Can neural models of cognition benefit from the advantages of connectionism?

Article *in* Behavioral and Brain Sciences · February 2006

DOI: 10.1017/S0140525X06379022

Friedrich T Sommer
University of California, Berkeley

Commentary on "Neural blackboard architectures of combinatorial structures in cognition" by Frank van der Velde and Marc de Kamps.

Abstract: *73* words
Main Text: *1026* words
References: *993* words
Total Text: *2092* words

Can neural models of cognition benefit from the advantages of connectionism?

Friedrich T. Sommer
Redwood Neuroscience Institute & UC Berkeley
1010 El Camino Real
Menlo Park, CA 94025
USA
+1 650 321 8282 Ext 241
fsommer@berkeley.edu
http://www.rni.org/ftsommer/FS/FTSommer.html

Pentti Kanerva
Redwood Neuroscience Institute
1010 El Camino Real
Menlo Park, CA 94025
USA
+1 650 321 8282 Ext 232
pkanerva@rni.org
http://www.rni.org/kanerva/homepg.html

Abstract
Cognitive function certainly poses the biggest challenge for computational neuroscience. As we will argue, past efforts to build neural models of cognition (this target article included) had a too narrow focus on implementing rule-based language processing. The problem with these models is that they sacrifice the advantages of connectionism rather than building on them. Recent and more promising approaches for modeling cognition build on the mathematical properties of distributed neural representations. These approaches truly exploit the key advantages of connectionism, that is, the high representational power of distributed neural codes and similarity-based pattern recognition. The architectures for cognitive computing that emerge from these approaches are neural associative memories endowed with additional mapping operations to handle invariances and to form reduced representations of combinatorial structures.

Introduction
There is an abundance of cognitive phenomena that our existing models are unable to explain. Jackendoff (2002) has singled out four from language, and the target article proposes to solve them with a neural blackboard architecture. How likely are we to come up with the right cognitive architecture by

starting with language? While language is a cognitive function without an equal, it is only a fraction of what makes up human cognition. It is also evolution's latest invention. In keeping with the principles of biological design, it rests fundamentally on earlier inventions that are found in mammals and many other animals. These include rich sensor integration, complex learned motor-action sequences, memory for places and things, learned judgment, memory for learned actions, mental imagery (imagination), learning from example, and signaling. In other words, neural mechanisms that predate language are capable of incredibly rich adaptive behavior. It therefore seems prudent to seek neural-based solutions to these more basic cognitive functions first, and then to model language by elaborating on the mechanisms for the basic functions. However, that is not the tradition in cognitive modeling (Feigenbaum & Feldman, 1963; Newell & Simon, 1972).

Local versus distributed representation
Language is the part of cognition that we are the most cognizant of. This has led us to modeling the rest of cognition with language and language-like systems such as logic, rather than the other way around. The target article exhibits the typical hallmark of these approaches, local representation, as de Kamps and van der Velde represent words by nonoverlapping cell populations. In nature, this single-neuron--single-function design is found only in relatively primitive animals, in the most basic life-sustaining functions of higher animals, and at the outer sensor and motor periphery -- that is, far from where cognition is assumed to take place.

Local representations make rigid assignments of what part in a connectionist architecture represents what entity. The more flexible and more promising alternative is distributed representation which means more than neuronal patterns involving activity in cell populations rather than single cells. A code is distributed if the neurons truly collaborate. In other words, not the activity in any single neuron in the network but only a distributed pattern of activity can fully specify a certain representation. Cell assemblies, by Hebb's (1949) definition, fulfill this property. Experimental neuroscience offers ample evidence for distributed codes. Braitenberg (1978) has pointed out that results of single-cell physiology suggest strongly of distributed representations. Many neurons respond to generic features (such as edges in visual input) that are shared by a large class of stimuli, rather than to one specific one. Recent techniques for recording the activity in many cells at the same time prove direct evidence for distributed patterns (Harris et al., 2003; Ikegaya et al., 2004). Quite confusingly, van der Velde and de Kamps use local representations but describe their model in the language of (distributed) cell assemblies and populations.

Advantages of distributed representation have been emphasized in early artificial neural-network models, such as neural associative memories. Distributed representation yields drastically increased representational power: A network with n neurons can represent n/k localist population codes with k neurons, but (n choose k) ~ (n/k)^k distributed representations. This means that a connectionist architecture with distributed representation can process more items than it has neurons, which might be crucial for many brain regions. Second, distributed representations form metric spaces where the metric is given by pattern similarity, such as pattern overlap or the inner product. In high-dimensional representational spaces, these metrics behave entirely differently from the Euclidean distance in two or three-dimensional space. For instance, a pair of random vectors is in general almost maximally dissimilar. The properties of these metrics are the mathematical basis for the pattern-recognition properties of artificial neural networks.

Neural architectures for processing distributed representations

Neural associative memories (Willshaw et al., 1969; Kohonen, 1977; Palm, 1982; Hopfield, 1982) are essentially the artificial neural-net implementation of cell assemblies (Sommer & Wennekers, 2003). In these models, cell assemblies are distributed memory patterns that can be formed by learning and can later be recalled. A cell assembly is formed by incrementing the synaptic strengths between all its active neurons, and it can be recalled by stimulating a large enough fraction of its cells. The recall is a pattern-recognition process based on the previously described pattern similarity: a noisy input pattern will evoke this most-similar cell assembly. Associative memories also provide a form of short-term memory: once a cell assembly is active, the recurrent excitation keeps it persistent (Amit, 1989).

However, the above described models of associative memory cannot perform invariant pattern recognition nor can they represent composite or combinatorial structures. The main problem is that memory recall hinges exclusively on overlaps with stored patterns. Input patterns that do not have meaningful overlaps with stored patterns cannot be recognized. Architecture that overcomes this limitation have been proposed (Anderson & van Essen, 1987; Dotsenko, 1989; Olshausen et al., 1993; Arathorn, 2002) and we will refer to them as feature-mapping memories. In these models the input is transformed by a parameterized mapping before it enters the associative memory. By tuning the mapping parameters while searching for a recognizable stored pattern, these systems can cope with invariance that is not reflected in the direct overlap-similarity of the input. The result of memory retrieval is then not only a stored pattern but also the appropriate mapping parameters for mapping the current input to the memory. These architectures have also been extended to more general graph-matching operations (v.d. Malsberg & Bienenstock, 1987; Kree & Zippelius, 1988).

The above feature-mapping memories can perform invariant pattern recognition but they do not provide for a more sophisticated working memory that would allow combinatorial structures to be composed and decomposed on the fly. This requires the introduction of an (invertible) mapping operation for binding (and decomposition) of representations. Smolensky (1990) proposed a map into higher-dimensional space, the tensor product of the representational spaces of the constituents. Such a dimension-expanding binding operation can keep the full structure of the constituents. But this comes at a high price. The problem is that the depth of nested structure is severely limited by affordable dimension and that atoms cannot be substituted for nested structures. Because of the downsides of dimension-expanding binding, Hinton (1990) proposed reduced representation of nested structures in the original space of the constituents. Different binding operations have been proposed for generating reduced representations, in particular, convolutions (Plate, 1994) elementwise multiplication (Kanerva, 1994, 1997; Gaylor, 1998) and permutation-based thinning (Rachkovskij & Kussul, 2000).

Gaylor (2003) has detailed how the Jackendoff problems can be solved with reduced representations. For instance, he uses binding of a representation with a permuted version of this representation to generate frames (for keeping the little star and the big star apart). It is important to note that systems of reduced representations have to be embedded in networks of cell-assembly memory as we described above. This is because the reduction in the binding process produces a noisy result upon decomposition.

The patterns resulting from decomposition have to be passed through a similarity-based filtering process, a clean-up process. In this regard the cognitive operations with reduced representations have close similarity to the recognition of a pattern in feature-mapping memories that we have earlier described.

Conclusions
We believe that connectionist models for cognitive processing should exploit the strengths of connectionism, that is, high representational power and the ability for pattern recognition. These strengths rely fundamentally on distributed representation and, as we have explained, are not realized in the blackboard architecture proposed in the target article. Neural associative memories, in their original form, incorporated the described strengths of connectionism but were too limited in function. Recently, however, memory-based architectures have been developed that can perform invariant pattern recognition in difficult perceptual problems (Arathorn, 2002). Thus, it seems promising to realize neural cognitive architectures by adding operations for processing reduced representations in similar memory systems. In general, we believe that serious progress in cognitive modeling will be based on understanding the general mathematical properties of high-dimensional representational spaces rather than on a specific solution to a relatively narrow set of challenge problems.

References

Amit, D. J. (1989). Modelling brain function. Cambridge University Press, Cambridge, UK.

Anderson, C. H. and Van Essen, D. C. B. (1987). Shifter circuits: A computational strategy for directed visual attention and other dynamic processes. *PNAS USA* 84, 6297-6301.

Arathorn, D. (2002). *Map-Seeking Circuits in Visual Cognition.* Stanford University Press, Palo Alto, CA.

Bienenstock, E. and von der Malsburg, C. (1987). A neural network for invariant pattern recognition. *Europhysics Letters* 4, 121-126.

Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In R. Heim and G. Palm (eds.), *Theoretical Approaches to Complex Systems,* pp. 171-188. Springer, Berlin.

Dotsenko, V. S. (1988). Neural networks: Translation-, rotation-, and scale-invariant pattern recognition. *Journal of Physics A* 21, L783-L787.

Feigenbaum, E. A. and Feldman, J. (eds.) (1963). *Computers and Thought.* New York: McGraw-Hill.

Gayler, R. W. (1998). Multiplicative binding, representation operators, and analogy [Abstract of poster].

In K. Holyoak, D. Gentner and B. Kokinov (eds.), *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences.* Sofia, Bulgaria: New Bulgarian University. Full text at http://cogprints.org/502/

Gayler, R. W. (2003). Vector Symbolic Architectures answer Jackendoff's challenges for cognitive neuroscience. In P. Slezak (Ed.), *Proceedings of the Joint International Conference on Cognitive Science.* Sydney, Australia: University of New South Wales. http://cogprints.org/3983/

Harris, K. D., Csicsvari, J., Buhl, D. and Buzsaki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature* 424, 552-556.

Hebb, D. O. (1949). *The Organization of Behaviour.* Wiley, New York.

Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence* 46, 47-76.

Hopfield, J. J. (1982). Neural networks and physical systems with emerging collective computational abilities. *PNAS USA* 79, 2554-2558.

Ikegaya Y., Aaron, G., Cossart, R., Aronov, D.. Lampl, I., Ferster D. and Yuste, R. (2004). Synfire chains and cortical songs: termporal modules of cortical activity. *Science* 304, 559-564.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford, UK: Oxford University Press.

Kanerva, P. (1994). The Spatter Code for encoding concepts at many levels. In M. Marinaro and P.G. Morasso (eds.), *ICANN '94, Proceedings of International Conference on Artificial Neural Networks* (Sorrento, Italy), vol. 1, pp. 226-229. London: Springer-Verlag. (http://www.rni.org/kanerva/icann94.pdf)

Kanerva, P. (1997). Fully distributed representation. In *Proceedings of 1997 Real World Computing Symposium* (RWC'97, Tokyo). Tsukuba-city, Japan: Real World Computing Partnership. (http://www.rni.org/kanerva/rwc97.ps.gz)

Kohonen, T. (1977). *Associative Memory: A System Theoretic Approach.* Springer, Berlin.

Kree, R. and Zippelius, A. (1988). Recognition of topological features of graphs and images in neural networks. *Journal of Physics A* 21, L813-818.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall.

Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. B. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience* 13, 4700-4719.

Palm, G. (1982). *Neural Assemblies.* Springer, Berlin.

Plate, T. A. (1994). *Distributed representations and nested compositional structure.* Doctoral dissertation. Department of Computer Science, University of Toronto, Toronto, Canada. (http://internet.cybermesa.com/~champagne/tplate/papers/plate.thesis.ps.gz)

Plate, T. A. (2003). *Holographic Reduced Representation: Distributed Representation for Cognitive Structures.* Stanford, USA: CSLI Publications.

Rachkovskij, D. A., and Kussul, E. M. (2001). Binding and normalization of binary sparse distributed representations by Context-Dependent Thinning. *Neural Computation* 13(2), 411-452. (http://cogprints.org/1240/)

Sommer, F. T. and Wennekers, T. (2003). Models of distributed associative memory networks in the brain. *Theory in Biosciences* 122, 55-69.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46, 159-216.