A Neural Network Model of Visual Object Recognition Impairments after Brain Damage

N.Davey[1], R.J.Frank[1], T.M.Gale[2,3], S.J.George[1]
Email: n.davey, r.j.frank, t.gale, s.j.george@herts.ac.uk

[1]Department of Computer Science,
[2]Department of Psychology,
University of Hertfordshire,
College Lane, Hatfield,
Hertfordshire,
AL10 9AB.

[3] Hertfordshire Neurosciences
Research Group,
Queen Elizabeth II Hospital,
Welwyn Garden City,
Hertfordshire.

Abstract
Dysfunction of the visual object recognition system in humans is briefly discussed and a basic connectionist model of visual object recognition is introduced. Experimentation in which two variants of this model are lesioned is undertaken. The results suggest that the well documented phenomenon of superordinate preservation is model independent. Differential category specific recognition deficits are also observed in this model, however these are sensitive to each particular variant.

Introduction
Connectionist models of neuropsychological phenomena can provide a potentially useful insight into the nature of cognitive information processing within the brain. In particular, the ability to damage a neural network and observe the resulting behaviour can throw light upon aspects of cognitive dysfunction. In this paper we show how two variations on a modular connectionist model, can be trained to map a pictorially represented object to a semantic feature vector. Both models are then lesioned and the results obtained are compared with known neuropsychological phenomena. One the major issues addressed is the extent to which phenomena in the model are persistent across the two variants.

Background
Visual Object Recognition (VOR), that is the ability to perceive and comprehend physical objects in our environment, is well known to be disrupted by a range of organic and non-organic brain disorders, such as Alzheimer's Disease [Hodges, Salmon & Butters, 1992; Done and Gale 1997] and Traumatic head injury [Funnell & Sheridan, 1992]. Although sources of injury often vary, the psychological functioning of visual agnosics can be remarkably consistent.

The best documented neuropsychological finding is loss of accuracy for fine-grained detail, in contrast to a marked preservation of general information. Patients will often use the superordinate term, such as fruit, when naming a specific object (e.g. apple). This phenomenon has been used as evidence for the disruption of the semantic system, where fine-grained knowledge is used to disambiguate similar objects. Possible interpretations of this finding are: (1) pre-semantic processes are intact, but the information needed to imbue the percept with meaning is either lost, degraded or inaccessible; (2) the inaccessible; (3) both perceptual and semantic processes are compromised.

A second finding is the emergence of category specific recognition deficits (CSRD). Patients with CSRD typically exhibit poor comprehension of some object classes (e.g. living things) yet have no difficulty with others (e.g. non-living things) [Farah, Meyer & McMullen, 1996]. As with preservation of superordinate knowledge it is not clear whether CSRDs arise through disrupted cognitive or perceptual mechanisms or a combination of both [Sartori & Job, 1988; Humphreys, Riddoch, Quinlan, 1988].

Artificial Neural Networks (ANNs) offer neuropsychologists increasing insight into the nature of disordered brain processes e.g. [Hinton and Shallice, 1991; Plaut and Shallice, 1993; Tippett, McAuliffe & Farah, 1995]. In the absence of neurophysiological data they facilitate refinement of theories and often generate hypotheses which can be tested in controlled patient studies. ANNs can be powerful tools when used to simulate the role of individual modules within a larger, more complex, system and this has strong implications for the study of normal and disordered VOR since it permits a line of investigation not possible with human subjects.

Details of the Model
We have developed two modular models of VOR that incorporate an unsupervised perceptual processing module and a supervised semantic memory module, see figure 2. An important feature of these models is that the perceptual processing module uses real pictorial data and represents the data using an unsupervised self organising feature map (SOFM), so that there is no inbuilt bias as to salient visual features.

The semantic memory module takes the form of a feedforward pathway ending in either a single layer attractor (type-A network) or a multiple layer attractor network (type-B network, such as that described by [Hinton & Shallice, 1991]). In the first case the output from the semantic units is passed through 30 clean up units, before back-propagation of errors, and in the second case the output is iterated eight times through the clean-up units with back-propagation through time, learning taking place on the last 3 iterations. Testing

robustness to training variation. Additionally the performance of the type-A model has been verified as consistent with data obtained from psychological, electrophysiological and primate learning studies of VOR [Gale, Done & Frank, submitted].

The separation of perceptual and semantic process in both models allows investigation of each process independently. This is an approach which is of relevance to study of preservation of superordinate information and CSRDs, but cannot be achieved with patients.

Representation of Information in the Model

Our model attempts to remove some of the problems which arise from the use of feature-based input representations by using real pictorial stimuli. The training set comprised a total of 560, 8 bit greyscale images deriving from the 4 superordinate categories of animals, musical instruments, clothing and furniture. Each object was depicted in a canonical perspective and all background detail was removed. Each superordinate category comprised 7 basic-level categories, so animal is subdivided into: bird, snake, spider, fish, deer, mouse and frog, Each basic-level category is further divided into 5 subordinate categories; for example fish are: pike, carp, salmon, herring and bass. These categorical levels reflected the tripartite hierarchy of increasing specificity originally proposed by [Rosch et al. 1976]. Each subordinate was represented by 4 versions of the same exemplar, which varied on dimensions of contrast and left-right inversion. Some example images are displayed in figure 1.

Figure 1: Examples of greyscale images presented to our modular architecture. Each image fits within a 50 by 50 pixel grid such that the principal dimension comes within one pixel of the grid border. These examples depict the 5 subordinate images of the basic level category 'clock' which, in turn, is one of the 7 basic level categories representing furniture.

Each image was processed by a self-organising feature map (SOFM) [Kohonen, 1982; 1988] similar to that used by [Schynns 1991]. The output of the SOFM, for each image, was presented to the semantic module.

The output vector for each pattern is a distributed binary representation of 32 bits. To prevent bias all patterns are coded by the same number of active bits and each unit is active for the same number of patterns. The 32 bits encode both superordinate and basic-level properties of each exemplar. Each unit plays an equal role in the representation of each feature type, thus removing any architectural distinction between superordinate and basic-level information. Complete counterbalancing within the

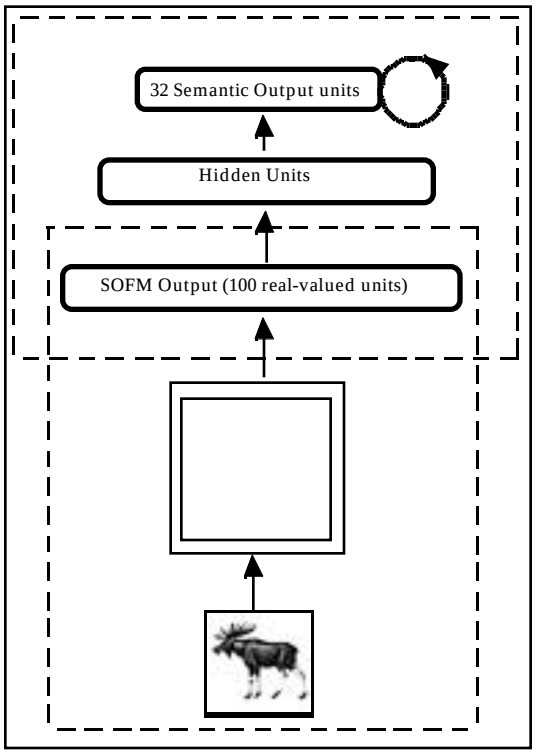

Figure 2: A representation of the full model showing intersection between the perceptual and semantic modules

Lesioning was operationalised in both types of the model by random removal of connections (i) between the SOFM and hidden layers and (ii) within the semantic memory module (i.e. deletion of intra-layer connections).

In the type-A model severity levels were set at 5%, 10%, 20% and 40%, deletion of connections at each lesion site. Each increase was calculated on a cumulative basis, so that connections deleted on earlier occasions were not re-instated on subsequent lesions. Furthermore, each lesion was performed in 10 random variations. The effects of lesioning were operationalised in terms of mean *unit output discrepancy* (UOD) for each feature type within each pattern. This involved taking the modulus of the discrepancy between the actual and desired activation value for each *semantic* output unit, and averaging across both sets of semantic output units (i.e. superordinate or basic level) for a single training pattern. Twelve semantic memory modules were trained, each with output from a different SOFM and each with a random configuration of starting weights. Results of lesioning were averaged across all random variations for all 12 networks, resulting in a mean basic-level and superordinate level UOD for each of the 4 categories (A, F, MI and C) at each of the 2 lesion sites.

In the type-B model an increasing percentage (10% through to 50%) of connections were deleted at each lesion site. Again each lesion was performed in 10 random variations for each lesion site. The results were averaged. In both cases a pattern was labelled as

vector against the target was above 0.8 and the next closest vector had an overlap of no more than 0.75.

Results

Superordinate level, basic-level and exemplar level information is lost in both types of model during early and late lesioning. Exemplar level information is initially lost more rapidly than superordinate information. This loss of information occurs independently of the site of any damage.

Networks lesioned late in the process, within the semantic memory module experienced a net gain in category level information throughout lesioning, see table 1. However lesioning early in the process exhibited a category dependant change to superordinate information. Within the musical instrument and animal categories both exemplar and superordinate level information was lost, whilst a minimal net gain in superordinate information was exhibited in each of the clothes and furniture categories.

	LESIONING SEVERITY				
LESION	**10%**	**20%**	**30%**	**40%**	**50%**
Early SO	165	182	173	175	178
Early EX	247	165	123	90	73
Late SO	51	87	107	133	154
Late EX.	500	461	436	404	370

Table 1: Number of patterns recalled correctly at Superordinate level only (SO) and Exemplar level (EX- including correct superordinate recall)

Category specific recognition deficits (CSRDs) were exhibited in both model types; however the extent of the recognition deficit was more widespread in the type-A model. In the type-A model CSRDs occurred for different categories as a function of both lesion site (early or late) and type of semantic representation (superordinate or basic-level), see table 2. Early lesions generated a higher number of errors for animal and musical instrument basic level categories. Late lesions have less impact on model accuracy overall (see figure 3). However, a superordinate CSRD for furniture was consistently observed. It is also interesting to note that clothing items are the most robust category to semantic layer lesioning. In this respect they seem to behave in a similar way to animate categories and musical instruments. However, this is perhaps unsurprising given that the perceptual qualities of clothes are largely dictated by the shape of body parts, a category that is considered by most to be biological in nature.

		LESIONING SEVERITY			
LESION	**Cat-egory**	**5%**	**10%**	**20%**	**40%**
Early SO	A	0.07	0.19	0.30	0.21
Early SO	F	0.04	0.14	0.31	0.27
Early SO	MI	0.04	0.19	0.22	0.63
Early SO	C	0.05	0.15	0.26	0.49
Early BL	A	0.17	0.33	0.41	0.47
Early BL	F	0.06	0.21	0.30	0.38
Early BL	MI	0.11	0.30	0.37	0.46
Early BL	C	0.05	0.15	0.25	0.37
Late SO	A	0.005	0.045	0.07	0.19
Late SO	F	0.025	0.08	0.15	0.22
Late SO	MI	0.01	0.05	0.105	0.15
Late SO	C	0.01	0.05	0.08	0.075
Late BL	A	0.04	0.075	0.18	0.215
Late BL	F	0.01	0.09	0.145	0.24
Late BL	MI	0.015	0.075	0.155	0.165
Late BL	C	0.015	0.045	0.075	0.14

Table 2: Mean UOD of superordinate (SO) and basic level (BL) classifications of each taxonomic category (A: Animals; F: Furniture; MI: Musical Instruments; C: Clothes) after lesioning in type-A model.

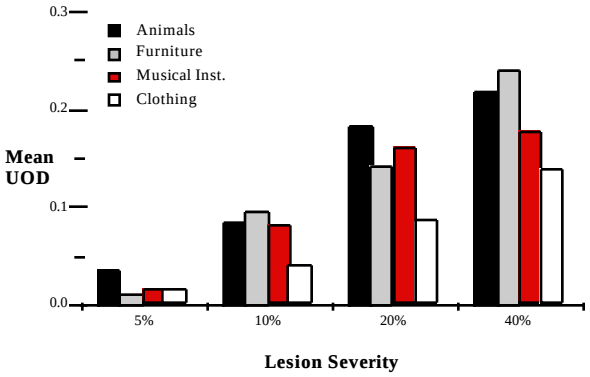

Figure 3: An example of the errors made with late lesioning in a type-A network. The figure shows the mean basic level UOD for each taxonomy as semantic layer lesion severity increases. At mild severity levels, animal basic-level representations were more severely affected. Across all other severity levels clothing basic-level representations were consistently better preserved.

Discussion

At all levels of lesion severity and for all lesion sites, superordinate information was better preserved. This concurs with a wealth of neuropsychological data suggesting that relative preservation of superordinate knowledge is consistent across different forms of visual agnosia. In this type of model, superordinate features are better preserved because they are activated more frequently than basic-level features and hence form stronger inter-connection weights.

Both type-A and type-B models suffer semantic errors

Early lesions generated a higher number of errors for animal and musical instrument basic level categories. The most likely reason for this is that these categories are perceptually homogenous [Gaffan and Heywood, 1993; Gale et al., 1998] meaning that their exemplars differ from each other on only a few perceptual dimensions. Early damage results in less perceptual information passing through the network so there is a high risk that these exemplars will no longer be discriminable from their associates. Conversely, for perceptually heterogeneous categories where exemplars are differentiated over many dimensions, a reduction in information is less likely to prevent the model from discerning between category members.

During late lesioning a superordinate CSRD for furniture was consistently observed in the type-A model. The most plausible explanation for this is that furniture is a perceptually diverse category characterised by high variation on input dimensions. Whilst this may render fine-grained discrimination between furniture patterns easy, it creates difficulties when the model has to group these exemplars together into one superordinate category. The connections in the semantic layer seem to be particularly important in this task, in contrast to other categories where superordinate coherence seems to be achieved earlier on in the processing cycle.

The two models exhibited a differential generation of error when submitted to late lesioning. It is thought that within the type-A model the semantic attractor layer was performing more of the semantic association than the same layer in the type-B model. The multiple semantic attractor layers in type-B model may be sharing the development of a final attractor state and consequently the errors produced when the single layer is lesioned will be more significant than those errors produced when a single layer from a multiple layer attractor system is lesioned.

Conclusions

The fundamental underlying behaviour of superordinate preservation is evident in both type-A and type-B models. This preservation is independent of the site of damage.

Category specific damage was evident in both type-A and type-B models however, the nature of the errors appears to be specific to the variant of the model used. The differential generation of errors within our models indicates that, contrary to previous work [Hinton and Shallice, 1991], the nature of the model's architecture and training regime can influence the exact nature of the error. Results from this study and others (e.g. [Gale et al., 1998]) suggest that CSRDs arise through categorical structure inherent in some perceptual representations rather than through anatomical separation of category representations in the brain.

References

Done, D.J. and Gale, T.M. (1997) Attribute verification in dementia of Alzheimer's Type: Evidence for the preservation of distributed concept knowledge. *Cognitive Neuropsychology,14,* 547-571.

Farah, M.J., Meyer, M.M., and McMullen, P.A. (1996) The living/non-living dissociation in not an artifact:

Funnell, E., and Sheridan, J. (1992) Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cognitive Neuropsychology, 9,* 135-153.

Gaffan, D., and Heywood, C.A. (1993) A spurious category-specific visual agnosia for living things in normal human and non human primates. *Journal of Cognitive Neuroscience, 5,* 118-128.

Gale, T.M., Done, D.J., and Frank, R.J. (1998) Modelling visual object recognition with a modular neural network architecture. *Cognitive Science,* submitted.

Hinton. G.E., and Shallice, T. (1991) Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74-95.

Hodges, J.R., Salmon, D.P., and Butters, N. (1992) Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia, 4,* 301-314.

Humphreys, G.W., Riddoch, M.J., and Quinlan, P.T. (1988) Cascade processes in picture identification. *Cognitive Neuropsychology, 5,* 67-103.

Kohonen, T. (1982) Self-organised formation of topologically correct feature maps. *Biological Cybernetics, 43,* 59-69.

Kohonen, T (1988) *Self-Organisation and Associative Memory.* Berlin: Springer-Verlag.

Plaut, D.C., and Shallice, T. (1993) Deep Dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377-500.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., and Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439.

Sartori, G., and Job, R. (1988) The oyster with four legs: A neuropsychological study on the interaction between vision and semantic information. *Cognitive Neuropsychology, 5,* 105-132.

Schynns, P.G. (1991) A modular neural network model of concept acquisition. *Cognitive Science, 15,* 461-508.

Tippett, L.J., McAuliffe, S., and Farah, M.J. (1995) Preservation of categorical knowledge in Alzheimer's disease: A computational account. *Memory, 3,* 519-533.