The local is running on the express track: Localist models better facilitate understanding of nervous system function

Paul A. Koch[a] and Gerry Leisman[b]

[a]*School of Engineering, New York Institute of Technology, Old Westbury, NY 11568;* [b]*Carrick Institute for Clinical Ergonomics, Rehabilitation and Applied Neurosciences, School of Engineering, State University of New York at Farmingdale, Farmingdale, NY 11735-1021.* **pkoch@liii.com**
leismag@farmingdale.edu

Abstract: Artificial neural networks have weaknesses as models of cognition. A conventional neural network has limitations of computational power. The localist representation is at least equal to its competition. We contend that locally connected neural networks are perfectly capable of storing and retrieving the individual features, but the process of reconstruction must be otherwise explained. We support the localist position but propose a "hybrid" model that can begin to explain cognition in anatomically plausible terms.

Whether they use localist or distributed representations, artificial neural networks as they are currently conceived have a serious weakness if they are able to be applied as models of natural cognition. It is well known from anatomical and physiological studies that neurons have a finite connection range beyond which the probability of connection of two cells decreases exponentially. This range is about 250 μ in the visual cortex (Krone et al. 1986) and 400 μ in the hippocampus (Traub et al. 1988). On the other hand, neural networks are constructed so that almost all cells in a given layer are connected to all the cells in some other layer. This situation could only be approximately applicable if all the cells fell within a single connection range. Using the figures of Wilson and Cowan (1973), about 10,000 cells could fit within a cylinder of radius of one connection length and thus constitute such a locally *connected* network.

In contrast, the cortex of the edible frog has about 6 million cells (Wilson & Cowan 1973). Making all allowances for the efficiency of scientifically constructed learning algorithms compared with what has been contrived by nature, it is not likely that a conventional neural network operating under the constraints of brain anatomy could have computational power greater than that of a frog. This is not to detract from the demonstrated abilities of neural networks whose nodes and/or activation states are given semantic significance; we simply question their direct relevance to animal psychology.

For the purposes of this discussion, we accept the arguments of the author that the localist representation is at least equal to its competition. The first question we must ask is how much information an actual neuron can store. There might very well be a "grandmother cell" because of the childhood implantation, but it is highly unlikely that anyone other than immediate family has a "Mike Page" cell. Yet we are capable of recognizing numerous relative strangers, largely through the reconstruction of distributively stored details (Bartlett 1932). We contend that locally connected neural networks are perfectly capable of storing and retrieving the individual features, but the process of reconstruction must be otherwise explained.

A rat negotiating a maze, for example, has memory known to be implanted through the action of single hippocampal "place" and "direction" cells (Muller et al. 1996; Rolls et al. 1998). The choice of direction at any point, however, must be assembled from this stored fragmentary data. Of particular significance is the recent work of Fried and colleagues (1997) who measured the activity of single hippocampal neurons during recognition tests. They found that some neurons responded differentially to faces and inanimate objects, some to different emotional expressions, and some to combinations of expression and gender. This would be, in our opinion, the right amount of content a node of an anatomically constrained localist neural network could hold.

What mechanism can connect the elementary localist net-

works? Perhaps a network that takes Euclidean distance into account can be devised. None exists as far as we know, so questions about how ability is degraded by systematic incompleteness of connections cannot be answered. There have been some simulations incorporating connection range (Krone et al. 1986; Traub et al. 1987; 1988; 1989), but these have not been given any cognitive content; rather, they are concerned with the general behavior of the neural medium.

This area of study was first suggested by Wilson and Cowan (1973), who created what can be called a continuum theory of neural tissue. The differential elements of this medium contain many cells; in modern terms they can be called *neural networks*, although such terms as "cell assemblies" (Hebb 1949) or "groups" have been applied. These elements may or may not be functional wholes; their constituent cells might change with task. However, it is logical to give each, at any time, the computational abilities of a localist neural network.

According to the theory, the elements are interconnected by probabilistically distributed connections of type e-e, e-i, i-e, and i-i, between (literal or figurative) "excitatory" and "inhibitory" "cell" species. The existence of e-e connections signifies an internal source of (electrochemical) energy, so that the medium can support growing disturbances. The conditions of growth give rise to waves with sharply defined wavelength spectra (Koch & Leisman 1990; 1996; 2001; Leisman & Koch 2000). The preferred wavelengths are controlled by the synaptic parameters, and they are always significantly longer than the largest average synaptic connection range.

We postulate that this phenomenon provides a mechanism by which the active nodes of the elementary networks can be "connected." Unlike a neural network, the connection is not causal; instead, two nodes are simultaneously active when the distance between them is an integral multiple of the favored wavelength, which can be changed by changes in the state of the relevant brain region.

We have applied this concept to models of the brain stem, cortex, and hippocampus (Koch & Leisman 1990; 1996; 2001; Leisman & Koch 2000). The last studies are of particular interest for the present discussion. The growth properties of hippocampal waves change as synapses are modified by Hebbian conditioning (Hebb 1949; 1972). Such a process exists in the hippocampus (Eccles 1986), manifested by strengthening in synapses between geometrically and synaptically neighboring cells. This is certainly akin to local weight modification in neural networks, but it has important global effects. Not only does it change the nature of the waves from decay to growth, but it also controls the amplified wavelength and therefore the connections among the networks.

Hence our title: The local ultimately ramifies throughout cognitive regions. We hope to see in the future a "hybrid" model that can begin to explain cognition in anatomically plausible terms.

Local versus distributed: A poor taxonomy of neural coding strategies

Michael W. Spratling

Centre for Brain and Cognitive Development, Birkbeck College, London, WC1E 7JL, United Kingdom. **m.spratling@bbk.ac.uk**

Abstract: Page 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. In his Response to the commentaries, 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). On that ground, 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 (Jelasity 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 or to a specific letter appearing at any location within the receptive field, or to both. 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 & 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 & Gardner-Medwin 2000).

A feature detector responds to a set of stimuli to which we can apply a meaningful label, but 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 (e.g., weighted summation followed by thresholding) are used in both cases, and similar learning rules can also be used; that is, 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 & 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 & Sejnowski 1997; van Hateren & 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 & 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 recurring patterns of stimuli are likely to correspond to features of the environment (Barlow 1972; 1994; Edelman & 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 the 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 therefore 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 & 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 versus 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, and 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.

Although 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.

Editors' Note: There is no Author's Response to this commentary.

References

Barlow, H. B. (1972) Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* 1:371–94. [MWS]
(1994) What is the computational goal of the neocortex? In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis, Ch. 1. MIT Press. [MWS]
Barlow, H. B. & Gardner-Medwin, A. (2000) Localist representation can improve efficiency for detection and counting. *Behavioral and Brain Sciences* 23:467–68. [MWS]
Bartlett, F. E. (1932) *Remembering.* Cambridge University Press. [PAK]
Bell, A. J. & Sejnowski, T. J. (1997) The independent components of natural scenes are edge filters. *Vision Research* 37:3327–38. [MWS]
Clark, A. & Thornton, C. (1997) Trading spaces: Computation, representation and the limits of uninformed learning. *Behavioral and Brain Sciences* 20:57–66. [MWS]
Eccles, J. C. (1986) Mechanism of learning in complex neural systems. In: *Handbook of physiology: The nervous system V,* ed. F. Plum, pp. 137–67. Williams & Wilkins. [PAK]
Edelman, S. & Duvdevani-Bar, S. (1995) Similarity, connectionism, and the problem of representation in vision. *Neural Computation* 9:701–20. [MWS]

Continuing Commentary

Feldman, J. A. (1990) Computational constraints on higher neural representations. In: *Computational neuroscience,* ed. E. L. Schwartz. MIT Press. [MWS]

Földiák, P. (1990) Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics* 64:165–70. [MWS]

Fried, I., MacDonald, K. A. & Wilson, C. L. (1997) Singe neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18:735–65. [PAK]

Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. (1986) Neuronal population coding of movement direction. *Science* 233:1416–19. [MWS]

Hebb, D. O. (1949) *The organization of behavior.* Wiley. [PAK]

  (1972) *Textbook of psychology.* W. B. Saunders. [PAK]

Hummel, J. E. (2000) Localism as a first step toward symbolic representation. *Behavioral and Brain Sciences* 23:480–81. [MWS]

Jelasity, M. (2000) Instance-based manifesto? *Behavioral and Brain Sciences* 23:482–83. [MWS]

Koch, P. & Leisman, G. (1990) A continuum model of activity waves in layered neuronal networks: A neuropsychology of brain-stem seizures. *International Journal of Neuroscience* 54:41–62. [PAK]

  (1996) Wave theory of large-scale organization of cortical activity. *International Journal of Neuroscience* 86(3–4):179–96. [PAK]

  (2001) Effect of local synaptic strengthening on global activity-wave growth in the hippocampus. *International Journal of Neuroscience* 108(1–2):127–46. [PAK]

Krone, G., Mallot, H., Palm, G. & Schuz, A. (1986) Spatiotemporal receptive fields: A dynamical model derived from cortical architectonics. *Proceedings of the Royal Society of London, B: Biological Sciences* 226:421–44. [PAK]

Leisman, G. & Koch, P. (2000) Continuum model of mnemonic and amnesiac phenomena. *Journal of the International Neuropsychological Society* 6:589–603. [PAK]

Logothetis, N. (1998) Object vision and visual awareness. *Current Opinion in Neurobiology* 8:536–44. [MWS]

Logothetis, N. & Sheinberg, D. L. (1996) Visual object recognition. *Annual Review of Neuroscience* 19:577–621. [MWS]

Muller, R. U., Ranck, J. B., Jr. & Taub, J. S. (1996) Head direction cells: Properties and functional significance. *Current Opinion in Neurobiology* 6:196–206. [PAK]

Newsome, W. T., Britten, K. H. & Movshon, J. A. (1989) Neuronal correlates of a perceptual decision. *Nature* 341:52–54. [MWS]

Oja, E. (1982) A simplified neuron model as a principal component analyser. *Journal of Mathematical Biology* 15:267–73. [MWS]

Page, M. (2000a) Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences* 23:443–67. [MWS]

  (2000b) Sticking to the manifesto. (Author's Response to commentary.) *Behavioral and Brain Sciences* 23:496–505. [MWS]

Perrett, D. I., Hietanen, J. K., Oram, M. W. & 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. [MWS]

Rolls, E. T., Treves, A., Robertson, R. G., Georges-Francois, P. & Panzeri, S. (1998) Information about spatial view in an ensemble of primate hippocampal cells. *Journal of Neurophysiology* 79(4):1797–813. [PAK]

Sanger, T. D. (1989) Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks* 2:459–73. [MWS]

Tanaka, K. (1996) A continuous map of higher-level visual features of objects in monkey inferotemporal cortex. In: *Coincidence detection in the nervous system,* ed. A. Konnerth, R. Tsien, K. Mikoshiba & J. Altman, pp. 143–51. Human Frontier Science Program. [MWS]

Traub, R. D., Miles, R. & Wong, R. K. S. (1987) Models of synchronized hippocampal bursts in the presence of inhibition. I. Single population events. *Journal of Neurophysiology* 58:739–51.

  (1988) Large scale simulations of the hippocampus. *IEEE Engineering in Medicine and Biology Magazine* 7:31–38.

  (1989) Model of the origin of rhythmic population oscillations in the hippocampal slice. *Science* 243:1319–25.

Traub, R. D., Whittington, M. A., Colling, S. B., Buzaki, G. & Jefferys, J. G. R. (1990) Analysis of gamma rhythms in the rat hippocampus *in vitro* and *in vivo. Journal of Physiology* (London) 493:471–84.

Tsotsos, J. K. (1995) Behaviourist intelligence and the scaling problem. *Artificial Intelligence* 75:135–60. [MWS]

van Hateren, J. H. & 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. [MWS]

Wallis, G. & Rolls, E. T. (1997) A model of invariant object recognition in the visual system. *Progress in Neurobiology* 51:167–94. [MWS]

Wilson, H. R. & Cowan, J. D. (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetic* 13:55–80. [MWS]

Wilson, S. W. (1991) The animat path to AI. In: *From animals to animats: Proceedings of the First International Conference on the Simulation of Adaptive Behaviour (SAB91),* ed. J.-A. Meyer & S. W. Wilson, pp. 15–21. MIT Press. [MWS]

Commentary on **J. Allan Hobson, Edward F. Pace-Schott, & Richard Stickgold (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. BBS 23(6):793–842.**

Abstract of the original article: Sleep researchers in different disciplines disagree about how fully dreaming can be explained in terms of brain physiology. Debate has focused on whether REM sleep dreaming is qualitatively different from nonREM (NREM) sleep and waking. A review of psychophysiological studies shows clear quantitative differences between REM and NREM mentation and between REM and waking mentation. Recent neuroimaging and neurophysiological studies also differentiate REM, NREM and waking in features with phenomenological implications. Both evidence and theory suggest that there are isomorphisms between the phenomenology and the physiology of dreams. We present a three-dimensional model with specific examples from normally and abnormally changing conscious states.

Drug induced alterations in dreaming: An exploration of the dream data terrain outside activation-synthesis

Jim F. Pagel

Department of Family Practice, University of Colorado Medical School, Pueblo, CO 81003. **pueo34@juno.com**

Abstract: Two meta-analyses of pharmacological research are presented, demonstrating that psychoactive drugs have consistent effects on EEG and sleep outside of their effects on REM sleep, and demonstrating that drugs other than those affecting sleep neurotransmitter systems and REM sleep can also alter reported nightmare occurrence. These data suggest that the neurobiology data terrain outside activation-synthesis may include sleep and dream electrophysiology, cognitive reports of dreaming, effects of alterations in consciousness on dreaming, immunology and host defense, and clinical therapies for sleep disorders.

The most accepted approach to addressing the obvious complexity of known components of central nervous system (CNS) electrophysiology and neurochemistry is to approach analysis of the system deductively, using selected data from many different areas to support a theoretical construct. Unfortunately, if this approach is utilized to present a purportedly broad-based review for prospective theorists in the field, data that are inconsistent or noncontributory to that theoretical construct (an amended activation-synthesis hypothesis) are excluded and ignored. This approach attempts to guarantee that future researchers and theorists in sleep and dreaming will work within the constraints of that model – a model currently requiring extensive restructuring and remodeling to encompass the experimental data of its supporters (see Nielsen 2003; Hobson et al. 2003). This commentary presents pharmacological data from two of these excluded areas: (i) psychoactive drug alteration of sleep stages and background EEG frequencies inclusive of alterations in REM sleep (REMS), and