (Tippett, McAuliffe, & Farah, 1995).

In contrast, our account claims that strongly inter-correlated perceptual and functional features will be better preserved following brain damage than those that are weakly inter-correlated. This generates the following predictions about the model's identification and categorisation performance.

a) Distinctive perceptual information is required to successfully discriminate between living things. Since this information is not strongly correlated with functional information, it will not be robust to damage. We therefore predict that the damaged model will perform badly on identification of living things. However, for artefacts, this will not be a problem, since distinctive form and function are most strongly correlated, and so will be preserved well enough to support discrimination among category members.

b) Categorisation depends on shared properties. Thus, categorising living things should be good because shared properties are preserved by the inter-correlations with functional information. Artefacts, however, whose shared properties are not supported by correlations with functional information should be more vulnerable to damage.

To test these predictions, the lesioned network's output was compared on a vector by vector basis with the complete target set to establish the closest neighbour. A successful identification in this context means that the lesioned network reproduced a pattern of activation whose nearest neighbour was the

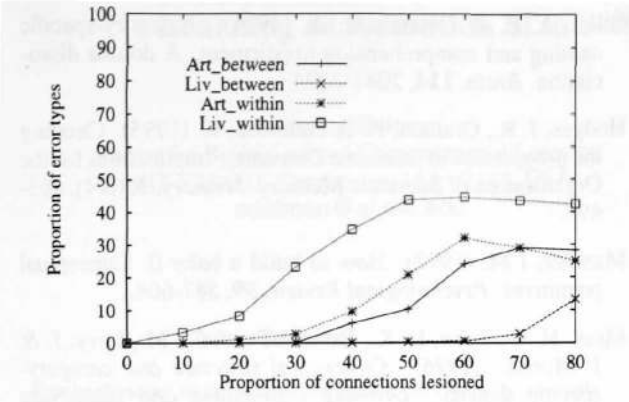

Figure 6: Between and within-category errors averaged over all 300 simulations.

target.

Figure 5 shows how lesioning affects the model's mean performance over 300 simulations on accurately identifying a target vector at progressively more severe levels of damage. Accuracy on identifying living things quickly falls off from the first level of lesioning severity compared with artefact identification performance. Identification performance for both then declines at a similar rate though performance for artefacts is consistently better than for living things up until half of the inter-layer connections have been lesioned. This models the classic dissociation between knowledge for artefacts and living things widely reported in the neuropsychological literature (Warrington & McCarthy, 1987; Warrington & Shallice, 1984; Saffran & Schwartz, 1992).

When greater than 60% of the connections between layers have been lesioned the trend is reversed. Identification accuracy for artefact vectors is worse than for living things, a pattern occasionally reported although much less frequently than the reverse dissociation (Hillis & Caramazza, 1991; Warrington & McCarthy, 1987).

The type of error made by a patient provides a useful insight as to what type of semantic information has been lost and what has been preserved. For example, purely within-category errors (e.g., confusing *dog* and *cat* within the category of mammals) indicate the degradation of distinctive features whereas between-category errors (e.g., confusing a bird with a cat) indicate the loss of category information.

Figure 6 plots the type of error made against lesioning severity. The network consistently makes more within-category errors for living things than for artefacts. This reflects the difference in robustness of distinctive perceptual features for artefacts and living things. Distinctive perceptual features for living things are not as robust to lesioning as artefact distinctive features because they are not correlated with functional information. In contrast, artefact between-category errors occur at an earlier stage of lesioning severity than the first between-category errors for living things. This is expected since the shared perceptual features for living things which locate an item within a category are supported by correlations with functional features. Artefact shared perceptual features are supported only by inter-correlations with each other. Category information (e.g., shared perceptual features and functional information) for living things is therefore more robust to damage than artefact category information as Figure 3 shows [4]

The deficit for artefacts at the later stages of lesioning is also due to the susceptibility of shared artefact perceptual features to damage as shown in Figure 5. As more damage accumulates to the network, the information which locates an artefact category member within a category becomes more degraded so that between-category errors become increasingly frequent. If the damaged net is unreliable in accurately determining a category, the chances of determining the correct category member are also reduced.

These simulation data show that category-specific deficits can be captured by damaging a distributed system. Neuroanatomically distinct subsystems for the storage of living and non-living things, or of perceptual and functional features are not necessary to produce category-specific deficit behaviour (Farah & McClelland, 1991).

Discussion

In our model, general, non-focal damage produced a pattern of performance with the following features:

- Distinctive perceptual information for artefacts better preserved than for living things.
- Shared perceptual properties for living things better preserved than artefact shared perceptual properties.
- Functional properties preserved although the relative robustness of artefact and biological functional information interacts with lesioning severity.

These patterns of preservation predict a neuropsychological deficit characterised by difficulty in identifying living things but relatively intact knowledge about artefact functional and biological functional information. Specific artefacts should therefore be identifiable and their function retrieved, but performance on tasks requiring the grouping of artefacts in different categories should be relatively impaired (e.g., sorting artefacts into categories on the basis of shared features). The data therefore model the classic dissociation between knowledge for artefacts and for living things (Warrington & Shallice, 1984) and more recent neuropsychological evidence concerning the robustness of functional information (Moss, Tyler, Patterson & Hodges, 1995; Tyler & Moss, in press). For example, Moss, Tyler, Durrant-Peatfield, Levy, & Morris, (1996), discuss RC, a HSE patient who had the typical 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 among artefacts. This was demonstrated, for example, in a picture sorting task; RC had no difficulty sorting pictures of living things according to category

[4] This finding is supported by a similar analysis we conducted on the Devlin *et al.* (in press) vector set which was derived from subject property generation data. See also Hodges, Graham & Patterson, (1995) for a discussion on between-category errors.

(e.g., identifying their shared properties) but was slightly impaired in sorting artefacts. In contrast, he had considerable difficulty when asked to sort the pictures of living things according to specific properties (e.g. fierce vs non-fierce animals), with significantly poorer scores than for artefacts (e.g. electrical vs non-electrical appliances). RC's definitions to words reported in Tyler & Moss (in press) included both biological functional information and shared perceptual features, which are similarly preserved following damage (see Figure 3), whereas distinctive perceptual knowledge is not readily accessible. For example, *bee* was defined as: "...two eyes of a see-through, or a hearing of two ears, of a mouth, of an eating and drinking". In contrast, for artefacts, RC's definitions contain precise descriptions of distinctive functional and perceptual properties: For example, *desk* was defined as: "...On the desk yeah, of an eating of a drinking of a working or a making things for the house for an interest if a use of , for whatever your interest...a desk, a desk is to, a wooden chair, and four legged and, and has a ...shelf to, to do writing on, to eat on, to build, to make things on."

Conclusion

Our account 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 correlatedness of features in the system's internal representations. We claim that it is the accompanying differential robustness of perceptual and functional information which underlies category-specific deficit behavior.

In conclusion, this research makes two important advances on Warrington & Shallice (1984). First, like artefacts, the perceptual features of living things are correlated with a type of information that represents the manner in which living things interact with the environment. We have referred to this information as *biological* functional. Second, the distinctiveness of functional properties is an important variable when modelling category-specific deficits.

References

Caramazza, A., Hillis, A. E., Rapp, B. C. & Romani, C. (1990). The multiple semantics hypothesis: multiple confusions? *Cognitive Neuropsychology*, **7**, 161-189.

Devlin, J., Gonnerman, L., Andersen, E. & Seidenberg, M. (in press). Category specific deficits in focal and widespread damage: A computational account. *Journal of Cognitive Neuroscience.*

De Renzi, E. & Lucchelli, F. (1994). Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex*, **30**, 3-25.

Durrant-Peatfield, M., Tyler, L. K., Moss, H. E., & Levy, J. P. (1997). Disintinctiveness and category structure. Manuscript in preparation.

Farah, M. J. & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, **120** (4), 339-357.

Hillis, A. E. & Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, **114**, 2081-2094.

Hodges, J. R., Graham, N. & Patterson, K. (1995). Charting the progression in Semantic Dementia: Implications for the Organisation of Semantic Memory. *Memory*, **3**, (3/4), 463-495.

Mandler, J M. (1992). How to build a baby II: Conceptual primitives. *Psychological Review*, **99**, 587-604.

Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., Levy, J. & J. Morris. (1996). *Conceptual structure and category-specific deficits: Drawing distinctions and identifying similarities*. Second International Congress on Memory, Padova, Italy.

Moss, H. E., Tyler, L. K., Hodges, J. Patterson, K. (1995). Exploring the loss of semantic memory in semantic dementia: Evidence from a primed monitoring study. *Neuropsychology*, **9**, 1, 16-26.

Saffran, E. J. & Schwartz, M. F. (1992). Of cabbages and things: Semantic memory from a neuropsychological view—A tutorial review. In Attention and Performance, XV.

Tippett, L. J., McAuliffe, S. & Farah, M. J. (1995) Preservation of Categorical Knowledge in Alzheimer's Disease: A Computational Account. *Memory*, **3** (4), 000-000.

Tyler, L.K. & Moss, H. (in press). Functional properties of word meanings: Studies of normal and brain-damaged patients. *Cognitive Neuropsychology.*

Warrington, E. K. & McCarthy, R. (1987). Categories of knowledge: Further fractionation and an attempted integration. *Brain*, **110**, 1273-1296.

Warrington, E. K. & Shallice, T. (1984). Category specific semantic impairments. *Brain*, **107**, 829-853.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, **27**, 635-657.