



Computational Constraints that may have Favoured the Lamination of Sensory Cortex

ALESSANDRO TREVES

SISSA—Programme in Neuroscience, Trieste, Italy

ale@sisssa.it

Received April 25, 2002; Revised September 20, 2002; Accepted November 11, 2002

Action Editor: Bruno A. Olshausen

Abstract. At the transition from early reptilian ancestors to primordial mammals, the areas of sensory cortex that process topographic modalities acquire the laminar 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 consider the hypothesis that granulation, and the differentiation between supra- and infra-granular pyramidal layers, may be advantageous to support fine topography in their sensory maps. 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. These patterns can be stored on recurrent collaterals in the cortex, and such memory can help substantially in the analysis of current sensory input. The simulation of a simplified network model demonstrates that a non-laminated patch of cortex must compromise between transmitting “where” information or retrieving “what” information. The simulation of a modified model including differentiation of a granular layer shows a modest but significant quantitative advantage, expressed as a less severe trade-off between “what” and “where”. The further connectivity differentiation between infra-granular and supra-granular pyramidal layers is shown to match the mix of “what” and “where” information optimal for their respective target structures.

Keywords: cortical layers, mammals, isocortex, neocortex, cortical organization, localization, attractor dynamics, recurrent collaterals

Unlike the cerebral cortex of reptiles and birds, the isocortex of mammals acquires during development a characteristic laminar structure that, in its basic traits, is common to different cortical areas and to different species (Ulinski, 1990; Rockel et al., 1980). Salient among these basic traits, especially in primary sensory cortices, is a granulation: the differentiation of a layer of granule cells sandwiched between two layers of pyramidal cells. The functional significance of this major qualitative step in evolution, which likely appeared at the transition from reptiles to mammals and was retained ever since, remains mysterious. Neuroscientists have speculated about it for a long time

(Creutzfeldt, 1977; Szentagothai, 1978; Allman, 1990; see also Super and Uylings, 2001), and recently computational modellers have been trying to address the issue with their methods (Grossberg, 1999). Common to most proposals, however, is the hope that the new structure might be explained *qualitatively*, in terms of distinct functions; for example, opponent inhibition as implemented by layer IV in primary visual cortex (Kayser and Miller, 2002). I explore here the more modest possibility that some basic traits of lamination may be there merely to produce *quantitative* advantages (Carroll, 1997) in carrying out certain information processing operations. Such a quantitative approach has turned out

to be useful in understanding another qualitative step in the evolution of mammalian brains, the insertion of the Dentate Gyrus at the front end of the hippocampal loop (Treves and Rolls, 1992; see Treves and Samengo, 2002), generating predictions in agreement with recent findings (Lassalle et al., 2000).

I focus here on the granulation of sensory cortex in early mammals, and on the differentiation between the connectivity of supra- and infra-granular layers, leaving aside the issue of whether the hypothesis I consider may account for cortical lamination in general. It has long been hypothesized that isocortical lamination appeared together with fine topography in cortical sensory maps, pointing at a close relationship between the two phenomena (Allman, 1990). In early mammals, sensory cortex was likely comprised of just one topographic area for each of the somatosensory, visual and auditory modalities (Diamond and Hall, 1969; Rowe, 1990). Each sensory map thus received its inputs directly from a corresponding portion of the thalamus, as opposed to the network of cortico-cortical connections which has been greatly expanded (Abeles, 1991; Braitenberg and Schuez, 1991) by the evolution of multiple, hierarchically organized cortical areas in each sensory system (Kaas 1993; Krubitzer, 1995). In the thalamus, a distinction has been drawn (Jones, 1998) between its matrix and core nuclei. The matrix, the originally prevalent system, projects diffusely to the upper cortical layers; while the core nuclei, which specialize and become dominant in more elaborate species (Erickson et al., 1967), project with topographic precision to layer IV, although their axons contact, there, also the dendrites of pyramidal cells whose somata lie in the upper and deep layers.

The crucial aspect of fine topography in sensory cortices is the precise correspondence between the location of a cortical neuron and the location, on the array of sensory receptors, where a stimulus can best activate that neuron. Some of the parameters characterizing a stimulus, those reflected in the position of the receptors it activates, are therefore represented continuously on the cortical sheet. I define them as providing *positional* information. Other parameters, which contribute to identify the stimulus, are not explicitly mapped on the cortex. For example, the exact nature of a tactile stimulus at a fixed spot on the skin, whether it is punctuate or transient or vibrating, and to what extent, are reflected in the exact pattern of activated receptors, and of activated neurons in the cortex, but not directly in the position on the cortical sheet. I define these parameters

as providing *identity* information. More elaborate cortices, like the primary visual cortex of primates, include complications due to the attempt to map additional parameters on the sheet, like ocular dominance or orientation, in addition to position on the retina (Bosking et al., 2002). This leads to the formation of so-called columns, or wrapped dimensions, and to the differentiation of layer IV in multiple sub-layers. They should be regarded as specializations, which likely came much after the basic cortical lamination scheme had been laid out, and will not be considered here.

The sensory cortices of early mammals therefore received from the thalamus, and had to analyze, information about sensory stimuli of two basic kinds: positional or “where” information, I_p , and identity or “what” information, I_i . Although I refer to I_p also as “where” information, I stress that the shorthand means “where on the receptor array”, not “where in the outside world”: the cochlea exemplifies the non-equivalence of the two meanings (Zatorre et al., 2002). These two kinds of information differ also in the extent to which cortex can contribute to the analysis of the stimulus. Positional information is already represented explicitly on the receptor array, and then in the thalamus, and each relay stage can only degrade it. At best, the cortex can try to maintain the spatial resolution with which the position of a stimulus is specified by the activation of thalamic neurons: if these code it inaccurately, there is no way the cortex can reconstruct it any better using a priori knowledge, because any other position would be just as plausible. The identity of a stimulus, however, may be coded inaccurately by the thalamus, with considerable noise, occlusion and variability, and the cortex can reconstruct it from such partial information. This is made possible by the storage of previous sensory events in terms of distributed efficacy modifications in synaptic systems, in particular on the recurrent collaterals connecting pyramidal cells in sensory cortex. Neural network models of autoassociative memories (Marr, 1971; Hopfield, 1982) have demonstrated how simple “Hebbian” rules modelling associative synaptic plasticity can induce weight changes that lead to the formation of dynamical attractors (Amit, 1995). Once an attractor has been formed, a partial cue corresponding e.g. to a noisy or occluded version of a stimulus can take the recurrent network within its basin of attraction, and hence lead to a pattern of activation of cortical neurons, which represents the stored identity of the original stimulus. Thus by exploiting in-homogeneities in the input statistics—some patterns of activity, those that

have been stored, are more “plausible” than others—the cortex can reconstruct the identity of stimuli, over and beyond the partial information provided by the thalamus. This analysis of current sensory experience in the light of previous experience is hypothesized here to be the generic function of the cortex, which thus blends perception with memory (Whitfield, 1979). Specialized to the non-topographic olfactory sense, this function may not require new cortical machinery to be carried out efficiently (Haberly, 1990). I explore here the possibility that a novel circuitry is instead advantageous, but only quantitatively, when the generic function is specialized to topographic sensory systems, which have to relay both where and what information, I_p and I_i .

Methods

The Simulated Model of an Isocortical Patch

A patch of cortex is modelled as a wafer of 3 arrays, each with $N \times N$ units. The 3 arrays are depicted as superimposed on each other, in Fig. 1, merely for visual convenience: they could well be imagined interspersed in the same layer, as my simulation does not address the vertical position of each cortical layer, but only their properties, and their being in spatial register with each other. The vertical arrangement is under genetic control (Bar and Goffinet, 2000), and to account for it is beyond the scope of this paper. Each unit, in all 3 arrays,

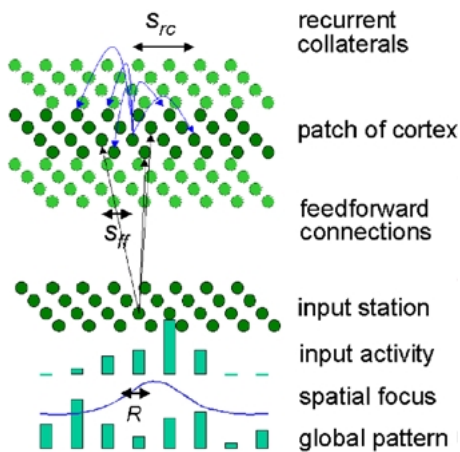


Figure 1. Scheme of the model patch. Parameters used in most of the simulations reported: $N \times N = 20 \times 20$, $R = 2$, $S_{rc} = 8$, $S_{ff} = 2-8$, $C_{ff} = 30$, $C_{rc} = 150$, $M = 12$, $N_{iter} = 10$. Neural representations are constrained to have sparsity $a = 0.3$ (Rolls and Treves, 1998) in each layer.

receives C_{ff} feedforward connections from a further array of $N \times N$ “thalamic” units, and C_{rc} recurrent connections from other units in the patch. Both sets of connections are assigned to each receiving unit at random, with a Gaussian probability in register with the unit itself, and of width S_{ff} and S_{rc} , respectively. Periodic boundary conditions are used, to limit finite size effects, so the patch is in fact a torus. To model, initially, a *uniform*, non-laminated patch, the 3 arrays are identical in properties and connectivity, so the C_{rc} recurrent connections each unit receives are drawn at random from all arrays. To model a laminated patch, later, different properties and connectivity will be introduced among the arrays, but keeping the same number of units and connections, to provide for a correct comparison of performance. The 3 arrays will then model supra-granular, granular and infra-granular layers of the isocortex.

A local pattern of activation is applied to the thalamic units, fed forward to the cortical patch and circulated for N_{iter} time steps along the recurrent connections, and then the activity of some of the units in the patch is read out. To separate out “what” and “where” information, the input activation is generated as the product of one of a set of M predetermined global patterns, covering the entire $N \times N$ input array, by a local focus of activation, defined as a Gaussian tuning function of width R , centered at any one of the N^2 units. The network operates in successive *training* and *testing* phases. In a training phase, each of the possible $M \times N \times N$ activations is applied, in random sequence, to the input array; activity is circulated in the output arrays, and the resulting activation values are used to modify connections weights according to a model associative rule. In a testing phase, input activations are the product of a focus, as for training, by a *partial cue*, obtained by setting a fraction of the thalamic units at their activation in a pattern, and the rest at a random value, drawn from the same general distribution used to generate the patterns. The activity of a population of output units is then fed into a decoding algorithm—external to the cortical network—that attempts to predict the actual focus (its center, p) and, independently, the pattern i used to derive the partial cue. I_i is extracted from the frequency table $P(i, i')$ reporting how many times the cue belonged to pattern $i = 1, \dots, M$ but was decoded as pattern i' :

$$I_i = \sum_{i,i'} P(i, i') \log_2 [P(i, i')/P(i)P(i')]$$

and a similar formula is used for I_p .

The exact “learning rule” used to modify connection weights was found not to affect results substantially. Those reported here were obtained with the rule

$$\Delta w_{ij} \propto r_j^{\text{post}} \bullet (r_i^{\text{pre}} - \langle r^{\text{pre}} \rangle)$$

applied, at each presentation of each training phase, to weight w_{ij} . Weights are originally set at a constant value (normalized so that the total strength of afferents equals that of recurrent collaterals), to which is added a random component of similar mean square amplitude, but asymmetrical, to generate an approximately exponential distribution of initial weights onto each unit. r denotes the firing rates of the pre- and postsynaptic units, and $\langle \dots \rangle$ an average over the corresponding array. In all the simulations shown, only recurrent weights were modified during training, although making feedforward weights modify as well did not affect substantially the results. To check that the sequential presentation of each local pattern during training was not a crucial factor, I have also presented random combinations of 4 local patterns simultaneously, thereby reducing training time by a factor of 4. Results were unchanged.

Among the several parameters that determine the performance of the network, I set $R \ll S_{rc}$, and concentrate on S_{ff} , as it varies from $S_{ff} \approx R$ up to $S_{ff} \approx S_{rc}$. It is intuitive that if the feedforward connections are focused, $S_{ff} \approx R$, “where” information can be substantially preserved, but the cortical patch is activated over a limited, almost point-like extent, and it may fail to use efficiently its recurrent collaterals to retrieve “what” information. If on the other hand $S_{ff} \approx S_{rc}$, the recurrent collaterals can better use their attractor dynamics, leading to higher I_i values, the spread of activity from thalamus to cortex means degrading I_p .

Differentiation of a Granular Layer

I have explored several modifications of the “null hypothesis” uniform model above, to try to find some that could result in a combination of I_p and I_i beyond the limit of the uniform model, exemplified in Fig. 2. A search of this kind cannot be exhaustive, of course, so I have tried in particular modifications that represent rough models of a granulated patch of cortex. Thus, all changes to the uniform model, at this stage, were designed to model solely the emergence of the granular layer, and not any further aspect of a fully laminated cortex. The values of I_p and I_i obtained with several

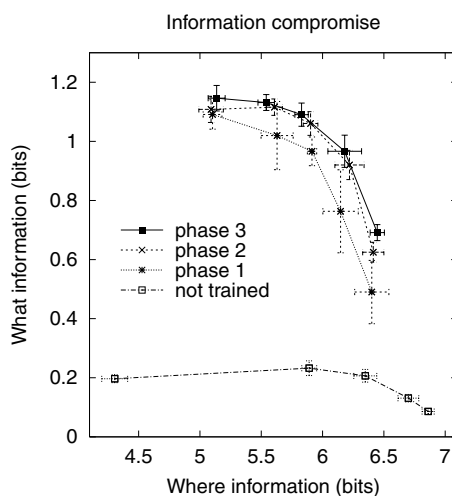


Figure 2. The combined I_i and I_p values obtained in simulations of the uniform model. The same general curve is obtained with other values for M , and using partial cues of different size. In this and the next figures, stimuli determined the firing of 40% of the thalamic units, while the remaining 60% had random activity. Nearly asymptotic values are reached after just 3 training phases. Error bars are calculated as the s.d. of the mean of 10 independent runs.

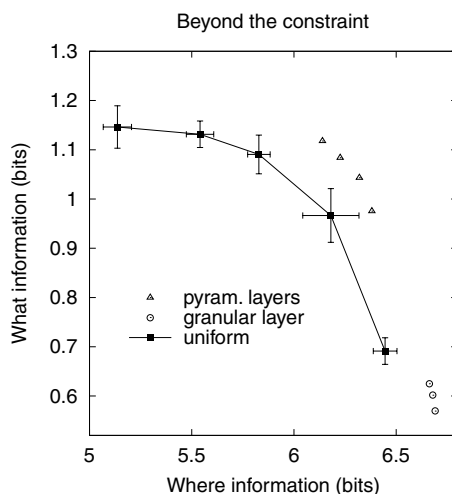


Figure 3. I_i and I_p values obtained, after 3 training epochs, with the uniform model, and with 4 different parameter choices for the granulated model. Note that the latter is not yet asymptotic after 3 epochs. One of the values decoded from the granular layer (circles) falls outside the graph, at $I_p = 6.66$, $I_i = 0.51$.

different simulations, all sharing the same three modifications, but differing in the values of some parameters, are reported in Fig. 3. For all such values of the parameters, the combined information values alleviate the conflict affecting the uniform model. Of course,

other parameter values can be found, that worsen the conflict. It must be stressed, though, that none of these three modifications alone, or in combination with just one of the other two, suffices to cross the boundary. All three together are required, at least in my experience. The three modifications are:

- (1) The thalamic afferents to the granular layer are focused, while those to the two pyramidal layers (still present, and still the same number, per unit) are diffuse. In the simulations shown, with $N = 20$, $S_{\text{ff}}(\text{IV}) = R = 2$, while $S_{\text{ff}}(\text{III}) = S_{\text{ff}}(\text{V}) = 8$. The reduced spread of inputs to layer IV is intended to model the restricted dendritic arborization of spiny stellate cells, rather than selectivity in the thalamic axons themselves.
- (2) The recurrent collateral system of the granular units is severely restricted. In particular, in the simulations reported here, the collaterals originating from layer IV units (and arriving at any layer) are focused ($S_{\text{rc}}(\text{IV}) = 2-3$, while $S_{\text{rc}}(\text{III}) = S_{\text{rc}}(\text{V}) = 8$ (Bosking et al., 1997)) and non-modifiable by training. Those arriving at layer IV units are fewer in number ($C_{\text{rc}}(\text{IV}) = 60$ while $C_{\text{rc}}(\text{III}) = C_{\text{rc}}(\text{V}) = 150$), thus in fact decreasing the total number of synapses in the laminated model with respect to the uniform one.
- (3) Model layer IV units are suppressed during training, and follow a non-adapting dynamics. This is implemented in the simulations by making their effect on postsynaptic units, whatever their layer, scale up linearly with iteration cycle (solely in the retrieval mode, when testing the system). Thus, compared to the model pyramidal units, whose firing rate would adapt over the first few interspike intervals, in reality (but is kept in constant ratio to the input activation, in the simulations), the firing rate of granule units, to model lack of adaptation, is taken to actually increase in time for a given input activation. This could also be taken to model non-depressing short-term plasticity at synapses originating from granule cells, an observation I owe to Prof. Haim Sompolinsky. Note that experimental evidence on either aspect is still inconclusive, and the dynamics implemented in the model has to be regarded as an assumption. Lately non-adapting firing has been attributed to granule cells in the hippocampus (Podlogar et al., 2001), while earlier reports of high release probability at synapses from spiny stellate cells (Feldmeyer et al.,

1999; Tarczy-Hornoch et al., 1999), indicative of depressing behavior, are still controversial.

Differentiating Infra- from Supra-Granular Connections

In addition to the difference in extrinsic projections, the *intrinsic* connectivity of supra- and infragranular layers also differs, although the exact pattern has been explored quantitatively only in some special cases (Nicoll and Blakemore, 1993). A useful summary of many ill-determined details is provided by the “canonical” cortical circuit model of Douglas and Martin (Douglas et al., 1989), which describes activity as propagating first to the supragranular and then to the infragranular layers, while being regulated by inhibitory feedback. In fact, while thalamic projections reach the dendrites of units in all three main layers, the subsequent preferential synaptic flow is $\text{IV} \rightarrow \text{III} \rightarrow \text{V}$ (Lubke et al., 2000). I have made an even cruder model of such flow by removing, in a second version of the laminated model, all direct projections from layer IV to layer V, and replacing them with an equal number of projections from layer III to layer V. All other parameters remain as in the first laminated model. With this further modification, layer III becomes the main source of recurrent collaterals (Nicoll and Blakemore, 1993; Yoshioka et al., 1992), which are spread out and synapse onto both supra- and infra-granular units and also, to a lesser degree, layer IV units.

Scaling with the Size of the Network

In addition to the simulations with $N = 20$, reported so far, I have conducted less exhaustive simulations with other network sizes. In particular, to explore the effects of scaling, I have run somewhat smaller nets ($N = 15$, for a total of 225×3 units in the patch), and somewhat larger nets ($N = 24$, for a total of 576×3 units). The other “length” parameters were scaled linearly, i.e. they were given values, in the first case, $R = S_{\text{ff}}(\text{min}) = S_{\text{rc}}(\text{min}) = 1.5$, $S_{\text{rc}}(\text{max}) = S_{\text{ff}}(\text{max}) = 6$, and in the second case $R = S_{\text{ff}}(\text{min}) = S_{\text{rc}}(\text{min}) = 2.4$, $S_{\text{rc}}(\text{max}) = S_{\text{ff}}(\text{max}) = 9.6$. The number of connections per unit, and the number of patterns used, were instead scaled roughly in proportion to the number of units, i.e. $M = 7$, $C_{\text{ff}} = 17$, $C_{\text{rc}} = 85$ with the smaller nets, and $M = 18$, $C_{\text{ff}} = 45$, $C_{\text{rc}} = 225$ with the larger ones. CPU times scaled approximately as 0.1:1:5, making the $N = 24$ simulations about the

largest that could be run in practice, with the required statistics.

Results

Non-Laminated Model

Preserving accurate coding of position conflicts with the analysis of stimulus identity. The conflict between I_p and I_i is depicted in Fig. 2, which reports their joint values extracted from simulations, as a function of the spread of the afferents, and of the training phase. What is decoded is the activity of all units in the upper array of the patch. Since the patch is not differentiated the other two arrays provide statistically identical information. Further, since information of both the what and where kinds is extracted from a number of units already well in the saturation regime (Treves, 2001), even decoding all units in all 3 arrays at the same time, or only, say, half of the units in any single array, does not alter the numbers of Fig. 2 significantly. Before any training occurs, little “what” information can be retrieved; after training (which with these parameters is already asymptotic with 3 epochs) I_i is monotonically increasing with S_{ff} . I_p , instead, decreases with S_{ff} , and as a result one can vary S_{ff} to select a compromise between what and where information, but *not* optimize both simultaneously. This conflict between what and where persists whatever the choice of all the other parameters of the network, although of course the exact position of the $I_p - I_i$ limiting boundary varies accordingly.

Differentiation of a Granular Layer

It is possible to go beyond the boundary represented in Fig. 2. The information values reported in Fig. 3 with triangles are decoded from the supragranular layer (layer III) in the granulated model. Decoding the activity of layer V, still at this stage statistically identical to layer III, gives the same results. Different values (the circles, with the same parameters as for the layer III triangles) are instead obtained by decoding the activity of layer IV, of course much more biased towards high I_p and low I_i , but still beyond the original constraint. Decoding the activity of all layers simultaneously yields somewhat intermediate values (not shown).

The three modifications, combined, thus produce a slight quantitative advantage in the joint I_p and I_i values that can be read off pyramidal cell activity. The advantage is small, but the model cortical patch used

is also tiny ($N \times N = 20 \times 20$), and the expectation is that the difference between uniform and laminated patches would scale up, as the size of the patch reaches realistic values. This issue is considered further below.

Can we understand the advantage brought about by lamination? The modifications required in the connectivity of layer IV are intuitive: they make granule units more focused in their activation, in register with the thalamic focus, while allowing the pyramidal units, that receive diffuse feedforward connections, to make full use of the recurrent collaterals. What is less intuitive is the requirement for non-adapting dynamics in the granule layer. It turns out that without this modification in the dynamics, the laminated network essentially averages linearly between the performances of uniform networks with focused and with diffuse connectivity, without improving at all on a case with, say, intermediate spread parameters for the connections. This is because the focusing of the activation and the retrieval of the correct identity interfere with each other, if carried out simultaneously, even if the main responsibility for each task is assigned to a different layer. Modifying the dynamics of the model granules, instead, enables the recurrent collaterals of the pyramidal layers to first better identify the attractor, i.e. the stored global pattern, to which the partial cue “belongs”, and to start the dynamical convergence towards the bottom of the corresponding basin of attraction (Amit, 1989). Only *later on*, once this process is—in most cases—safely underway, the granules make their focusing effect felt by the pyramidal units. The focusing action, by being effectively delayed after the critical choice of the attractor, interferes with it less—hence, the non-linear advantage of the laminated model. It should be stressed that layer IV units are suppressed during training, enabling the attractor structure to be stored on the recurrent synapse without the additional constraint of a more limited spatial focus.

Differentiating Infra- from Supra-Granular Connections

Why does isocortex have pyramidal layers both above and below the granule layer? In the granulated model considered above, the supragranular and infragranular layers are still identical in all their properties, and exactly the same mix of I_p and I_i can be read off the activity of both. In the real cortex, however, the supragranular and infragranular layers differ in several ways. One difference which likely goes back hundreds of millions

of years is in their efferent projections. The supragranular layers (denoted here as “layer III”, without any commitment about the why, how and when of the further differentiation between layers II and III) project mainly onward, to the next stage of processing. Layer III is also the major source of callosal projections, those to the other hemisphere (Innocenti, 1986), but those are not considered here. In elaborate mammalian species, ‘onward’ means they project to the next cortical areas in the sensory or motor stream (Felleman and Van Essen, 1991; Barbas and Rempel-Clower, 1997). In simpler mammals, and probably in the primordial species, which likely had only one sensory cortical area per modality (Rowe, 1990), they project strongly to the medial cortex, associated with multimodal integration and memory (Gloor, 1997). The infragranular layers (denoted here as “layer V”, again neglecting the differentiation of V from VI) project mainly backward (Batardiere et al., 1998), or subcortically. Among their chief target structures are the very thalamic nuclei from which projections arise to layer IV. It is clear that having different preferential targets would in principle favor different mixes of what and where information. In particular, cortical units that project back to the thalamus would not need to repeat to the thalamus “where” a stimulus is, since this information is already coded, and more accurately, in the activity of thalamic units. They would rather report in full the genuine contribution of cortical processing, that is, the retrieval of identity information. Thus, we expect infragranular units to maximize “what” information at the expense of “where” information. Units that project to further stages of cortical processing, on the other hand, should balance the “what” added value with the preservation of positional information—the mix that we have so far considered optimal for pyramidal units in general.

To model a patch of cortex with a differentiation between supra- and infragranular layers, I have partially altered the source of recurrent connections to the infragranular layer, as described under Methods. The effect of this differentiation can be appreciated by decoding the activity in the three layers, separately, as shown in Fig. 4. From layer IV one can extract, as before, a large I_p but limited I_i ; from layer III one obtains again a balanced mix, albeit now slightly biased towards favoring I_i —but a choice of somewhat different parameters would take the balance back to any desired value. From layer V, on the other hand, one can extract predominantly “what” information, I_i , at the price of a rather reduced I_p content. Thus, the further

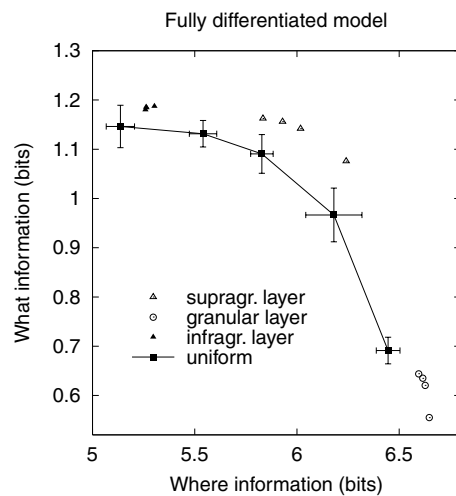


Figure 4. I_i and I_p values obtained, after 3 training epochs, with the uniform model, and with the 4 parameter choices of Fig. 3, but for the fully differentiated model. 3 of the data points for the infragranular layer (black triangles) are nearly superimposed.

connectivity change (which leaves total synaptic numbers unaltered), by effectively reducing the coupling between granular and infragranular layers, has made the latter optimize “what” information, while neglecting “where” information, of limited interest to their target structures.

Scaling

The very limited quantitative advantage produced, in the performance of the model patch, by either granulation alone or by the 3-way lamination, makes it imperative to consider how the advantage would scale up with larger patch sizes. The crucial factor, in scaling, is actually the number of recurrent connections each unit receives, which limits the number of global activity patterns which can be stored and retrieved (Rolls and Treves, 1998). In the simulations, C_{rc} is some two orders of magnitude below realistic cortical values, but it cannot be made much bigger in a patch of limited size $N \times N$ —and, on the other hand, scaling up both N^2 and C , which is in principle possible, rapidly makes simulations exceedingly long. The approach I have taken is to compare the performance of nets of three similar sizes. To make the extrapolation somewhat less arbitrary, I have scaled together “length” parameters, as described in Methods, while connections and number of patterns used have been scaled roughly with length square.

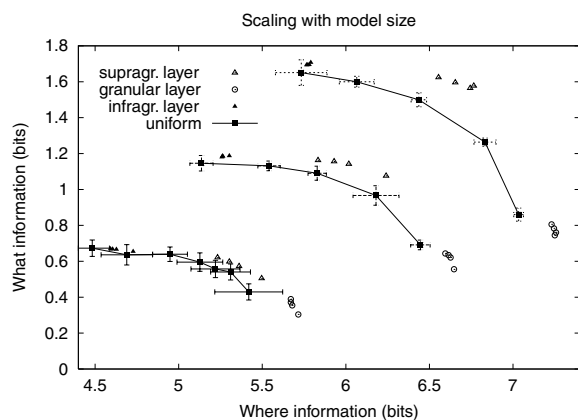


Figure 5. Same as Fig. 4, but including simulations run with networks of size 15×15 , 20×20 (as in Fig. 4) and 24×24 , to show how information values scale with network size, as discussed in the text.

The results show that the advantage of lamination grows in absolute value with network size, and is approximately constant in proportion to the information value itself (it appears in fact to grow slightly even as a percent of the information, from 3.5 to 4.5%, but the statistics is insufficient to demonstrate this increase). Since the fluctuations in the information measures derived from each independent simulation do not grow in absolute size with the size of the network (in fact decrease), the advantage of lamination could not be established statistically with the smaller nets (675 units in the patch), but is significant in the intermediate size nets (1200 units), and more clearly so in the larger nets (1728 units).

Discussion

The Simulation Approach

I propose that a small quantitative advantage in relaying combined positional and identity information may have driven the initial evolutionary differentiation of neocortical layers in mammals. The proposal, I stress, concerns the function of this lamination, not the genetic mechanisms that evolution has developed to carry it out (Allen and Walsh, 1999).

The proposal, while in line with speculations arising of traditional comparative neuroanatomy, is supported by the simulation of a very simplified neural network model—a methodology that requires justification. Connectionist models have a reputation for being

adaptive, which often means they can be designed ad hoc to demonstrate the validity of whatever hypothesis. Indeed, the plethora of parameters that have to be specified in even a simplified neural network model, like the one simulated here, is so large as to make exhaustive analysis impossible, and independent validation difficult. In the case of my simulations, not only the analysis of parameter space, but even most of the details of the model could not be reported here, for lack of space, and will be described in full elsewhere. The code used may be obtained from my website <http://www.sissa.it/~ale/limbo.html>. The truly important elements, however, are (i) the mutual constraint between relaying where information and retrieving what information, evident in the uniform model, and (ii) the quantitative comparison with a laminated model with no more units and/or connections. The what/where conflict is manifest also in rather different models, like those developed independently by Hamish Meffin (2000), and, essentially, it requires only the separate measurability of I_i and I_p to be demonstrated, whatever the remaining details of the model. The quantitative comparison is taken to be fair, here, since the laminated model remains identical to the uniform original, except for the modifications discussed, and in particular has the same number of units. The overall number of synapses in fact decreases slightly, to allow layer IV units, which receive only 2/5 of the original recurrent collaterals, to be more influenced by thalamic inputs in the laminated version. In the real cortex, which devotes most of its volume to synapses (Braitenberg and Schuez, 1991), it is likely that synaptic density per mm^2 is a true constraint to evolutionary expansion; it is also possible that the number of synaptic inputs on a pyramidal cell may be limited by biophysical constraints, to keep the effective electrotonic length of the dendrites short, and allow efficient integration by the cell. What might limit, instead, the number of cells per cortical mm^2 , in other words what could be the cost of adding layers *ad libitum*, is less clear. In the absence of any reasonable argument to estimate such cost, the procedure of keeping numbers fixed, as done here, ensures a conservative comparison: it controls the advantage of differentiating the properties of the different layers for the more trivial effects of adding more layers, or more units, or more connections.

The small advantage of the laminated patch appears to scale up, as described above, when the model, and in particular synaptic numbers, are scaled up towards realistic values. It should be considered, however, that

even a slight quantitative advantage may be selected for, once replicated over millions of sensory experiences per individuals, and over millions of generations in the course of mammalian evolution. Such a quantitative advantage can obviously be demonstrated only with computer simulations, which are precise and can also be replicated millions of times. It remains inaccessible to experimental observation, either in vivo or in vitro, even if it were possible to devise preparations that approximated laminated and uniform cortical patches with similar quantitative characteristics.

Experimental Falsifiability

The early differentiation among layers, discussed here, occurred in mammalian ancestors long extinct, and no present-day species, however primitive or rather regarded as primitive, will ever ensure an experimental check of the hypothesis under identical boundary conditions. In elaborate species, moreover, sensory cortices tend to have specialized much further at later stages in evolution, while higher-order cortices have become positioned in a much more complex network of cortico-cortical connections. With neocortical expansion, the divergence on the way from sensory receptors through the thalamus and up to the cortex has also been altered considerably among species, possibly favoring novel information processing operations. For example, it has been suggested that hyperacuity might rely also on temporal variability (tremor) at the receptor stage, to be then expressed as a pure spatial pattern on a much finer grid of cortical units (Hennig et al., 2001).

What are nevertheless the predictions arising from the proposal, that could be checked experimentally? Essentially, differences in the information content of the activity of populations of cells in different layers. With an appropriate experimental design, I_p and I_i can be measured in vivo from populations of tens of units (Treves, 2001) recorded in well-identified layers. While the relative values of I_p and I_i depend on the design and are not comparable, the model does predict that, very much as in Fig. 4, when separate measures are extracted from populations of equal size, $I_p(\text{V}) \ll I_p(\text{III}) \leq I_p(\text{IV})$, while $I_i(\text{V}) \geq I_i(\text{III}) \gg I_i(\text{IV})$. The differences in I_i should be manifest in cortical areas crucial in the processing of the stimuli to be discriminated, and the testing should involve the use of rather noisy stimuli (partial cues to retrieval).

One system in which a suitable experiment may be conducted is the barrel cortex in the rat. Taking

advantage of the neat topographical arrangement of neurons responding to the stimulation of different whiskers (Welker, 1971), the plan is to apply stimuli that involve groups of whiskers, at different, partially overlapping locations. This makes the coding of position nontrivial, and requires the decoding of the activity of an extended population, which is the one sampled by a fixed microelectrode array. Independently from variations in position, the stimulus can be varied in its temporal dimension, e.g. by using textures which the animal has been trained to discriminate well (Mathew Diamond, personal communication). This would represent stimulus identity in this particular preparation.

Another suitable system may be the auditory cortex of the cat. Here, position in the cortex means sound frequency (in one dimension, what is mapped in the other dimension is not quite clear (Rauschecker et al., 1995)). The plan is to apply stimuli that are superpositions of an extensive number of different narrow bandwidths, fixed at equally spaced intervals on a logarithmic scale (Israel Nelken, personal communication). Each combination of amplitudes would comprise a "pattern", with which the animal would be trained. Different patterns would realize different stimulus identities, while different "positions" would be obtained by multiplying each pattern with a broader Gaussian filter centered at one of several positions on the frequency scale. Again, using randomly sampled populations of units in the different layers, with the total number of units per layer fixed and each population extending over the relevant frequency range, should yield the 3-way inequalities above.

In addition, the time course of $I_p(\text{III})$ is expected to be delayed with respect to that of $I_i(\text{III})$. This is a more delicate point to be checked experimentally, as it requires analysis of the time course of information measures which, to be meaningful, require sampling spikes for a minimum time inversely proportional to the size of the population (Treves, 2001). Since the simplified model used here cannot give a reliable indication of the size of the expected delay, it is important that the minimum time required by the measurement be small enough to be within what may turn to be a relatively rapid local network dynamics, e.g. 10–20 ms (Thorpe and Imbert, 1989; Treves, 1993). Recording from tens of units at a time should ensure both reasonable sampling and reliable measurements of the information in spike patterns extending over 5 ms periods.

A different approach involves the observation of behavior in individuals with genetic or acquired maldevelopments of cortical circuitry (Noctor et al., 2001).

Such observation may indirectly help assess the hypothesis considered here, e.g. the lack of a layer IV may result in poorer behavioral discrimination of stimulus position. Nevertheless, animal preparations do not allow, obviously, a fair comparison between laminated and non-laminated cortices, keeping everything else the same. Therefore any behavioral testing is bound to be even more indirect than physiological experiments, where at least the information content of neural activity can be contrasted among different layers in the same animal.

Relation to Other Proposals

How does the present proposal relate to alternative accounts of the significance of neocortical lamination? Essentially, it does not interfere nor cooperate with the few accounts that, to my knowledge, have been proposed. For example, the 'RULER' model (McComas and Cupido, 1999) emphasizes the *dynamics* of activation in the different layers, in line with the "canonical" model (Douglas et al., 1989), but it does not attempt to really quantify function, or performance. The present model, which is extremely simplified in its dynamics, should be entirely compatible with a more accurate dynamical description. At the same time certain purely dynamical aspects, such as the exact time course of feedforward and feedback inhibition that provides optimal control of excitatory activity, may be important but only tangentially related to the informational properties considered here.

Other points of view, like the one convincingly reviewed by Super and Uylings (2001) contemplate "mechanical" advantages that radial thalamic projections offer over the earlier horizontally running afferents. Among these advantages may be the ease of generating, during development, layers with different properties; but what is the goal of differentiating layers is not addressed. Similar points of view are thus important and complementary to the one adopted here, in that they do not quantitate information processing, but discuss how the appropriate anatomical arrangement may be wired up.

A number of papers have been produced by Stephen Grossberg and collaborators (see e.g. Grossberg, 1999) that as a whole relate neural interactions between the various layers to mechanisms of visual perceptions, e.g. to promote the grouping together of V1 cells with similar orientation and disparity selectivity, or of V2 cells that represent similar edges, texture or shading. Similar

in spirit is the recent proposal by Kayser and Miller (2001). The mechanisms described are fairly complex and difficult to assess with quantitative comparisons between laminated and uniform models. While it remains possible that some of these mechanisms might be specific implementations of my generic account, or at least might be compatible with it, the perspective is clearly very different. It seems to me difficult to disentangle, from the sophisticated mechanisms that have evolved in visually advanced species, such as cats or monkeys, the primitive ones that may have been associated with the emergence of lamination, hundreds of millions of years before. Advanced species have evolved "what" and "where" streams, for example. This later improvement would make it much more difficult to relate positional and identity information to granulation, if one neglected to consider that granulation first occurred, in all likelihood, in species which had no separate sensory streams, or no streams at all.

A simpler strategy seems to be the one pursued here, of considering generic aspects of sensory information processing, pertinent to each topographic modality and to primitive species, and which lend themselves easily to accurate quantification, at least in terms of computer models.

Acknowledgments

I am indebted to Hamish Meffin, for embarking with me on this project, although our routes parted half way, and to Ford Ebner, Moshe Abeles, Valentino Braitenberg, Jim Knierim and Valeria Del Prete, for commenting on an earlier version of this manuscript. Partial support was from Human Frontier Science Program grant RG0110/1998-B. The hospitality of the Institute for Theoretical Physics at U. Cal. Santa Barbara (supported in part by N.S.F. Grant PHY99-07949), where this work was revised, is gratefully acknowledged.

References

- Abeles M (1991) *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press, Cambridge.
- Allen KM, Walsh CA (1999) Genes that regulate neuronal migration in the cerebral cortex. *Epilepsy Res.* 36: 143–154.
- Allman J (1990) Evolution of neocortex. In: EG Jones, A Peters, eds. *Cerebral Cortex, Vol. 8A: Comparative Structure and Evolution of Cerebral Cortex*. Plenum Press, New York. pp. 269–283.
- Amit DJ (1989) *Modelling Brain Function*. Cambridge University Press, Cambridge.
- Amit DJ (1995) The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behav. Brain Sci.* 18: 617–657.

- Bar I, Goffinet AM (2000) Evolution of cortical lamination: The reelin/Dab1 pathway. *Novartis Found. Symp.* 228: 114–128.
- Barbas H, Rempel-Clower N (1997) Cortical structure predicts the pattern of cortico-cortical connections. *Cereb. Cortex* 7: 635–646.
- Batardiere A, Barone P, Dehay C, Kennedy H (1998) Area-specific laminar distribution of cortical feedback neurons projecting to cat area 17: Quantitative analysis in the adult and during ontogeny. *J. Comp. Neurol.* 396: 493–510.
- Bosking WH, Crowley JC, Fitzpatrick D (2002) Spatial coding of position and orientation in primary visual cortex. *Nature Neurosci.* 5: 874–882.
- Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* 17: 2112–2127.
- Braitenberg V, Schuez A (1991) *Anatomy of the Cortex*. Springer-Verlag, Berlin.
- Carroll RL (1997) *Patterns and Processes of Vertebrate Evolution*. Cambridge University Press, Cambridge.
- Creutzfeldt OD (1977) Generality of the functional structure of the neocortex. *Naturwissenschaften* 64: 507–517.
- Diamond IT, Hall WC (1969) Evolution of neocortex. *Science* 164: 251–262.
- Douglas RJ, Martin KAC, Whitteridge D (1989) A canonical microcircuit for neocortex. *Neural Comput.* 1: 480–488.
- Erickson RP, Hall WC, Jane JA, Snyder M, Diamond IT (1967) Organization of the posterior dorsal thalamus of the hedgehog. *J. Comp. Neurol.* 131: 103–130.
- Feldmeyer D, Egger V, Lubke J, Sakmann B (1999) Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel' of developing rat somatosensory cortex. *J. Physiol.* 521: 169–190.
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1: 1–47.
- Grossberg S (1999) How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spat. Vis.* 12: 163–185.
- Haberly LB (1990) Comparative aspects of olfactory cortex. In: EG Jones, A Peters, eds. *Cerebral Cortex, Vol. 8B: Comparative Structure and Evolution of Cerebral Cortex*. Plenum Press, New York. pp. 137–166.
- Hennig MH, Kerscher N, Funke K, Woergoetter F (2001) Stochastic resonance in the visual system: Does the eye-tremor actually improve visual acuity? *Soc. Neurosci. Abs.* 821.29.
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational capabilities. *Proc. Natl. Acad. Sci. USA* 79: 2554–2558.
- Gloor P (1997) *The Temporal Lobe and Limbic System*. Oxford University Press, New York.
- Innocenti GM (1986) General organization of callosal connections in the cerebral cortex. In: EG Jones, A Peters, eds. *Cerebral Cortex, Vol. 5: Sensory-Motor Areas and Aspects of Cortical Connectivity*. Plenum Press, New York. pp. 291–353.
- Jones EG (1998) Viewpoint: The core and matrix of thalamic organization. *Neuroscience*. 85: 331–345.
- Kaas JH (1993) Evolution of multiple areas and modules within neocortex. *Perspect. Dev. Neurobiol.* 1: 101–107.
- Kayser AS, Miller KD (2002) Opponent inhibition: A developmental model of layer 4 of the neocortical circuit. *Neuron*. 33: 131–142.
- Krubitzer L (1995) The organization of neocortex in mammals: Are species differences really so different? *Trends Neurosci.* 18: 408–417.
- Lassalle J-M, Bataille T, Halley H (2000) Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task. *Neurobiol. Lear. Mem.* 73: 243–257.
- Lubke J, Egger V, Sakmann B, Feldmeyer D (2000) Columnar organization of dendrites and axons of single and synaptically coupled excitatory spiny neurons in layer 4 of the rat barrel cortex. *J. Neurosci.* 20: 5300–5311.
- Marr D (1971) Simple memory: A theory for archicortex. *Phil. Trans. R. Soc. Lond. B* 262: 24–81.
- McComas AJ, Cupido CM (1999) The Ruler model. Is this how the somatosensory cortex works? *Clin. Neurophysiol.* 110: 1987–1994.
- Meffin H (2000) Ph.D. Thesis, University of Sidney, Department of Mathematics.
- Nicoll A, Blakemore C (1993) Patterns of local connectivity in the neocortex. *Neural Comput.* 5: 665–680.
- Noctor SC, Palmer SL, McLaughlin DF, Juliano SL (2001) Disruption of layers 3 and 4 during development results in altered thalamo-cortical projections in ferret somatosensory cortex. *J. Neurosci.* 21: 3184–3195.
- Podlogar M, Dietrich D, Selke K, Kral T, Clusmann H, Schramm J (2001) Perforated patch-analysis of firing properties in hippocampal granule cells: Minimal sAHP and tonic firing. *Soc Neurosci abs.* 37.3.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111–114.
- Rockel AJ, Hiorns RW, Powell TBS (1980) The basic uniformity in structure of the neocortex. *Brain* 103: 221–244.
- Rolls ET, Treves A (1998) *Neural Networks and Brain Function*. Oxford University Press, Oxford.
- Rowe M (1990) Organization of the cerebral cortex in monotremes and marsupials. In: EG Jones, A Peters, eds. *Cerebral Cortex, Vol. 8B: Comparative Structure and Evolution of Cerebral Cortex*. Plenum Press, New York. pp. 263–334.
- Super H, Uylings HBM (2001) The early differentiation of the neocortex: A hypothesis on neocortical evolution. *Cerebral Cortex* 11: 1101–1109.
- Szentogthai J (1978) The neuron network of the cerebral cortex: A functional interpretation. *Proc. Roy. Soc. (London) B* 201: 219–248.
- Tarczy-Hornoch K, Martin KA, Stratford KJ, Jack JJ (1999) Intracortical excitation of spiny neurons in layer 4 of cat striate cortex in vitro. *Cereb Cortex* 9: 833–843.
- Thorpe SJ, Imbert M (1989) Biological constraints on connectionist models. In: R Pfeifer, Z Schreter, F Fogelman-Soulie, eds. *Connectionism in Perspective*. Elsevier, Amsterdam. pp. 63–92.
- Treves A (1993) Local neocortical processing: A time for recognition. *Int. J. Neural. Syst.* 3: 115–119.
- Treves A (2001) Information coding in higher sensory and memory areas. In: F Moss, S Gielen, eds. *Handbook of Biological Physics, Vol. 4: Neuro-Informatics and Neural Modelling*. Elsevier, Amsterdam. pp. 825–852.
- Treves A, Rolls ET (1992) Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 2: 189–199.

- Treves A, Samengo I (2002) Standing on the gateway to memory: Shouldn't we step in? *Cognitive Neuropsychology* 12: 557–575.
- Uliński PS (1990) The cerebral cortex of reptiles. In: EG Jones, A Peters, eds. *Cerebral Cortex, Vol. 8A: Comparative Structure and Evolution of Cerebral Cortex*. Plenum Press, New York. pp. 139–215.
- Welker C (1971) Microelectrode delineation of fine grain somatotopic organization of (Sml) cerebral neocortex in albino rat. *Brain Res.* 26: 259–275.
- Whitfield IC (1979) The object of the sensory cortex. *Brain Behav. Evol.* 16: 129–154.
- Yoshioka T, Levitt JB, Lund JS (1992) Intrinsic lattice connections of macaque monkey visual cortical area V4. *J. Neurosci.* 12: 2785–2802.
- Zatorre RJ, Bouffard M, Ahad P, Belin P (2002) Where is 'where' in the human auditory cortex? *Nature Neurosci.* 5: 905–909.