Review Article

Learning to Predict Through Adaptation

Alessandro Treves

SISSA, Cognitive Neuroscience sector, Trieste, Italy; NTNU, Centre for the Biology of Memory, Trondheim, Norway; E-mail: ale@sissa.it

Abstract

Common themes underlying three recent studies of mine on disparate topics are reviewed: the lamination of sensory cortex; the differentiation into sub-fields of the mammalian hippocampus; and the neuronal dynamics that might underlie the faculty for language in the human frontal lobes. These studies all discuss the evolution of cortical networks in terms of their computations, quantified by simulating simplified formal models. They all dwell on the interrelationship between qualitative and quantitative change. Finally, they all include, as a necessary ingredient of the relevant computational mechanism, a simple feature of pyramidal cell biophysics: firing rate adaptation.

AU: please provide 4–7 Index Entries

Looking at the Past Through a Spin Glass

Over the last few years, I have worked at a number of different problems relative to the evolution of cortical networks in mammals. In each case, I have used the simulation of drastically simplified network models as the primary tool for analysis. While the details of the models used were specific and adapted to the problem being considered, the underlying approach has been similar across studies, and this is what I want to briefly discuss in this review, together with the most interesting dissimilarity.

A belief motivating my approach is that the most important steps in the evolution of the nervous system are those that address *computational* demands, demands that are part of the "job specification" of the brain as an information processing system, rather than those steps that address, say, physiological or anatomical constraints. The folding of the human or dolphin cortex into an elaborate pattern of sulci and gyri, for example, I regard as a very interesting problem but one only collateral to information processing. Primarily anatomical in nature, the cortex folding problem has been addressed with

concepts of mechanical tension along axons (Van Essen, 1997). It may still have functional implications, as mechanical forces result in distinct morphologies of the neural elements in different parts of the cortical landscape, and certainly it follows appealingly simple quantitative laws with respect to the ratio of white to gray matter (Zhang and Sejnowski, 2000).

Among genuine information processing problems, one that has been quantified through the use of formal models is the limit on the storage of memories that is imposed by the connectivity of a system of neuron-like units. Considering this limit is partly motivated by the observation that most gray matter volume appears to be devoted to synaptic contacts (Braitenberg and Schüz, 1991), as if the cortex had evolved to maximize connectivity and, ultimately, memory storage. The mathematical procedures that have been used to obtain a proper quantification of the relation between connectivity and memory were originally developed to analyze the physics of a class of materials known as spin glasses (*see* e.g., Amit, 1989). Spin glasses are endowed with interactions that can be characterized as disordered, and hence as interfering with each other, somewhat like, in a neural network, distinct memory representations interfere with each other at retrieval. Although spin glasses have nothing deeper in common with memory systems than this analogy and the mathematical procedures useful in analyzing them, the effectiveness and generality of these procedures have lead some of us to approach many information processing problems by relying on the analysis of spin glasses as a basic paradigm. Unwrapped from its technicalities, the spin glass approach reduces essentially to the idea that cortical systems face a crucial connectivity constraint on extensive memory storage, that the constraint results from interference among memories, and that to analyze such interference we can borrow techniques from statistical physics.

The three problems I consider here are all, to some extent, spin glass problems in disguise.

The Lamination of Sensory Cortex

At the transition from early reptilian ancestors to primordial mammals sensory cortex laminates, in those areas that process topographic modalities, acquiring the layer structure of isocortex. A prominent step in lamination is granulation, whereby the formerly unique principal layer of pyramidal cells is split by the insertion of a new layer of excitatory, but intrinsic, granule cells, layer IV. I have formulated a hypothesis (Treves, 2003a) that accounts for granulation, and for the differentiation between supra- and infra-granular pyramidal layers, as advantageous to support fine topography in the sensory maps that mammals have evolved, over and beyond the gross topography that limits the usefulness of sensory maps in reptiles. Fine topography implies a generic distinction between "where" information, explicitly mapped on the cortical sheet, and "what" information, represented in a distributed fashion as a distinct firing pattern across neurons. Memory patterns can be stored on recurrent collaterals in the cortex, and such memory can help substantially in the analysis of current sensory input. The effective use of recurrent collaterals, because of the "spin glass" limit on memory storage load, requires afferent projections to the cortex that are spread over a large patch; whereas the precise localization of a stimulus on the sensory map requires narrowly focused afferents (*see* Treves, 2003a for the complete argument). The simulation of a simplified network model demonstrates that a non-laminated patch of cortex, with a single characteristic spread of afferent connections, must compromise between transmitting "where" information or retrieving "what" information. The differentiation of a granular layer affords a quantitative advantage, by allowing focused afferents to the granular

units together with widespread afferents to pyramidal units. For this purely anatomical differentiation to be effective, however, it must be accompanied by a physiological differentiation: pyramidal units must adapt their firing, that is decrease their response to steady inputs, much more than granular units. With this further difference, the pyramidal layers can select the correct attractor for memory retrieval before the granular layer, which adapts less, partially takes over the dynamics and focuses activity on the cortical spot that most accurately reflects the position of the sensory input.

Adaptation thus effectively separates out in time, albeit only partially, two information-processing operations that occur in different spaces: the retrieval of memories in the abstract space of attractors, and the accurate relay of stimulus position in the physical space of the cortical surface. The advantage of the differentiation is quantitatively minor, and thus the hypothesis is that a major qualitative step, the transition from a simpler paleocortex to a more elaborate isocortex, came about to gain a few percent more bits in the combined value of "what" and "where" information.

The differentiation of the CA3 and CA1 Fields of the Hippocampus

The differentiation between the CA3 and CA1 fields of the mammalian hippocampus is one of the salient traits that set it apart from the organization of the homolog medial wall in reptiles and birds. CA3 is widely thought to function as an autoassociator, but what do we need CA1 for? Based on evidence for a specific role of CA1 in temporal processing, I have explored the hypothesis that the differentiation between CA3 and CA1 may help solve a computational conflict (Treves, 2003b). The conflict is between *pattern completion*, or integrating current sensory information on the basis of memory, and *prediction*, or moving from one pattern to the next in a stored

sequence. CA3 would take care of the former, while CA1 would concentrate on the latter. I have found the hypothesis to be only weakly supported by neural network simulations. The conflict indeed exists, but two mechanisms that would more directly relate to a functional CA3-CA1 differentiation were found unable to produce genuine prediction. Instead, a simple mechanism based on firing frequency adaptation in pyramidal cells was found to be sufficient for prediction, with the degree of adaptation as the crucial parameter balancing retrieval with prediction. The differentiation between the *connectivity* of CA3 and CA1 has a minor but significant, and positive, effect on this balance, and in particular for a fixed anticipatory interval it significantly increases, in the model, the information content of hippocampal outputs. Different degrees of adaptation in CA3 and CA1 cells, however, were not found to lead to better performance, further undermining the notion of a full functional dissociation. Therefore, there may be just a plain quantitative advantage in differentiating the connectivity of the two fields, just as the hypothesis about lamination holds that there may be just a plain quantitative advantage in differentiating isocortical connectivity across layers.

As for the lamination study, the analysis of this hypothesis about the differentiation of hippocampal subfields was based on the simulation of simplified models of both an unifferentiatiated and a differentiated hippocampus, keeping the same number of units and connections. Each model is tested on memory for a single spatial environment. The accuracy of spatial memory retrieval is subject to the general "spin glass" limit, and it is further modulated by connectivity details. Recent results obtained recording the activity of multiple hippocampal cells, in the labs of Edvard Moser and of James Knierim, indicate a potentially much more dramatic differentiation between CA3 and CA1 units, which has

to do with their ability to distinguish among several spatial environments. These results stimulate the development of more elaborate computational models, that however still revolve around the spin glass limit on memory retrieval, this time extended to memory for multiple environments (Battaglia and Treves, 1998). In their present form (Treves, 2003b), the models would be unable to capture the essential advantage brought about by the connectivity differentiation, if it has to do with multiple maps, while they are sufficient to analyze the essentials of the mechanism linking firing rate adaptation to the prediction of spatial position within a single environment.

The Evolution of Frontal Latching Networks

Understanding the neural basis of higher cognitive functions, such as those involved in language, requires a shift from mere localization to an analysis of network operation. A recent proposal (Hauser et al., 2002) points at *infinite recursion* as the core of several higher functions, including language, and thus challenges cortical network theorists to describe network behavior that could subserve infinite recursion. Building on a variant of the notion that language may have evolved out the semantic and procedural memory systems (Ullman, 2001), I have been exploring the hypothesis that a capacity for infinite recursion may be associated with the natural adaptive dynamics of large semantic associative networks (Treves, 2004). I have used a network of Potts (multistate) units to simulate a semantic memory system distributed over many cortical modules, and I have tested its joint ability to both retrieve a semantic memory based on a partial cue, and, subsequently, when deprived of further inputs, to follow a latching dynamics in attractor space, jumping from one memory to the next with structured transition probabilities. While the

retrieval ability is limited by an appropriate variant of the spin glass constraint (first considered by Kanter, 1988), the latching ability requires a sufficient density of attractors. Since the spin glass constraint limits the number of attractors proportionally to the connectivity, the joint ability can be realized only once the connectivity of the modular system becomes, in evolution, sufficiently extensive. At that point, perhaps following a kind of phase transition, the system is both able to retrieve and to support structured transition probabilities between global network states. The crucial development endowing a semantic system with a non-random dynamics would thus be an increase in connectivity, perhaps to be identified with the dramatic increase in spine numbers recently observed in the basal dendrites of pyramidal cells in human and Old World monkey frontal cortex (Elston, 2000).

Quality vs Quantity and the Need to Adapt

All three studies require firing rate adaptation as a crucial ingredient in producing, respectively: a separation between the processing of "what" and "where" information; the prediction of future locations in a spatial environment; and transitions to different semantic attractor states. In all three, memory retrieval is limited by the "spin glass" constraint.

A fundamental dissimilarity is in the relation between qualitative and quantitative changes. In the first two studies, the hypothesis is that a major *qualitative* structural change may have served to produce solely a *quantitative* functional advantage. Although the first such hypothesis seems a posteriori more convincing than the second, both are methodologically valid a *priori*, and in fact it has been noted (Carroll, 1988) that often in evolution major steps may subserve only "small" improvements in survival ability. In the third study, the hypothesis considered has the

opposite flavor: a quantitative change in connectivity (admittedly, a *major* change) would be enough to produce a phase transition to an entirely novel computational faculty, infinite recursion, with its collateral effects including the emergence of language in humans. Although all these hypotheses require much further testing, they serve to underscore the often subtle relations between structure and function that can apply to cortical networks, mediated by the collective, emergent dynamics of large populations of neurons.

Acknowledgments

I wish to thank the organizers of the excellent Cambridge workshop (where I discussed only the second of the above studies) for giving me the opportunity, with this short review, to express a general viewpoint that is usually squeezed out of individual papers.

References

AU: more info for refs. Amit, Braitenberg, Carroll?

Amit, D. J. (1989) Modeling Brain Function. Cambridge Univ. Press, Cambridge.

Battaglia, F. P. and Treves, A. (1998) Attractor neural networks storing multiple space representations: a model for hippocampal place fields. Physical Review E 58, 7738–7753.

Braitenberg, V. and Schüz, A. (1991) Anatomy of the Cortex. Springer-Verlag Berlin.

Carroll, R. L. (1988) Vertebrate Paleontology and Evolution. W H Freeman & Co., New York.

Elston, G. N. (2000) Pyramidal cells of the frontal lobe: all the more spinous to think with. J Neurosci 20, RC95(1–4).

Hauser, M. D., Chomsky, N., and Fitch W. T. (2002) The faculty of language: what is it, who has it, and how did it evolve? Science 298; 1569–1579.

Kanter, I. (1988) Potts-glass models of neural networks. Phys. Rev. A. 37, 2739–2742.

Treves, A. (2003) Computational constraints that may have favoured the lamination of sensory cortex. J. Comput. Neurosci; 14, 271–282.

Treves, A. (2003b) Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation. Hippocampus 13; in press.

Treves, A. (2004) Frontal latching networks: a possible neural basis for infinite recursion. Cogn. Neuropsych. under review.

Ullman, M. T. (2001) A neurocognitive perspective on language: the declarative/procedural model. Nat. Rev. Neurosci. 2; 717–726.

Van Essen D. C. (1997) A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature 385; 313–318.

Zhang, K. and Sejnowski, T. J. (2000) A universal scaling law between gray matter and white matter of cerebral cortex. Proc. Natl. Acad. Sci. USA 97; 5621–5626.

AU: update

AU: update