Multimodular Networks and Semantic Memory Impairments

David Horn and Nir Levy

School of Physics and Astronomy,

Raymond and Beverly Sackler Faculty of Exact Sciences,

Tel Aviv University, Tel Aviv 69978, Israel

horn@neuron.tau.ac.il   nirlevy@post.tau.ac.il

Eytan Ruppin

Departments of Physiology & Computer Science,

Sackler School of Medicine & Faculty of Exact Sciences,

Tel Aviv University, Tel Aviv 69978, Israel

ruppin@math.tau.ac.il

September 10, 1998

Abstract

We present a multi-modular approach to neural modeling of associative memory. By segregating between intra-modular and inter-modular synaptic transmission, and subjecting the latter to non-linear dendritic processing, we can successfully store memories encoded on different numbers of modules. This model has a striking capability of memory retrieval from partial inputs when the appropriate neurons in only few modules, or even a single module, are activated by afferent connections. Hence, if modular afferents are lesioned, memories that are encoded in more modules are more resilient to damage. Assuming that memories of concrete objects are encoded in a larger number of modules than abstract ones, our results provide a novel explanation of the phenomenon of category specific semantic impairments in patients with deep acquired dyslexia.

1 Introduction

Imaging studies (Martin et al. 1995, Martin et al. 1996) support the notion that the knowledge of concrete objects is stored in the brain in a distributed network of cortical areas. Attributes that define an object seem to be represented close to the cortical regions that mediate perception of these attributes. Hence, if we wish to represent the memory of some object using an attractor neural network, we better let its architecture reflect this general structure. This we try to do in our model, in which the network is constructed of different modules, assuming each module to represent a different cortical area. There exists evidence for cortical modules in the somatosensory (Mountcastle 1957), visual (Hubel and Wiesel 1977) and association (Hevner 1993) cortices. Cortical modules were also shown to function as memory units (see (Amit 1995) for a review).

Given a multi-modular structure one may now wonder how easy it is to store in it memory patterns that are coded in a *variable* number of modules. This problem was first introduced by (Lauro-Grotto et al. 1994). They found that they have to separately define the category to which each memory belongs, thus determining its level of modular coding, and then to modulate synapses and neuronal thresholds during retrieval. This process has led (Lauro-Grotto et al. 1994) to conclude that storage and retrieval of memories that are coded in a variable number of modules must involve consciousness. In a recent work (Levy et al. 1998), we have introduced a low-level mechanism that solves this problem. It is based on a functional distinction between intra-modular and inter-modular synaptic couplings. The post-synaptic currents induced by inter-modular synapses are supposed to undergo additional dendritic non-linear processing before reaching the soma. This new *squashing* function on the inputs coming from other modules, eliminates the difference between contributions to the postsynaptic potential generated by memories with different levels of activity. The biological motivation for intra/inter modular synaptic segregation hinges on the observation that neurons from distant modules synapse onto the distal part of the dendritic tree (Markram et al. 1997, Yuste et al. 1994, Hetherington and Shapiro 1993).

Given this structure we can readily investigate a well-known psychological problem, namely category specific impairments in the presence of focal damage, such as in deep acquired dyslexia (Warrington and Shallice 1984). It is quite easy to understand in our model why memories with higher activity are less susceptible to damage. Following (Hinton and Shallice 1991) one may see this as the reason for the higher resilience of memories of natural objects to focal damage. A recent invetigation reports a reverse situation observed for patients of accute Alzheimer's disease for whom artifact memories are more robust (Gonnerman et al. 1997). We will show how this fact can also be accounted for in our model.

2 The Multi-modular Model

In this section we present the mathematical description of our model. It contains M memory patterns in L modules of N binary neurons each. Each memory η^μ is defined on a subset of size Ω^μ of the L modules. We refer to Ω^μ as its *modular coding*. The sparse coding level inside a module is $p << 1$.

The synaptic efficacy ${J_{ij}}^{lk}$ between the jth (presynaptic) neuron from the kth module and the ith (postsynaptic) neuron from the lth module is chosen in a Hebbian manner

$$ {J_{ij}}^{lk} = \frac{1}{Np} \sum_{\mu=1}^{M} {\eta^{\mu}}_{il} {\eta^{\mu}}_{jk} \ . \tag{1} $$

This synaptic matrix is a natural extension of (Tsodyks 1989) formulation for the case of multi-modular network. The updating rule for the activity state ${V_i}^l$ of the ith binary neuron in the lth module is given by

$$ {V_i}^l(t+1) = \Theta \left[{h_i}^l(t) - \theta_n \right] \ , \tag{2} $$

where $\Theta[x]$ is the step function and θ_n is the neuronal threshold. The neuron's local field, or membrane potential, has two components

$$ {h_i}^l(t) = {h_i}^l{}_{internal}(t) + {h_i}^l{}_{external}(t) \ . \tag{3} $$

The internal field is

$$ {h_i}^l{}_{internal}(t) = \sum_{j \neq i}^{N} {J_{ij}}^{ll} {V_j}^l(t) - \gamma_n \mathcal{Q}^l(t) \ , \tag{4} $$

with inhibition proportional to the total activity inside the module

$$ \mathcal{Q}^l(t) = \frac{1}{Np} \sum_{j}^{N} {V_j}^l(t) \ . \tag{5} $$

The external field is

$$ {h_i}^l{}_{external}(t) = \mathcal{G} \left[\sum_{k \neq l}^{L} \sum_{j}^{N} {J_{ij}}^{lk} {V_j}^k(t) - \gamma_d \sum_{k \neq l}^{L} \mathcal{Q}^k(t) - \theta_d \right] \ , \tag{6} $$

where $\mathcal{G}[x]$ represents dendritic processing of post-synaptic currents from neurons situated in other cortical modules.

The retrieval quality at each recall trial is measured by the overlap function that describes the similarity between the final state V the network converges to and the memory pattern η^μ that is cued in each trial. It is defined by

$$m^\mu(t) = \frac{1}{p(1-p)N\Omega^\mu} \sum_{k=1}^{L} \sum_{i=1}^{N} ({\eta^\mu}_{ik} - p)\, {V_i}^k(t) \ , \tag{7}$$

where Ω^μ is the modular coding of η^μ.

3 The Importance of Non-linear Dendritic Processing

Most conventional neural network models of associative memory are limited to storing memories with fixed coding levels. It is quite a computational challenge to store memories with variable coding levels in the same network. In our network, where we use a fixed coding level p inside the module, this question turns into one of storing memories with different levels of modular codings (i.e. different values of Ω^μ specifying the number of modules in which the memory μ is encoded). It is here where the non-linear character of the dendritic transfer function plays a crucial role. Representing it by a Heaviside threshold function has the important advantage that the neuronal input that is due to activity in other modules will always have a standard strength. Thus a neuron activated during memory retrieval feels the input of all other neurons in its own module as well as a standard strong input reflecting the fact that several other modules of the same memory pattern are active, but it is not sensitive to the exact number of other active modules.

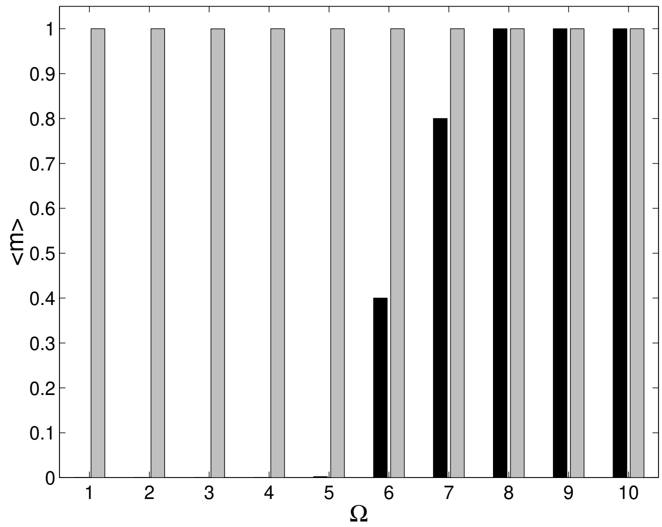

Figure 1: Quality of retrieval *vs.* memory modular coding. The dark shading represents the mean overlap of retrieved memories with different modular codings achieved by a network with linear synaptic couplings. The light shading represents the mean overlap of a network with non-linear processing of the inter-modular connections. The latter achieves perfect recall of all memory patterns. The simulation parameters are: $L = 10$, $N = 500$, $M = 50$, $p = 0.05$, $\lambda = 0.7$, $\theta_d = 2$ and $\theta_n = 0.6$. The encoded memories are distributed homogeneously over all possible modular codings.

An example is presented in Figure 1 where we compare the case of $\mathcal{G}[x] = x$, corresponding to no dendritic processing, with the non-linear $\mathcal{G}[x] = \lambda\Theta[x]$, where Θ is the Heaviside step function. It shows the performance of the two networks when the stored memories have different levels of modular codings. We measure the network's performance by the mean overlaps of the retrieved memory patterns. We choose the modular coding Ω^μ to be homogeneously distributed with equal proportions for all memories e.g., we store 5 memories with $\Omega^\mu = 1$, another 5 memories with $\Omega^\mu = 2$, and so on. As can be seen, for the given choice of threshold parameters, the conventional network can only sustain memories with high modular codings. In order to retrieve patterns with low Ω^μ, non-linear inter-connections are needed.

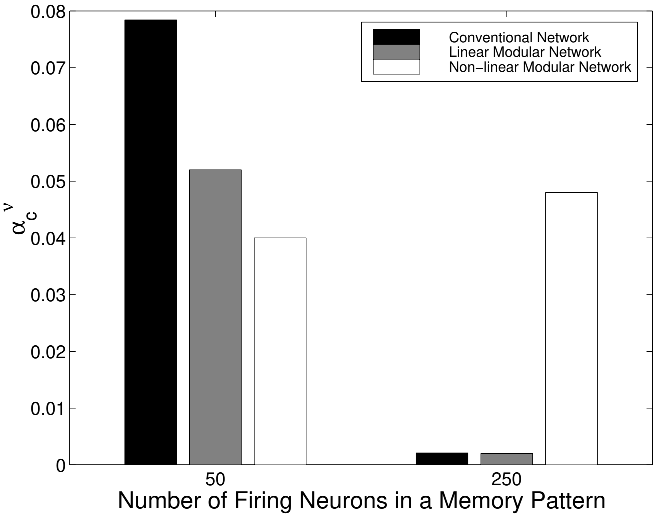

Figure 2: Critical capacities for conventional (single module), linear and non-linear modular networks storing two populations of memories. All three networks have 5000 neurons, and the activities of the two memory populations are 50 and 250 neurons correspondingly (that is, diffused activities of 0.01 and 0.05 in the single module network and two populations of $\Omega = 1, 5$ and $p = 0.05$ in the modular networks). Other parameters are $L = 5$, $\theta = \theta_n = 0.7$, $\theta_d = 2$ and $\lambda = 0.5$.

In (Levy et al. 1998) we have studied the critical capacities of three types of models: single-module with two levels of coding-sparseness, multi-modular network without dendritic processing, and one with non-linear dendritic processing. Only the last one was found to have the desired property of a uniform level of capacity for memories with different coding levels. An example of these results is shown in Figure 2. As evident, the non-linear modular network has a lower critical capacity of the small memories but is significantly better for the large memories.

4 Fault Tolerance to Afferent Focal Damage

The modular organization of our network allows us to study the interesting question of fault-tolerance to focal afferent damage. Assuming spatially localized damage, and setting the input cues to some of the modules to zero, we can easily obtain the pattern completion quality of our model.

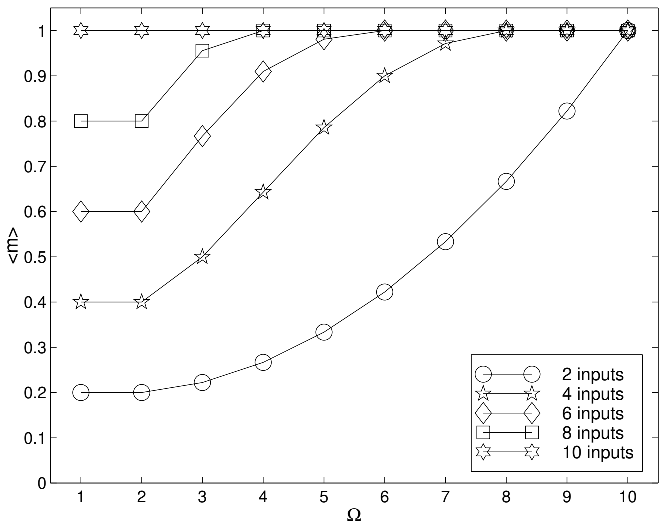

Figure 3: Performance of networks with different modular codings Ω for different cases of focal afferent damage. The curves show the average overlaps obtained when different numbers of modular inputs were left intact.

The basic important message from these results is that memories encoded in more modules are much more resilient to focal afferent damage than memories that are encoded in few modules. It is then interesting to make a connection with the observations of (Warrington and Shallice 1984) who have shown that objects that have more semantic features are better preserved. To be specific, the comparison was made between memories of natural objects and those of artificial objects. Psychological studies have demonstrated that the former are associated with a larger number of semantic features than the latter. Based on this realization, (Hinton and Shallice 1991) have constructed a neural model with various feedforward and feedback features in which they could demonstrate that memories with higher activity are less susceptible to damage. In our model this result is immediate. It follows naturally from the squashed representation of inputs from other modules.

5 Fault Tolerance to Dendritic Weakening

The analysis in the previous section was carried out under conditions in which afferents were damaged but all lateral connections were left intact. Obviously there are many kinds of damage that can be afflicted on neuronal systems, and there exist natural ways for the system to defend itself against most of them (see (Horn et al. 1998) for a discussion of neuronal regulation, a mechanism that has obtained recent experimental support (Turrigiano et al. 1998)). In this section we would like to point out one specific problem that may be easily parametrized and studied in our model, the weakening of the dendritic transport function. In our model this is simply represented by the parameter

λ that multiplies the Heaviside step function describing non-linear dendritic processing. Obviously, the resilience of memories with high modular coding, described in the previous section, crucially depends on λ being large, leading to a strong signal from other activated modules. If, however, λ is weakened, the situation will be reversed. Highly coded memories will lose their resilience to afferent focal damage. In fact, low-coded memories have then a higher chance of survival.

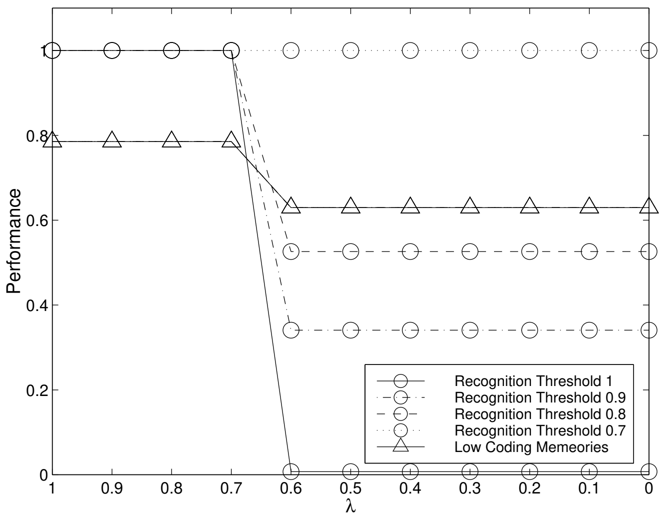

Figure 4: For $\lambda = 1$ low-coded memories ($\Omega = 1, 2, 3$) perform worse than high-coded memories ($\Omega = 8, 9, 10$) when two modular afferents are absent. This situation is reversed when the strength of the signal transmitted from the apical dendrites becomes too weak.

An example of this state of affairs is shown in Fig. 4. Starting out with high-λ values, high-coded memories show their robustness in a situation in which two (of ten) modules lose their inputs. However, when λ decreases below some threshold (0.6 in this example), and memory reconstruction is demanded above some overlap (0.7 here), low-coded memories will outperform high-coded ones on average. This may explain an interesting observation in patients of Alzheimer's disease (Gonnerman et al. 1997). It seems that for low-damage patients memories of natural items are better retained than artifacts, while for high-damage patients artifact memories are more robust. The transition between the two groups of patients looks like the effect described here.

6 Summary

We have studied an associative memory model within a multi-modular approach. By introducing non-linear dendritic processing, we were able to successfully store memories encoded on different numbers of modules. Thus we solved a difficult question within conventional neural network modelling, that of accomodating in the same network memories with vastly different neuronal activities.

Our model has a striking capability of memory retrieval from partial inputs. Even if only few modules are activated by afferent connections the whole memory can be retrieved. Naturally this leads to the conclusion that memories that are encoded in more modules are more resilient to afferent damage. Hence we can easily explain interesting psychological results: assuming that memories of concrete objects are encoded in a larger number of modules than abstract ones, this provides a novel explanation of the phenomenon of category specific semantic impairments in patients with deep acquired dyslexia.

This explanation hinged upon the assumption that the relevant patients have impaired afferents, but the other elements of their memory system stay intact. One novel element that we have introduced in our model is that of non-linear dendritic transmission. Allowing for some deterioration in its strength we move into a different region of anomalous functioning. In fact, it reverses the scenario described above, making memories with high coding more vulnerable. This can account for semantic impairments of patients with accute Alzheimer's disease, who seem to suffer less memory loss of artificial objects than of natural ones.

References

Amit, D. J. (1995), 'The Hebbian paradigm reintegrated: local reveberations as internal representations', *Behavioural and Brain Science* **18**, 617.

Gonnerman, L. M., Andersen, E. S., Devlin, J. T., Klemper, D. and Seidenberg, M. S. (1997), 'Double dissociation of semantic categories in alzheimer's disease', *Brain and Language* **57**(2), 254–279.

Hetherington, P. A. and Shapiro, L. M. (1993), 'Simulating Hebb cell assemblies: the necessity for partitioned dendritic trees and a post-not-pre LTD rule', *Network* **4**, 135–153.

Hevner, R. F. (1993), 'More modules', *TINS* **16**(5), 178.

Hinton, G. E. and Shallice, T. (1991), 'Lesioning at attractor network: investigations of acquired dyslexia', *Psychological Review* **98**(1), 74–95.

Horn, D., Levy, N. and Ruppin, E. (1998), 'Memory maintenance via neuronal regulation', *Neural Computation* **10**, 1–18.

Hubel, D. H. and Wiesel, T. N. (1977), 'Functional architecture of macaque visual cortex', *Proc. R. Soc. Lond. B* **198**, 1–59.

Lauro-Grotto, R., Reich, S. and Virasoro, M. A. (1994), The computational role of conscious processing in a model of semantic memory, *in* 'Proceedings of the IIAS Symposium on Cognition Computation and Consciousness'.

Levy, N., Horn, D. and Ruppin, E. (1998), 'Associative memory in a multimodular network', *preprint* .

Markram, H., Lübke, J., Frotscher, M., Roth, A. and Sakmann, B. (1997), 'Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex', *J. Physiol.* **500**(2), 409 – 440.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L. and Ungerleider, L. G. (1995), 'Discrete cortical regions associated with knowledge of color and knowledge of action', *Science* **270**, 102 – 105.

Martin, A., Wiggs, C. L., Ungerleider, L. G. and Haxby, J. V. (1996), 'Neural correlates of category-specific knowledge', *Nature* **379**, 649 – 652.

Mountcastle, V. B. (1957), 'Modality and topographic properties of single neurons of cat's somatic sensory cortex', *J. Neurophysiol* **20**, 408–434.

Tsodyks, M. V. (1989), 'Associative memory in neural networks with the hebbian learning rule', *Modern Physics Letters B* **3**(7), 555–560.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. and Nelson, S. B. (1998), 'Activity-dependent scaling of quantal amplitude in neocortical neurons', *Nature* **391**, 892–895.

Warrington, E. K. and Shallice, T. (1984), 'Category specific semantic impairments', *Brain* **107**, 829–854.

Yuste, R., Gutnick, M. J., Saar, D. and Tank, K. R. D. D. (1994), 'Ca^{2+} accumulation in dendrites of neocortical pyramidal neurons: An apical band and evidence for two functional compartments', *Neuron* **13**, 23 – 43.