

Behavioral and Brain Sciences

**Combining distributed and localist computations
in real-time neural networks**

**A commentary on: “Connectionist modelling in psychology:
A localist manifesto,” by Mike Page**

Gail A. Carpenter

*Department of Cognitive and Neural Systems, 677 Beacon Street, Boston University,
Boston, Massachusetts 02215. gail@cns.bu.edu <http://cns.bu.edu/~gail/>*

November, 1999

Technical Report CAS/CNS TR-99-028
Boston, MA: Boston University

Abstract: In order to benefit from the advantages of localist coding, neural models that feature winner-take-all representations at the top level of a network hierarchy must still solve the computational problems inherent in distributed representations at the lower levels.

By carefully defining terms, demonstrating strong links among a variety of seemingly disparate formalisms, and debunking purported shortcomings of winner-take-all systems, Page has made a significant contribution toward the creation of a functional classification of the growing array of neural and cognitive models. One important feature of the target article is a clarification of terminology. For example, a model is here labeled “localist” when the representation at the top level (n) of a network hierarchy is localist (section 2.6, paragraph 1). This definition is based on the logical conclusion that, once a code representation has reached the limit of winner-take-all compression, additional network levels would be redundant. Conversely, any non-redundant localist system would normally have distributed representations at the lower levels $1 \dots n-1$. By considering systems in their hierarchical configurations, Page shows that models and related data previously viewed as “distributed” in fact derive essential properties from localist mechanisms.

Page’s hierarchical definition of localist networks implies that any such system with more than two levels could inherit the computational drawbacks, as well as the benefits, of distributed networks. As Page points out (section 7.1), many distributed models are subject to catastrophic interference and require slow learning and multiple interleaved presentations of the training set. One of my research goals in recent years has been the development of real-time neural network systems that seek to combine the computational advantages of fully distributed systems such as multilayer perceptrons (Rosenblatt, 1958, 1962; Rumelhart, Hinton, & Williams, 1986; Werbos, 1974) with the complementary advantages of localist systems such as adaptive resonance theory (ART) networks (Carpenter & Grossberg, 1987, 1993; Carpenter, Grossberg, & Reynolds, 1991; Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992). An initial product of this ongoing project was the distributed ART (dART) family of neural networks (Carpenter, 1996, 1997; Carpenter, Milenova, & Noeske, 1998), which permit fast as well as slow learning, and distributed as well as localist code representations, without catastrophic forgetting. Where earlier ART models, in order to help stabilize memories, employed strongly competitive activations to produce winner-take-all coding, dART code representations may be distributed across any number of nodes. In order to achieve its computational goals, the dART model includes a new configuration of the network architecture, and replaces the traditional path weight with a *dynamic weight*, which is a joint function of current coding node activation and long-term memory (LTM). The

dART system also employs new learning rules, which generalize the instar (equation (10), section 4.4, paragraph 3) to the case where the target node activation patterns at layer L_2 may be fully distributed. The original instar equation implies that, unless learning is very slow, all weight vectors \mathbf{w}_j would converge to the same input pattern \mathbf{a} at every location where the target L_2 node is active ($a_j > 0$). With the distributed instar learning rule, dynamic weights automatically bound the sum of all LTM changes, even with fast learning. The computational innovations of the dART network would allow distributed representations to be incorporated at levels $1 \dots n-1$ in a network hierarchy while retaining the benefits of localist representations at level n .

In contrast to the aim of the dART research program, which is to define a real-time, stand-alone neural network with specified properties, the primary aim of the target article is to unify diverse computational and conceptual themes. In the service of this goal, the corresponding learning module (section 4.1) is, by design, skeletal. However, such a partially specified model might risk being unduly rejected on the basis of what it seems not to do, and some of the model's properties are subject to misinterpretation if taken at face value. For example, Page's localist model permits learning only at an uncommitted node, which then encodes the current input. The decision whether to activate an uncommitted node depends upon the value of the threshold \square , which is somewhat analogous to the vigilance matching parameter \sqcup in an ART model. In particular: "If the threshold is set slightly lower [than 1], then only activation patterns sufficiently different from previously presented patterns will provoke learning." (section 4.1, paragraph 2) Page points out that this construction would help solve the problem of catastrophic interference, since coding a new pattern does not affect previous learning at all. On the other hand, this feature might also be the basis for rejecting this model, and by extension other localist models, since each category can be represented only as a single exemplar: there is no opportunity for new exemplars that correctly activate a given category to refine and abstract the initial learned representation. In contrast, a more fully specified localist model could permit controlled learning at committed nodes as well as at uncommitted nodes, hence creating prototype as well as exemplar memories while still retaining the ability to resist catastrophic interference. Even though this capability is not part of Page's simplified model, the possibility of learning at committed nodes is implied later in the article (section 4.5, paragraph 3): "...when at least one of the associates is learned under low-vigilance (cf. prototype) conditions, remapping of items to alternative associates can be quickly achieved by rapid reconfiguration of connections to and from the mapping layer."

Similarly, a reader may be misled who takes seriously the assertion: "The extension [of the learning module in the target article] to continuous activations will

usually be necessary and is easily achieved.” (section 4.1, paragraph 1) This statement is true, but defining an extension of the simplified system is not a matter of straightforward substitution. In particular, the learning module is defined only for the case of binary inputs, and the validity of its computational properties relies implicitly on the assumption that $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}| = \|\mathbf{a}\|^2$, which is true only when \mathbf{a} is binary.

In summary, the simplified localist learning module defined by Page is a valuable tool for unifying and clarifying diverse formalisms, but a more complete computational development is needed to define stand-alone neural network systems that realize the promise of the localist analysis.

References

- Carpenter, G.A. (1996). Distributed activation, search, and learning by ART and ARTMAP neural networks. *Proceedings of the International Conference on Neural Networks (ICNN'96): Plenary, Panel and Special Sessions* (pp. 244-249). Piscataway, N.J.: IEEE Press.
- Carpenter, G.A. (1997). Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks*, *10*, 1473-1494.
- Carpenter, G.A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, *37*, 54-115.
- Carpenter, G.A., & Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neuroscience*, *16*, 131-137.
- Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., & Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, *3*, 698-713.
- Carpenter, G.A., Grossberg, S., & Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, *4*, 565-588.
- Carpenter, G.A., Milenova, B.L., & Noeske, B.W. (1998). Distributed ARTMAP: a neural network for fast distributed supervised learning. *Neural Networks*, *11*, 793-813.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*, 386-408.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. Washington, D.C.: Spartan Books.
- Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition*, I (pp. 318-362). Cambridge, Mass.: MIT Press.
- Werbos, P.J. (1974). Beyond regression: new tools for prediction and analysis in the behavioral sciences. Unpublished Ph.D. thesis. Cambridge, Mass.: Harvard University.