

Local vs distributed: a poor taxonomy of neural coding strategies

M. W. Spratling

Centre for Brain and Cognitive Development, Birkbeck College, London. UK.

October 29, 2001

Abstract

Page (2000a) is to be congratulated for challenging some misconceptions about neural representation. However, his target article, and the commentaries to it, highlight that the terms 'local' and 'distributed' are open to misinterpretation. These terms provide a poor description of neural coding strategies and a better taxonomy might resolve some of the issues.

Consider a neural network in which individual nodes represent single letters of the alphabet. When a letter is presented to this network a node responds providing a localist representation. However, if two (or more) letters are simultaneously presented, two (or more) nodes will be active and the representation would now be described as distributed. The same network thus generates local or distributed representations in different circumstances. Page (2000b) makes this point himself: "under the appropriate task demands, such a layer might process a stimulus as a distributed pattern of activation across localist representations" (p. 497). Thus local or distributed representations are not a property of the network (alone) and it is questionable whether such labels are useful in describing models. Furthermore, there are two dimensions along which neural coding strategies need to be classified: the tuning properties of the nodes and the number of nodes involved in the representation (Jelasy 2000).

Tuning properties. In the above example of a network representing letters, the preferred stimulus of each individual node has a clear interpretation. Such a node might be narrowly tuned to a specific stimulus or it might be widely tuned responding over a range of inputs with varying degrees of similarity to the preferred stimulus. Such a node might respond to the same letter written in a variety of fonts and/or to a specific letter appearing at any location within the receptive field. However, in all of these cases it is possible to ascribe meaning to the node's activity in isolation: it acts as a feature detector or a matched filter (Barlow and Gardner-Medwin, 2000). An alternative response property which has been described as '(fully/dense) distributed', 'compact' or 'ensemble' coding requires unique combinations of active nodes to represent meaningful features of the environment. In such a scheme the responses of isolated nodes do not have an obvious interpretation (Page, 2000a). An example is the binary ASCII code in which any individual bit will be ON for an arbitrary set of characters (Barlow and Gardner-Medwin, 2000).

A feature detector responds to a set of stimuli to which we can apply a meaningful label, while in a compact code a node responds to stimuli for which there is no such simple classification (Hummel, 2000). That the distinction is purely one of interpretation should come as no surprise. Similar activation functions (weighted summation followed by thresholding) are used in both cases, and similar learning rules can also be used (pseudo-Hebbian learning can generate compact codes such as Principal Components (Oja, 1982; Sanger, 1989), but is also widely used to learn feature detectors (Földiák, 1990; Wallis and Rolls, 1997)). The arbitrariness of the distinction is further illustrated by differing interpretations of the response properties of simple cells in the primary visual cortex. The receptive fields of these cells have been shown to be similar to the independent components of natural images (Bell and Sejnowski, 1997; van Hateren and Ruderman, 1998), however, they are also routinely described as edge detectors. Physiological evidence for correlations between the activity of individual neurons and behaviour (Georgopoulos et al., 1986; Newsome et al., 1989), and between the activity of individual neurons and sensory stimuli (Logothetis and Sheinberg,

1996; Perrett et al., 1992; Tanaka, 1996) suggests that cortical cells are generally tuned to meaningful categories. Such tuning properties may be readily learnt since re-occurring patterns of stimuli are likely to correspond with features of the environment (Barlow, 1972, 1994; Edelman and Duvdevani-Bar, 1995).

Coding density. For the network responding to letters, described above, individual letters are ‘explicitly’ represented whereas combinations of letters are ‘implicitly’ represented. Similarly, in a hierarchical network information implicitly represented by the response of many nodes at one level may be explicitly represented by the response of individual nodes at a higher level. A single node thus explicitly represents a particular feature of the input stimulus and it may do so independently of the activity of other nodes or as part of an implicit representation of the stimulus as a whole.

The cortex explicitly and implicitly encodes information. For example, cells in area V1 provide explicit representations of oriented edge features at specific locations. However, they must also provide an implicit representation of all visual stimuli that can be distinguished by the visual system and which may (or may not) be explicitly represented in higher cortical areas. Since all visual information is available in V1 (and for that matter in the LGN and at the retina) it must be advantageous to recode information so that more abstract entities become more explicitly represented. The appropriate level of abstraction will vary between tasks: it would be equally impractical to explicitly represent every possible event using single nodes as it would be to use a representation in which all events were only implicitly represented using low-level feature detectors (Feldman, 1990; Tsotsos, 1995; Wilson, 1991). Furthermore, such recoding makes learning tractable by transforming complex, relational, tasks into simpler, statistical, problems over the recoded data (Clark and Thornton, 1997). What information is explicitly represented should depend on the environment and the importance of those stimuli to the task/animal (Logothetis 1998).

Conclusions. Describing networks along a single local vs distributed dimension fails to make explicit the properties of the coding strategy employed. Specifically, Page (2000a) uses the term local to describe tuning properties without making a commitment to the coding density, while he uses the term distributed to refer to the coding density without specifying tuning properties. He is therefore correct in his assertion that these terms are not dichotomous but only because they are being used to describe independent properties.

While most models employ nodes with similar tuning properties there are significant differences between models in terms of coding density. Constraints on the sparsity or density of activity influence whether information is encoded explicitly or implicitly. For example, a winner-takes-all (WTA) network will need to explicitly encode each stimulus that is to be represented, while a k-WTA network will be biased towards forming implicit representations. Such constraints are commonly employed in unsupervised learning algorithms resulting in networks unable to cope with tasks that require an arbitrary number of active nodes: even tasks as simple as responding to single or multiple letters. Such constraints result in poor models of cortical coding and may have contributed to the delusion that the cortex is also committed to using either local or distributed representations.

ACKNOWLEDGEMENT

This work was funded by MRC Research Fellowship number G81/512.

References

- Barlow, H. and Gardner-Medwin, A. (2000). Localist representation can improve efficiency for detection and counting. *Behavioural and Brain Sciences*, 23(4):467–8.
- Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1:371–94.
- Barlow, H. B. (1994). What is the computational goal of the neocortex? In Koch, C. and Davis, J. L., editors, *Large-Scale Neuronal Theories of the Brain*, chapter 1. MIT Press, Cambridge, Massachusetts; London.
- Bell, A. J. and Sejnowski, T. J. (1997). The ‘independent components’ of natural scenes are edge filters. *Vision Research*, 37:3327–38.

- Clark, A. and Thornton, C. (1997). Trading spaces: computation, representation and the limits of uninformed learning. *Behavioural and Brain Sciences*, 20(1):57–66.
- Edelman, S. and Duvdevani-Bar, S. (1995). Similarity, connectionism, and the problem of representation in vision. *Neural Computation*, 9:701–20.
- Feldman, J. A. (1990). Computational constraints on higher neural representations. In Schwartz, E. L., editor, *Computational Neuroscience*. MIT Press, Cambridge, Massachusetts; London.
- Földiák, P. (1990). Forming sparse representations by local anti-hebbian learning. *Biological Cybernetics*, 64:165–70.
- Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233:1416–9.
- Hummel, J. E. (2000). Localism as a first step toward symbolic representation. *Behavioural and Brain Sciences*, 23(4):480–1.
- Jelasy, M. (2000). Instance-based manifesto? *Behavioural and Brain Sciences*, 23(4):482–3.
- Logothetis, N. (1998). Object vision and visual awareness. *Current Opinion in Neurobiology*, 8:536–44.
- Logothetis, N. and Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, 19:577–621.
- Newsome, W. T., Britten, K. H., and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341:52–4.
- Oja, E. (1982). A simplified neuron model as a principal component analyser. *Journal of Mathematical Biology*, 15:267–73.
- Page, M. (2000a). Connectionist modelling in psychology: a localist manifesto. *Behavioural and Brain Sciences*, 23(4):443–67.
- Page, M. (2000b). Sticking to the manifesto. *Behavioural and Brain Sciences*, 23(4):496–505.
- Perrett, D. I., Hietanen, J. K., Oram, M. W., and Benson, P. J. (1992). Organisation and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London*, 335:23–30.
- Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–73.
- Tanaka, K. (1996). A continuous map of higher-level visual features of objects in monkey inferotemporal cortex. In Konnerth, A., Tsien, R., Mikoshiba, K., and Altman, J., editors, *Coincidence Detection in the Nervous System*, pages 143–51. HFSP, Strasbourg.
- Tsotsos, J. K. (1995). Behaviourist intelligence and the scaling problem. *Artificial Intelligence*, 75:135–60.
- van Hateren, J. H. and Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B*, 265:2315–20.
- Wallis, G. and Rolls, E. T. (1997). A model of invariant object recognition in the visual system. *Progress in Neurobiology*, 51:167–94.
- Wilson, S. W. (1991). The animat path to AI. In Meyer, J.-A. and Wilson, S. W., editors, *From Animals to Animats: Proceedings of the First International Conference on the Simulation of Adaptive Behaviour (SAB91)*, pages 15–21. MIT Press.