VISUAL CATEGORISATION IN A MODIFIED SELF ORGANISING MAP: IMPLICATIONS FOR CATEGORY-SPECIFIC AGNOSIA

**Tim Gale[1,2,3] Lorna Peters[1] Ray Frank[1]
Neil Davey[1]**

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

3. Hertfordshire Neurosciences Research Group (HNRG), QEII Hospital, Howlands, Welwyn-Garden-City, Herts, UK, AL7 4HL.

All correspondence to first author
e-mail t.gale@herts.ac.uk
fax: +44 (0)1707 335168
telephone: +44 (0)1707 365383

1. ABSTRACT

There are many reports of patients who, after sustaining brain damage, exhibit a selective recognition deficit for certain categories of object. There has been much controversy as to whether this is informative about the neural organisation of knowledge in the human brain. In this paper we describe an unsupervised neural network model that is trained to process images from a variety of different object categories. Analysis of the unsupervised representations reveals some interesting distinctions between different classes of object. We contend that this model indicates a natural perceptual distinction between certain object categories, which may become exaggerated by the effects of human brain damage.

2. INTRODUCTION

There are numerous reports in the literature of brain-damaged patients who present with selective deficits in recognising certain classes of objects (for a recent review see Caramazza and Shelton, 1998). The most commonly observed pattern of deficit is where patients have difficulty in recognising and naming pictures of living things relative to non-living things (e.g. Warrington and Shallice, 1984) although the reverse dissociation has also been reported (e.g. Sacchett and Humphreys, 1992). Whilst there is good evidence that some such cases may be explained by poorly controlled experiments (e.g. Funnell and Sheridan, 1992; Stewart, Parkin and Hunkin, 1992), other cases defy such an explanation (e.g. Farah, Meyer and McMullen 1996). It remains unlikely that neurological substrate in the brain honours a clear-cut distinction between living and non-living things since there are very few cases where a pure deficit for one or the other is found. For example, patients who are poor at recognising living things also tend to be poor at recognising musical instruments. Conversely, patients who are poor at recognising non-living things often have similar difficulties with body-parts. Such anomalies offer a serious challenge to the view that knowledge in the brain is organised taxonomically.

In a primate learning study, Gaffan and Heywood (1993) demonstrated that monkeys had greater difficulty in discriminating between picture-pairs of living things, relative to pairs of non-living things. They concluded that the category of living things was 'visually crowded' and that this could lead to an apparent category-specific visual agnosia. They did not present similar data for musical instruments but did show that, when normal subjects were asked to name degraded

line drawings, they showed a relative deficit for both animals and musical instruments. Thus, it is a reasonable hypothesis that visual properties inherent within these categories render them more difficult to recognise when visual processing resources are constrained by brain damage.

We test this hypothesis here using an unsupervised neural network model that is able to process greyscale images. We present it with a series of images deriving from categories of clothing, animals, musical instruments and furniture and examine the nature of representations that develop in the model. Any qualitative difference in representations between categories can be due only to visual aspects of the stimuli since there is no semantic information in the model.

3. THE MODEL

The model is based on Kohonen's self-organising feature map [SOFM] (Kohonen, 1982a; 1982b; 1988) but differs in the nature of competitive learning. In Kohonen's original SOFM, pattern classification was achieved by activation of a single 'winning' unit whereas, in our model, classification is distributed across all units in the feature map, thereby generating a *contoured* internal representation for each input vector. The 'winning' unit in our model is still highly important in the representation of an input vector but, unlike Kohonen's model, does not have *exclusive* diagnostic capability. Our model is structurally and functionally similar to the model described by Schynns (1991). An illustration of a small SOFM can be seen in figure 1.

The model comprises an n-dimensional input vector (A), where n is the number of pixel values in each training image - in this case 2500 (i.e. 50 by 50), which is fully connected to an output map of 10 units square. Each output unit (o_i) in the SOFM is connected to the input vector with an n-dimensional weight vector (W_i). Each input vector (A_i) comprises 2500 integer values representing the 8-bit greyscale information (i.e. the level of 'greyness' between 0 and 255) and spatial location of *each* pixel in the input vector. At each iteration, the activation map of the output layer is computed by comparing the Euclidean distance between the input vector (A) and the weight vector (W_i) for each output unit. The output unit with the lowest Euclidean distance from the input vector (and

hence the highest output value) is regarded as the winner. The output map is computed using a transfer function that returns a value in the range of 0 to 255 where 0 represents a unit whose weight vector is far away from the input vector and 255 represents a unit whose weight vector is identical.
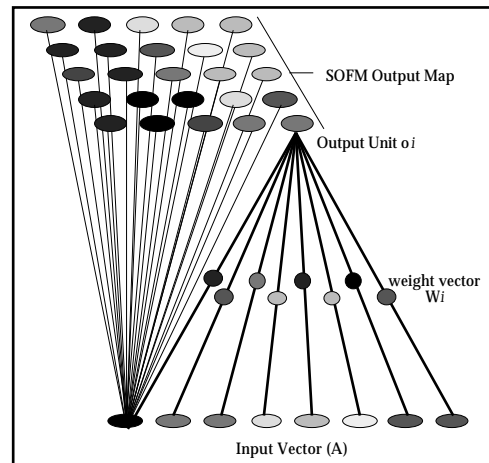

Figure 1 A SOFM connected to an input vector (A) by weight vector (W_i). This diagram shows (i) how each SOFM output unit connects to every input unit and (ii) how each output unit will have its own n-dimensional weight vector, where n is the number of values in the input vector. Greyscale information depicts variability in both unit activation and weight values.

The training rule updates the weights of the winning unit, and also those of a neighbourhood surrounding the winner, moving them closer to the input vector. The neighbourhood is characterised by a Gaussian function such that weight vectors of units *closer* to the winner tend to approximate the input vector to a greater degree than those further away. It is these correlated zones of activation that facilitate self-organisation (Kohonen, 1982a; Schynns, 1991). As training time (t) increases, the neighbourhood size decreases linearly such that, eventually, only the winning unit's weight vector is updated. Initially, to ensure global order, each unit's neighbourhood is larger than one half of the output map but is reduced over the full epoch time. Similarly, to ensure learning stability, the training rate also declines over t. In our simulations, SOFMs were trained over 1500 epochs with an initial neighbourhood of size 6 by 6 units and an initial learning rate of 0.5. Presentation of training set exemplars was randomised within each epoch. The neighbourhood size and learning rate decreased linearly every 250 and 150 epochs respectively. In summary, our model *reduces the dimensionality* of the input vector but

preserves *training pattern topology*. Images that are close in Euclidean distance will, therefore, generate SOFM contour patterns that are similar.

4. THE TRAINING IMAGES

Our training images were derived from the 4 categories of Animals, Musical Instruments, Clothing and Furniture. All selected images were colour photographs that depicted an object or animal in a typical orientation. Seven basic level categories were chosen for each of the 4 superordinate categories and, for each chosen basic-level category, 5 images were selected that depicted different but not atypical subordinate members (e.g. for the basic category of *Fishes* we selected salmon, carp, pike, bass and cod). Thus, each superordinate category was represented by 35 different subordinate images.

Each image was scanned at a resolution of 100dpi. All images were then edited to (i) remove any background detail, (ii) convert to 8-bit greyscale and (iii) reduce in size such that each object's maximal dimension fitted exactly within a 50 by 50 pixel grid. Some example images of the basic level category 'clock' are displayed in figure 2.
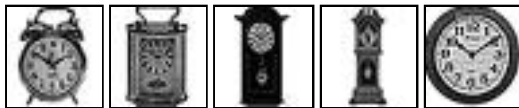

Figure 2 **Examples of greyscale images presented to our model. Each image fits within a 50 by 50 pixel grid such that the principal dimension comes within one pixel of the grid border. These examples depict the 5 subordinate images of the basic level category 'clock' which, in turn, is one of the 7 basic level categories representing furniture.**

Before presenting the images to our model, two additional control measures were undertaken. Firstly, a left-right inversion of each image was created to overcome the fact that some asymmetrical objects (e.g. fishes) tended always to be depicted in the same orientation (e.g. with head to the left and tail to the right), whereas others (e.g. chairs) had a more variable orientation. Secondly, for each left-right inversion, a novel image was created by shifting the distribution of greyscale values by ±20%. Images whose originals were subjectively judged to be dark were made lighter, whilst images that were judged to be light were made darker. These manipulations generated 4 versions of each subordinate exemplar, increasing the total image set size to 560.

5. THE EXPERIMENT

Method Images were presented randomly over 1500 epochs to ten randomly configured SOFMs. Activation values (in the range 0-255) were recorded for all 100 output units in each of the 10 SOFMs, for each training image in turn. For analytical purposes, the possible activation values (i.e. integers in the range 0 to 255) were grouped into 5 ordinal bands as follows: (1) 0-49, (2) 50-99, (3) 100-149, (4) 150-199, (5) 200-255. The frequencies of unit activation values falling within each band were recorded for each training exemplar, averaged across all 10 SOFMs. These values were then averaged across the 4 superordinate classes. Average distribution of unit activation values for each category is displayed in figure 3.
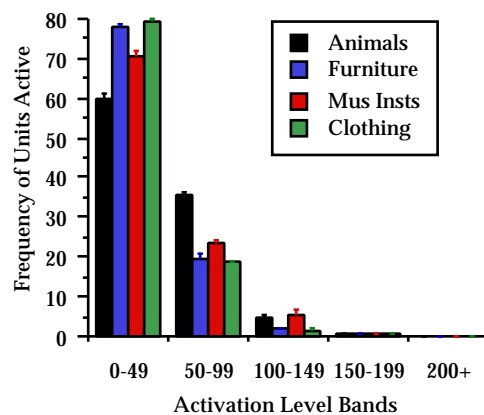

Figure 3 **The distribution of mean activation values across 10 SOFMs for images from the 4 superordinate categories. Error bars display the standard deviation across SOFM.**

Results There were significant differences between categories in the frequencies of activation values under all 5 bands (For 0-49, $F[3,559] = 48.2$, $p < 0.0001$; For 50-99, $F[3, 559] = 46.3$, $p < 0.0001$; For 100-149, $F[3, 559] = 43.4$, $p < 0.0001$; For 150-199, $F[3,559] = 2.7$, $p < 0.05$; For 200+, $F[3,559] = 6.2$, $p < 0.0005$). Post-hoc tests confirmed that *both* musical instruments and animals generated significantly less activation values in the 0-49 band ($p < 0.0001$) and significantly more activation values in the 50-99 and 100-149 bands ($p < 0.001$). The 'visual' impact of these differences is more clearly illustrated in figure 4, which displays the SOFM representations of four training images. Light and dark regions indicate high and low levels of activation respectively.
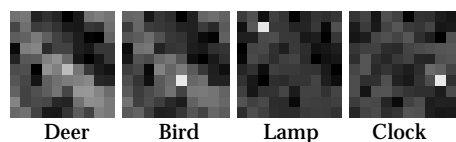
| Deer | Bird | Lamp | Clock |

Figure 4 Examples of SOFM output contour maps for 4 training patterns. Note how the furniture training patterns generate sharper peaked contour maps than the animal training patterns which tend to generate plateaux of moderate activation.

6. DISCUSSION

The fact that an unsupervised neural network appears able to distinguish broad category boundaries in a reasonably large set of visual stimuli would suggest that visual properties may underlie or, at least contribute to, the emergence of category-specific deficits after brain damage. Of particular relevance is the fact that musical instrument representations are qualitatively similar to animal representations: in our model, furniture and clothing exemplars tended to generate SOFM representations that were highly distinctive (i.e. each image was associated with a different 'winning unit'). This is indicative that images from these categories do not share much perceptual overlap. Conversely, animal and musical instrument exemplars tended to generate SOFM representations that were broadly similar (i.e. regions of higher activity were 'shared' amongst many exemplars) suggesting that both these categories are 'visually crowded' (Gaffan and Heywood, 1993). The co-occurrence of recognition deficits for both living things and musical instruments has always been problematic for theories of category-specific recognition deficits. We contend that our model offers the most parsimonious explanation for why these deficits co-occur and suggest that simple visual similarities may have an important role to play.

9. REFERENCES

Caramazza, A., and Shelton, J.R. (1998) Domain-specific knowledge in the brain: The animate-inanimate distinction. Journal of Cognitive Neuroscience, 10:1, 1-34.

Farah, M.J., Meyer, M.M., and McMullen, P.A. (1996) The living non-living dissociation in not an artefact: Giving an a-priori implausible hypothesis a strong test. *Cognitive Neuropsychology, 13,* 137-154.

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.

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

Kohonen, T (1982b) Clustering, taxonomy and topological maps of patterns. in M. Land (Ed.), *Proceedings of the Sixth International Conference on Pattern Recognition.* Silver Spring, MD: IEEE Computer Society Press.

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

Sacchett, C., and Humphreys, G.W. (1992) Calling a squirrel a squirrel but a canoe a wigwam: A category-specific deficit for artefactual objects and body parts. *Cognitive Neuropsychology, 9,* 73-86.

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

Stewart, F., Parkin, A.J., and Hunkin, N.M (1992) Naming impairments following recovery from Herpes Simplex Encephalitis: Category-specific? *Quarterly Journal of Experimental Psychology, 44A,* 261-284.

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