THE CORTICAL PYRAMIDAL CELL AS A SET OF INTERACTING ERROR BACKPROPAGATING DENDRITES: A MECHANISM FOR DISCOVERING NATURE'S ORDER

Oleg V. Favorov[1], Dan Ryder[2], Joseph T. Hester[1], Douglas G. Kelly[3] and Mark Tommerdahl[1]

Departments of [1]Biomedical Engineering, [2]Philosophy, and [3]Statistics

University of North Carolina at Chapel Hill

Corresponding author:
Oleg V. Favorov, Ph.D.
Department of Biomedical Engineering
CB #7575
University of North Carolina
Chapel Hill, NC 27599-7575, USA
Phone: 919-966-1291
Fax: 919-966-6927
E-mail: favorov@med.unc.edu

ABSTRACT

Central to our ability to have behaviors adapted to environmental circumstances is our skill at recognizing and making use of the orderly nature of the environment. This is a most remarkable skill, considering that behaviorally significant environmental regularities are not easy to discern: they are complex, operating among multi-level nonlinear combinations of environmental conditions, which are orders of complexity removed from raw sensory inputs. How the brain is able to recognize such high-order conditional regularities is, arguably, the most fundamental question facing neuroscience. We propose that the brain's basic mechanism for discovering such complex regularities is implemented at the level of individual pyramidal cells in the cerebral cortex. The proposal has three essential components. (1) Pyramidal cells have 5-8 principal dendrites. Each such dendrite is a functional analog of an error backpropagating network, capable of learning complex, *nonlinear* input-to-output transfer functions. (2) Each dendrite is trained, in learning its transfer function, by all the other principal dendrites of the same cell. These dendrites teach each other to respond to their separate inputs with *matching* outputs. (3) Exposed to different but related information about the sensory environment, principal dendrites of the same cell tune to different nonlinear *combinations* of environmental conditions that are *predictably related*. As a result, the cell as a whole tunes to a set of related combinations of environmental conditions that define an orderly feature of the environment. Single pyramidal cells, of course, are not omnipotent as to the complexity of orderly relations they can discover in their sensory environments. However, when organized into feed-forward/feedback layers, they can build their discoveries on the discoveries of other cells, thus cooperatively unraveling nature's more and more complex regularities. If correct, this new understanding of the pyramidal cell's functional nature offers a fundamentally new insight into the brain's function and identifies what might be one of the key neural computational operations underlying the brain's tremendous cognitive and behavioral capabilities.

I. INTRODUCTION

When a student first becomes attracted to neuroscience, it is often because the great mystery of the mind-body problem beckons. “How could that wrinkly lump of cells produce… this!” he might exclaim. “It *must*, I know that, but how?” And he imagines the culmination of his career to be attaining just that understanding.

We all chuckle. Much later, we think, when he comes to appreciate the realities of single neuron physiology, anatomical tracing, genetic analysis, behavioral testing, and (perhaps most importantly) grant writing, he will realize the foolishness of his dream. Neuroscience is in the details. And there are too many details to hope for “understanding the mind/brain” (whatever that might be!) before many generations hence. And even then it will probably be a collective understanding, where Professor C. understands the vermis of the cerebellum, Dr. M. gets a handle on the intricacies of primary motor cortex control of pianist finger movements, and the Y Institute has mapped out the important G-proteins to be found in a typical layer V pyramidal cell in V1.

Is this what we have to look forward to? Possibly, if the brain is, as is commonly believed, an amalgam of many different circuits constructed on different principles to carry out different functions. We should then expect that an understanding of how the brain works will be attained only after a painstaking process of identifying those circuits, discovering how they function, how they are built, and, ultimately, learning how they work together to endow us with our remarkable mental and behavioral abilities.

Older scientific traditions in such disciplines as physics, chemistry, and biology stand in stark contrast to this vision of a balkanized neuroscience. These older traditions are unified and driven forward by *key theoretical insights*, insights that uncover fundamental properties which lie at the causal roots of a natural kind of phenomena. Therein lies their explanatory power. For example, the consistent causal structure that DNA imposes allowed Mendel, Darwin, and Crick & Watson to furnish us with key biological insights. Similarly, the systematic structure of the atom allowed for the theoretical advances of Mendeleev and Bohr. Mechanics was a subject just asking to be described by basic laws (Newton), electromagnetism could be understood with a small set of equations (Maxwell), and even linguistics – once a disunified field like neuroscience – now has the basic organizing principles of transformational grammar (Chomsky).

If the brain really is an amalgam of functional circuits built on different, function-specific principles, then we should expect many insights restricted to those circuits, rather than one or just a few truly fundamental insights, as was the case in other scientific disciplines. There is, however, the alternative possibility, one that seems promising to us: that the brain’s functional capabilities can be explained by a small but fundamental set of principles that we have not yet recognized, and our ignorance of these principles makes understanding the brain seem much more difficult to us than it really is. Neuroscience has had a long tradition of search for such a set of principles (e. g., James, Sherrington, Pavlov, Hebb, Lashley, Barlow, Edelman). Thus far, this search has failed to achieve its goal. Some even judge this strategy to be bankrupt. We shall address some of their principal objections in the final section of this paper. But first we will describe the approach we have taken in pursuing the fundamental principles of neuroscience, and present what we believe are very promising results.

Defining the problem

Where might we look for an insight into the brain’s most fundamental principles? It seems to us that those principles might be discerned most readily by considering what makes the brain uniquely special. We suggest that what distinguishes the brain and is at the very roots of its various accomplishments is the ability to make predictions (in the broadest sense).

Prediction and expectation are closely connected. For example, a passing glance at a mostly occluded but familiar object is frequently sufficient to infer its identity, thus incurring expectations for how the object would appear in full view from another angle, or how it would respond to manipulation. Similarly, we can infer the presence of a particular object (e. g., a deer) without actually seeing any part of it, just from circumstantial evidence (e. g., hoof prints). We acquire the expectation that, if we followed the tracks, we would find a deer. In familiar and even not-so-familiar situations we know what will happen next, what to expect. Our own actions are predictive in their very nature: they are carried out in expectation of certain outcomes.

All these are remarkable feats, considering that the brain receives through its senses only very limited and fragmentary information about the outside world. Though the ability to predict comes easily to us, it is computationally difficult: witness the very minimal successes achieved so far by attempts to emulate in artificial systems even the most basic perceptual and motor tasks.

The ability to predict is what makes it possible for humans and other animals to have behaviors successfully adapted to their environments and circumstances. In turn, what enables us to make predictions is the fact that nature is to a large degree orderly. Animals evolved to exploit, via their behavioral interactions with the surroundings, regularities in their environments. The brain is the organ whose *raison d'etre* is to recognize and exploit nature's orderliness. Some regularities are taken advantage of by neural mechanisms that are instinctive. More importantly (especially for advanced animals), other regularities are discovered by the individual, through its sensory experiences and interactions with its surroundings.

What is the mechanism by which the brain discovers regularities? While the job description for this mechanism is simple – the ability to associate related environmental conditions – filling it is far from simple. The main challenge is posed by the fact that the predictable relations among natural phenomena that are most useful behaviorally operate not between pairs of basic environmental conditions, readily detectable by our sensory receptors, but among *combinations* of conditions of various degrees of complexity. To recognize predictable relations among combinations of conditions, the neural mechanism will need to know among which combinations of conditions to look. Unfortunately, there is practically an infinite number of possible combinations to choose from, and most of them do not yield useful regularities.

A common suggestion has been that regularities among environmental conditions are learned by Hebbian connections among neurons participating in representation of these conditions in the brain. The Hebbian rule is an associative synaptic plasticity rule that varies the strength of connections according to temporal correlation in behaviors of the pre- and postsynaptic cells (Hebb, 1949; Brown et al., 1990). However, as stated above, the environmental conditions that are most useful behaviorally are actually complex nonlinear combinations of simple conditions. Is there a mechanism that enables neurons to tune to precisely those combinations of conditions - among the infinite repertoire of possible ones - that are predictable, i.e. associated with other such combinations? This is a very difficult question, for which there is currently no answer (Phillips and Singer, 1997).

In this paper we propose a radically different mechanism, SINBAD learning, for discovering predictable relations in the environment (SINBAD is an acronym for a *Set of INteracting BAckpropagating Dendrites* referred to in the paper's title). The virtue of this mechanism is that it does not separate the task of neurons' tuning from that of associative learning; associating *combinations* of environmental conditions and tuning neurons to those *combinations* are two outcomes of the same operation. We will first outline our proposal and then explain the details of its implementation and biological justification.

How does the brain discover orderly relations?

We propose that the basic mechanism is implemented at the level of individual neurons. With the cerebral cortex being the part of the brain primarily engaged in the task of discovering regularities, we focus our proposal on the main type of neurons composing the cerebral cortex, pyramidal cells. Pyramidal cells have 5-8 principal dendrites originating from the soma, including 4-7 basal dendrites, and the apical dendrite with its side branches (Feldman, 1984). Each principal dendrite sprouts an elaborate, tree-like pattern of branches and is capable of complex forms of integration of synaptic inputs it receives on its branches from other neurons (Mel, 1994). Through its synapses, each principal dendrite receives information from other neurons about different environmental conditions. Exposed to this information during different situations experienced by the animal, each principal dendrite gradually learns (by adjusting its synaptic connections) to respond in a particular way to patterns of synaptic inputs it receives (Singer, 1995).

Our proposal is that each principal dendrite learns to combine information it receives about different environmental conditions so as to respond to a particular nonlinear combination of them. These combinations are not chosen randomly: each principal dendrite is influenced in its choice by all of the cell's other principal dendrites. Under their mutual influences, principal dendrites in each pyramidal cell choose different, but *co-occurring* or *successive* combinations of environmental conditions. Thus, our basic proposal is that orderly relations in the environment are discovered by individual pyramidal cells, with their principal dendrites each identifying one of several related (and thus mutually predictive) combinations of environmental conditions.

None of the biological learning mechanisms presently known to neuroscience is up to the task of learning nonlinear input-to-output transfer functions (e.g. logical functions, such as *exclusive-OR* or more complex combinations; Phillips and Singer, 1997) and we suggest, as a part of our proposal, a new learning mechanism (which is, in another context, very familiar). It is experimentally well-established that synapses on the dendrites of pyramidal cells are capable of modifying their efficacy under the influence of activities of the pre- and postsynaptic cells (Artola et al., 1990; Kirkwood et al., 1993; Markram et al., 1997b). This synaptic plasticity is currently believed to be Hebbian (Singer, 1995). We propose that it resembles Hebbian learning in that it acts in accordance with the Hebbian rule in the context of the experimental conditions under which it has been studied, but that it is actually a more complex form of learning. According to the form of synaptic plasticity that we propose, the strength of a synapse is controlled not only by the two factors that are used in Hebbian plasticity (output activities of the pre- and postsynaptic cells), but also by the output activity of the principal dendrite on which that synapse resides. Synaptic strength is controlled on the postsynaptic side by the *difference* between the output activity of the postsynaptic cell and the output activity of the host principal dendrite. Such a form of learning is known in the artificial neural network literature as *error-correction* learning (Widrow and Hoff, 1960; Rumelhart et al., 1986). The postsynaptic controlling factor - the difference between the cell's and the dendrite's outputs - is traditionally called the *error*, or *delta*, signal and the effect of this type of learning is to minimize the error signal (Widrow and Hoff, 1960).

As a part of synaptic strength modification, the "error" signal is propagated back to each synapse. If local dendritic branches integrate their inputs nonlinearly (Mel, 1994), then to be effective, the error signal delivered to each synapse will have to be modified in a certain way according to the dendritic conditions along the path from the synapse to the soma. The nature of this modification is well understood theoretically and is known as *error backpropagation* (Rumelhart et al., 1986). Thus we propose that dendrites of pyramidal cells learn by "error" backpropagation; in other words, a principal dendrite is a form of the well-known backpropagation ("backprop") network.

In standard applications of backpropagation networks, the desired output patterns, used to train the network, are provided by an external "teacher." In the dendritic application of backpropagation learning, the role of such a teacher is played by other dendrites: each principal dendrite is "taught" by all the other principal dendrites of the same cell. The purpose of this arrangement is for each principal dendrite in a given pyramidal cell to learn to predict, on the basis of the synaptically-transmitted information available to *it*, what the cell's *other* principal dendrites are doing in response to *their* inputs. The functional significance of the dendrites learning to predict (i. e., match) each other's activities is that the dendrites will identify different combinations of environmental conditions that are predictive of each other. The cell, as a whole, tunes to a set of related combinations of environmental conditions, which define an orderly feature of the environment, such as an object, a complex property, a causal relation, etc.

Combinations of environmental conditions learned by principal dendrites have a number of useful properties. They are *predictable*: that is how they are identified in the first place. They are *informative*: being involved in orderly relations, they are predictive of other conditions taking part in those relations. Finally, they can be useful as *building blocks* in the construction - by other pyramidal cells - of higher-order combinations of environmental conditions and the discovery of regularities involving those higher-order conditions. The path to discovery by the cerebral cortex of high-order regularities in the environment is through discoveries of lower-order regularities, from simple to progressively more complex, because in nature higher-order regularities are built from lower-order regularities. Because lower-order regularities are simpler, they will be easier for cortical pyramidal cells to recognize. In turn, the recognized regularities will make it easier for pyramidal cells in higher-level cortical areas to recognize higher-order regularities that involve them. The recognized higher-order regularities will enable recognition of regularities of even higher order (and in addition via feedback connections they might help recognize other lower-order regularities as well), and so on. Thus, for example, recognition of lines, edges, and textures by some pyramidal cells will enable recognition of surfaces and figures by other cells, which in turn will enable recognition of different types of objects and their states by yet other cells, which will enable recognition of different types of situations (involving interacting objects), etc.

In conclusion, we propose that pyramidal cells of the cerebral cortex are devices for discovery of orderly features of the environment. An individual pyramidal cell obviously has a limited ability to recognize orderly features in the information it receives from other cells. But, when organized into sequences of cortical networks, pyramidal cells can build their discoveries on the discoveries of the preceding cells, thus gradually unraveling nature’s more and more complex relations.

II. IMPLEMENTATION OF THE PROPOSAL

How might error backpropagation learning be implemented in dendrites?

Our proposed mechanism for discovering orderly environmental properties is built on dendrites of pyramidal cells learning by error backpropagation. Figure 1 illustrates how a principal dendrite of a pyramidal cell can be treated as a backpropagation network. Figure 1A shows a drawing of an idealized principal dendrite emanating from a pyramidal cell soma. This dendrite has one primary, two secondary, and four tertiary dendritic branches. Locations of synaptic contacts on those branches are also shown. The individual dendritic branches can be thought of as separate compartments (i. e., electric circuits; Segev et al., 1989), and the entire dendritic tree can be thought of as a system of such interconnected compartments (Figure 1B). Each compartment receives direct synaptic contacts from other cells, and it is also connected with other, contiguous compartments.

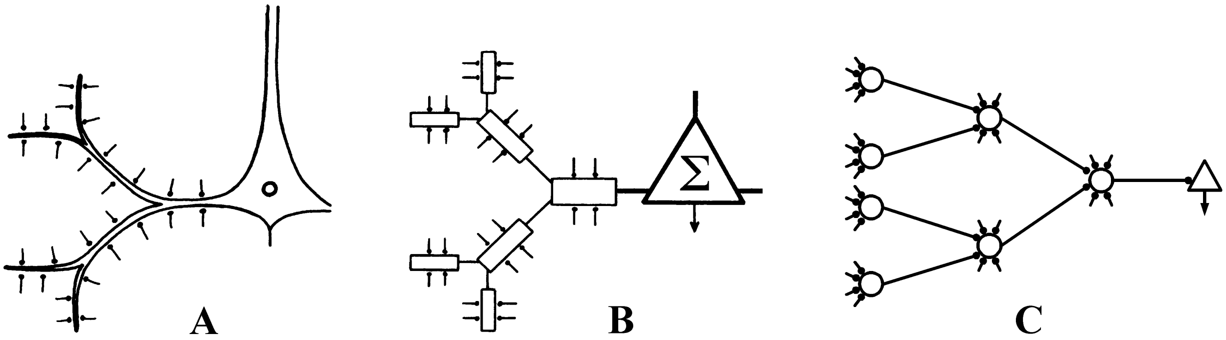


Figure 1. A principal dendrite of a pyramidal cell (**A**) viewed as a set of connected compartments (**B**) and as a multilayered backpropagation network (**C**). Also shown are synaptic connections, distributed throughout the dendritic tree.

The final step in the transformation from the anatomical view of principal dendrites to their representation as a backpropagation network is shown in Figure 1C. As shown here, each dendritic compartment can be treated as a hidden unit of a backpropagation network. Each compartment receives “input layer” (not drawn explicitly) connections (i.e., synaptic connections) and, if it is the primary or a secondary dendritic compartment, it also receives connections from the preceding “hidden layer” (i.e., from more distal compartments).

The key proposal here is to treat the link between two dendritic compartments as an activity-modifiable connection. In physiological terms, the strength of the connection between two compartments is measured as a conductance, and it can be modified in a number of ways, including changes in passive membrane conductance or local changes in the dendritic diameter. The last mechanism is quite interesting because it would predict that learning involves not only changes in synaptic efficacy, but also changes in the sizes and the morphology of dendrites, possibly including retractions as well as sproutings of local dendritic branches (Quartz and Sejnowski, 1997).

In accordance with the backpropagation design, dendritic compartments integrate their inputs nonlinearly, a biologically realistic proposition (Mel, 1994). Dendritic compartments should integrate their inputs nonlinearly because this will enable principal dendrites to learn more complex, linearly inseparable combinations of environmental conditions and will give the dendrites incomparably greater information-processing powers. Some variations in nonlinear properties in different parts of dendritic trees are acceptable: e. g., while distal dendritic compartments should integrate inputs nonlinearly, the proximal compartments can be linear.

The primary dendritic compartment plays the role of the output unit of the backpropagation network; its output is compared with the somal output (which plays the role of a “teacher” signal). The primary compartment, being the stem of the entire dendritic tree, is in an advantageous position to compute the difference between the outputs of the soma and the principal dendrite. First, all the synaptic inputs onto the entire tree travel and converge at the stem, where they summate and thus compute the output of the entire tree. And second, coming from the opposite direction, from the soma, the stem of the dendritic tree also receives information about the cell’s output in the form of action potentials. We propose that the computed difference, or the error signal δ, is then back propagated to all the more distal dendritic compartments and is used to adjust local connections, both the synaptic connections and the connections between compartments.

For demonstration purposes we modeled error backpropagation in the pyramidal cell shown in Figure 1. The cell was given external inputs from four other cells. Each input cell made connections with every compartment of the dendritic tree (it is not necessary for the success of backpropagation learning that each input cell should connect to every compartment, but in this first, simple example we have very few input cells and very few compartments). These connections were controlled by the backpropagation algorithm. The primary and secondary dendritic compartments also had connections - again controlled by the backpropagation algorithm - from two more distal compartments, as shown in Figure 1C.

The activity of each tertiary dendritic compartment was computed as a sigmoid function of its inputs (in this case, hyperbolic tangent with a range between -1 and 1):

$$A_i = \tanh\left(\sum_{j=1}^{4} w_{ji} \cdot IN_j\right), \qquad (1)$$

where A_i is the activity of compartment i, j is an input cell and IN_j and w_{ji} are its activity and its connection strength to compartment i. Activities of the secondary dendritic compartments, which had additional connections from tertiary compartments, were computed analogously:

$$A_i = \tanh\left(\sum_{j=1}^{4} w_{ji} \cdot IN_j + \sum_{k=1}^{2} w_{ki} \cdot A_k\right), \qquad (2)$$

where k is a compartment distal to compartment i, and A_k and w_{ki} are its activity and its connection strength to compartment i. The activity of the primary dendritic compartment was computed as a sum of its inputs:

$$A_1 = \sum_{j=1}^{4} w_{j1} \cdot IN_j + \sum_{k=1}^{2} w_{k1} \cdot A_k, \qquad (3)$$

where k is a secondary compartment, and A_k and w_{k1} are its activity and its connection strength to the primary compartment.

The activity of the primary compartment is also the output of the entire principal dendrite. It contributes to the output of the entire pyramidal cell, i. e. the somal output:

$$OUT = TR + A_1. \tag{4}$$

Here, *TR* is the training signal, representing the net contribution to the somal output from all of the cell's *other* principal dendrites.

The associative learning task we chose for this demonstration is learning a logical function that cannot be learned by a linear neural network (because the training stimulus patterns are not linearly separable), but is easily learned by a standard nonlinear backpropagation network. The function we chose is:

$$TR = (IN_1 \text{ XOR } IN_2) \ \& \ (IN_3 \text{ XOR } IN_4). \tag{5}$$

This function describes a relationship (& and XOR are the logical functions *AND* and *exclusive-OR*, respectively) between the training signal *TR* and the pattern of activities of the four input cells IN_1 - IN_4. The entire training set of 16 stimulus patterns is shown in Figure 2.

After initially setting all the adjustable connections to randomly chosen strengths *w*'s, the cell was activated with a randomly chosen sequence of 16 training stimulus patterns. The connections were adjusted according to the error backpropagation algorithm (Rumelhart et al., 1986) after each stimulus presentation. Specifically, the error signal δ was first computed for the primary dendritic compartment as:

$$\delta = OUT - \alpha \cdot A_1, \tag{6}$$

where α is a scaling constant that has to be greater than 1 in order for the dendrite's output to have a net negative contribution to the error signal; in our demonstration $\alpha = 2$.
For the secondary and tertiary basal compartments δ was computed as:

$$\delta_i = \delta_j \cdot w_{ij} \cdot A_i', \tag{7}$$

where *i* is the compartment for which δ is computed, and *j* is the more proximal compartment to which compartment *i* connects. Connection strengths were adjusted by:

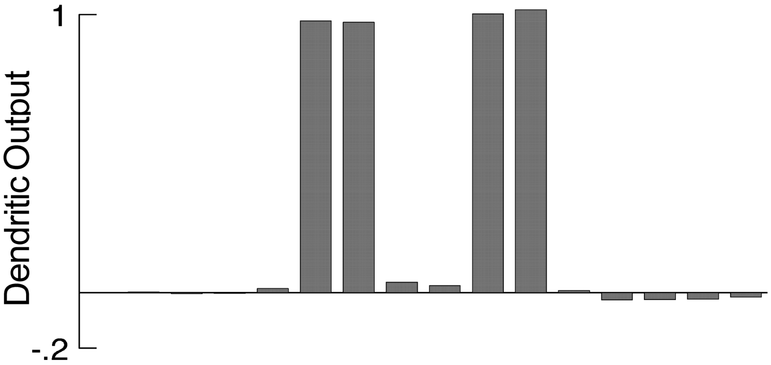
$$w_{ij}(t+1) = w_{ij}(t) + \mu \cdot A_i \cdot \delta_j, \tag{8}$$

where μ is learning rate constant (μ=0.1 in our simulations), *i* is a more distal dendritic compartment or an input cell (in which case $A_i = IN_i$), and *j* is a compartment that receives the connection.

It would seem to be a departure from biological realism that this connection-adjusting algorithm (equations 6-8) can allow a change in the sign of a connection (or, in physiological terms, it can allow a change from excitatory to inhibitory and vice versa). With regard to synaptic connections, for biological realism our setup should be interpreted as follows: each input cell in the model actually stands for two input cells that have identical activities but different physiological sign - one is excitatory, the other is inhibitory (for experimental evidence of plasticity of inhibitory connections in the cortex, see Komatsu and Iwakiri, 1993; Komatsu, 1994; Pelletier and Hablitz, 1996). These cells cannot change the sign of their connections; for example, when a connection weight of an excitatory cell should, according to equation 8, go below zero, it is just set to zero. At the same time, the connection of the inhibitory counterpart cell - which until now has been set to zero - now acquires negative (i. e., inhibitory) value. Thus, a connection weight *w* in the model actually describes the weights of two functionally identical, excitatory and inhibitory input cells.

With regard to connections between dendritic compartments, their weights should only be positive, since they represent positive longitudinal conductances along the dendrites. Accordingly, in the model the weights of intercompartmental connections were never reduced below a certain minimal positive value, which in the case of this demonstration was set at 0.01.

Training of the connections (the training stimuli were presented in a random sequence) continued until their strengths came close to stable values. At this time we evaluated the performance of the dendritic tree by presenting all 16 training stimulus patterns and observing the dendritic output A_1. Results are plotted in Figure 2, showing that the dendrite successfully learned to accurately predict the training signal: when *TR* = 1, A_1 was very close to 1, and when *TR* = 0, A_1 was very close to 0.



Training signal	**TR**	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Input	**2**	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
cells	**3**	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
	4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Input Patterns

Figure 2. Output of the principal dendrite, A_1, in response to the 16 training stimulus patterns. The training patterns are shown in the table below, aligned with the plot.

This exercise demonstrates that dendrites of pyramidal cells can, in principle, implement the error backpropagation learning algorithm, enabling them to learn complex, linearly inseparable functions, and to capture complex relations that are beyond the reach of the Hebbian rule. What is not resolved by this demonstration, however, is whether dendrites do in fact learn by error backpropagation. The current understanding, based on extensive experimental work (Singer, 1995), is that synaptic learning in the cerebral cortex is controlled by the Hebbian rule, according to which the strength of a synapse is a function of the correlation in behavior of the pre- and postsynaptic cells. However, this experimental work does not preclude the possibility that the synaptic learning rule in the cortex is the error-correcting one: the two rules, while greatly different in their functional consequences, are quite similar in their details. Experimentally the error-correction rule can easily resemble the Hebbian rule. In fact, the error-correction rule can be described as an elaboration of the Hebbian rule by an additional controlling factor, namely, the local dendritic activity along the path from the synapse down the dendritic tree to the soma.

Whether it is Hebbian or error-correcting, the synaptic learning rule requires that a signal describing the output activity of the postsynaptic cell be backpropagated from the axon hillock (where output action potentials originate) to each individual synapse. The most likely means by which information about a cell's output reaches the sites of individual synapses on distal dendrites is via spikes that are actively back propagated from the soma up the dendrites (Markram et al., 1997b; Stuart et al., 1997). In the Hebbian rule, the postsynaptic activity signal delivered to a synaptic site should reflect accurately and without distortions the cell's output activity. In the error-correction rule this signal is modified in two ways. The first is by subtracting the activity of the primary dendritic compartment (i.e. the most proximal portion of the dendrite) from the cell's output activity (see eq. 6). Thus the error-correction learning rule can be seen as a version of the Hebbian rule where each principal dendrite's output activity attenuates the postsynaptic signal back propagated through it more that it contributes to it. The second modification of the postsynaptic signal is due to its re-scaling by local longitudinal conductances (probably reflecting diameters of dendritic branches) and by local activity along the dendritic path from the soma to the synapse (see eq. 7). Both modifications are physiologically very plausible, and experimental and theoretical evidence suggests that some of such modifications of the postsynaptic signal do take place (Stuart et al., 1997).

In conclusion, there are grounds to be optimistic that synaptic learning in dendrites of cortical pyramidal cells is of the error correction type, implemented by error backpropagation in the dendrites. To evaluate this possibility, we need experiments in which the efficacy of synaptic connections is studied as a function of not only the pre- and postsynaptic output activities, but also the activity of the host dendrite. The actual learning rule used by the dendrites probably will deviate in its details from the one we describe here (equations 6-8). It might be more effective than the modern backpropagation algorithms used in artificial neural networks. On the other hand, it might be less effective, since learning demands placed on individual dendrites are likely to be quite modest, and the learning power of the entire cortex arises from having large numbers of cells, and from having multiple hierarchically organized cortical areas. What is required from the type of learning we propose here is that each dendrite is capable of tuning to *nonlinear* combinations of its inputs that will *minimize* the difference between the output of that dendrite and the output of the entire cell.

How can dendrites be set up to teach each other?

Interpreting the principal dendrites of cortical pyramidal cells as functional equivalents of backpropagation networks is the first part of our proposal. Backpropagation learning is a supervised form of learning, and the second part of the proposal is that the role of a teacher for a principal dendrite is played by the other principal dendrites of the cell. To explain how dendrites can teach each other and what they can learn in the process, we set up a model of a single pyramidal cell with two principal dendrites. Real pyramidal cells have 5-8 principal dendrites, but two dendrites are sufficient for presenting our basic idea.

Each principal dendrite is modeled in this demonstration as a standard backpropagation network, rather than as a dendritic tree, shown in Figure 1. The reason for this choice is that in our view the principal dendrite is essentially a functional equivalent of the backpropagation network, and the backpropagation network is easier to model and understand. The design of the cell is shown in Figure 3. Each dendrite consists of 10 hidden units and one output unit representing the primary compartment of the dendrite. The hidden units of each dendrite receive input connections from two other cells, or *input channels*, that carry information about environmental conditions. In the present demonstration the two principal dendrites receive connections from two different pairs of input channels, $IN_{1,1}$ and $IN_{1,2}$ vs. $IN_{2,1}$ and $IN_{2,2}$.

Activity of a hidden unit h in dendrite d is computed as a sigmoid function of activities of its two input channels:

$$H_{d,h} = \tanh(w_{d,1,h} \cdot IN_{d,1} + w_{d,2,h} \cdot IN_{d,2}), \qquad (9)$$

where $w_{d,1,h}$ and $w_{d,2,h}$ are the weights of connections of two input channels $d,1$ and $d,2$ on hidden unit h of dendrite d.

Activity of the output unit, i. e. the output of dendrite d, is:

$$D_d = \sum_{h=1}^{10} w_{d,h} \cdot H_{d,h}, \qquad (10)$$

where $w_{d,h}$ is the weight of the connection of hidden unit d,h to the output unit.

Outputs of the two dendrites are summated to produce the somal input $SOM = D_1 + D_2$. The output of the entire cell is:

$$OUT = \tanh(\gamma \cdot SOM), \qquad (11)$$

where γ is a variable that adjusts the cell's output by whether somal input is greater or smaller than the average somal input. Specifically, $\gamma = 1.2$ if $SOM > \overline{SOM}$ and $\gamma = 0.8$ if $SOM \leq \overline{SOM}$. This adjustment drives the cell to expand the dynamic range of its output values. For demonstrating our idea, it is not important here that, in deviation from biological realism, the cell's output can be either positive or negative.

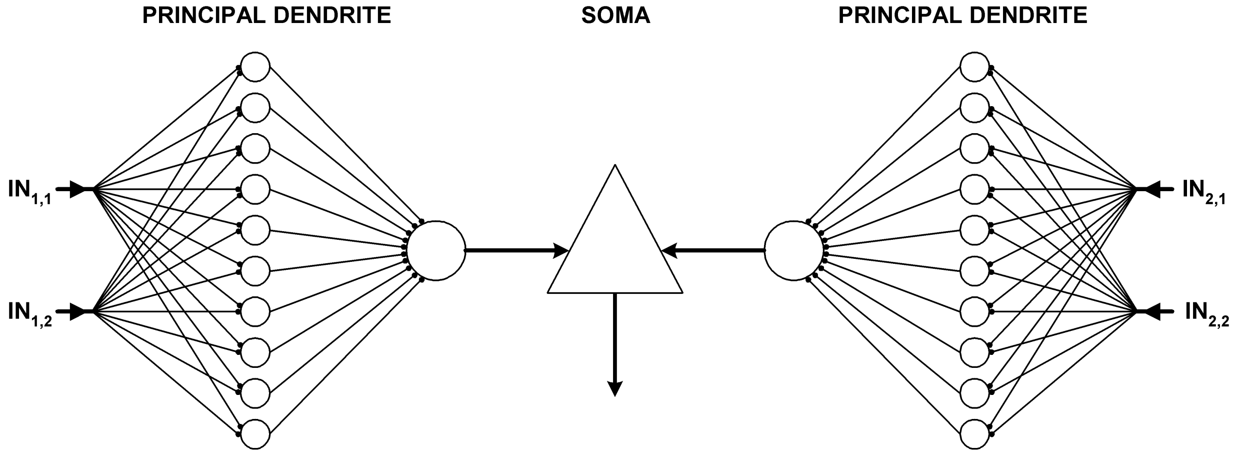


Figure 3. The SINBAD model of a pyramidal cell with two principal dendrites connected to the soma (shown as a triangle). Each principal dendrite is modeled as an error backpropagation network with one output unit, a single layer of ten hidden units, and two input channels.

The sensory environment of the cell was conceived to involve a single orderly entity, object *X*, that can be either present or absent in any given situation (i. e., *X* = 1 or *X* = 0). Object *X* manifests itself in two combinations of environmental conditions, each of which is represented by the activity (0 or 1) of one of the input channels. Specifically,

$X = IN_{1,1}$ *exclusive-OR* $IN_{1,2}$ and (12)

$X = IN_{2,1}$ *exclusive-OR* $IN_{2,2}$. (13)

That is, when present, object X can reveal itself by activating either input channel $IN_{1,1}$ or $IN_{1,2}$, but not both of them, and also by activating either channel $IN_{2,1}$ or $IN_{2,2}$, but not both of them.

There are eight possible patterns of input channel activities that satisfy equations 12 and 13, and they define the entire repertoire of environmental situations distinguished by the input channels. All these patterns are shown in Figure 4B. An inspection of these patterns will reveal that the sensory environment is orderly, but this order is not reflected in activities of single channels, and is only apparent at the level of specific (*exclusive-OR*) combinations of two pairs of channels. Can the cell discover that the combining functions defined by equations 12 and 13 are predictive of each other and also informative of their underlying causal source, object *X*? Note that the existence of object *X* is not indicated to the cell in any direct way, but it is only hinted at indirectly by the regularities hidden in the input patterns.

After initially setting all the adjustable connections to randomly chosen strengths *w*'s, the cell was activated with a randomly chosen sequence of the eight training input patterns. The connections were adjusted according to the error backpropagation algorithm after each input pattern presentation. Specifically, the error signals δ_d were first computed for the two dendrites as:

$$\delta_d = (OUT - EST_d) \cdot EST'_d = (OUT - EST_d) \cdot (1 - EST_d^2), \qquad (14)$$

where EST_d is the dendrite *d*'s estimate of the cell's output *OUT*. It was computed as:

$$EST_d = \tanh(2 \cdot D_d). \qquad (15)$$

For the hidden units, δ was backpropagated as:

$$\delta_{d,h} = \delta_d \cdot w_{d,h} \cdot H'_{d,h}. \qquad (16)$$

Connection weights were adjusted by:

$$\Delta w_{d,i,h} = \mu_i \cdot IN_{d,i} \cdot \delta_{d,h} \text{ and} \qquad (17)$$

$$\Delta w_{d,h} = \mu_h \cdot H_{d,h} \cdot \delta_d, \qquad (18)$$

where μ_i and μ_h are learning rate constants for the input and hidden unit connections ($\mu_i = 6$ and $\mu_h = 0.003$ in our simulations).

Before we turn to the results of simulation of this model neuron, it might be helpful to consider the nature and dynamics of the learning task we set up for the two dendrites. Being a backpropagation network, each dendrite can, in principle, learn a large variety of input-to-output transfer functions. Its actual choice will be dictated by the teaching signal, coming from the other dendrite. But that dendrite also has a large choice of possible transfer functions, and it itself relies on the first dendrite for its own guidance. Thus, the two dendrites will teach each other how to respond to input patterns while continuously changing their own behaviors and their own teaching signals. Such a teaching/learning process will continue until the two dendrites discover such transfer functions that will enable them to have identical responses to their *different* co-present inputs. In other words, the process of connection strength adjustments will continue until each dendrite will learn to predict - on the basis of its own inputs - the responses of the other dendrite to its inputs.

Of course, if the input patterns applied to one dendrite do not relate in any way to the input patterns applied at the same time to the other dendrite, but are accidental in their co-occurrence, then it will be impossible for the dendrites to discover any matching transfer functions, and the process of connection strength adjustments will continue indefinitely. On the other hand, if there is some consistent, and therefore predictable, relationship between co-occurring patterns of the input to the two dendrites, then the dendrites might be able to discover transfer functions predictive of each other's outputs. Success will not be guaranteed, but will depend on the complexity of the orderly relations between the two sets of inputs. For our demonstration here we chose an intermediate level of complexity of the orderly relationship between input channels $IN_{1,1}$ - $IN_{1,2}$ and $IN_{2,1}$ - $IN_{2,2}$; as described above (equations 12-13), this relationship is not discernable at the level of individual channels, but only at the level of their *exclusive-OR* logical combinations.

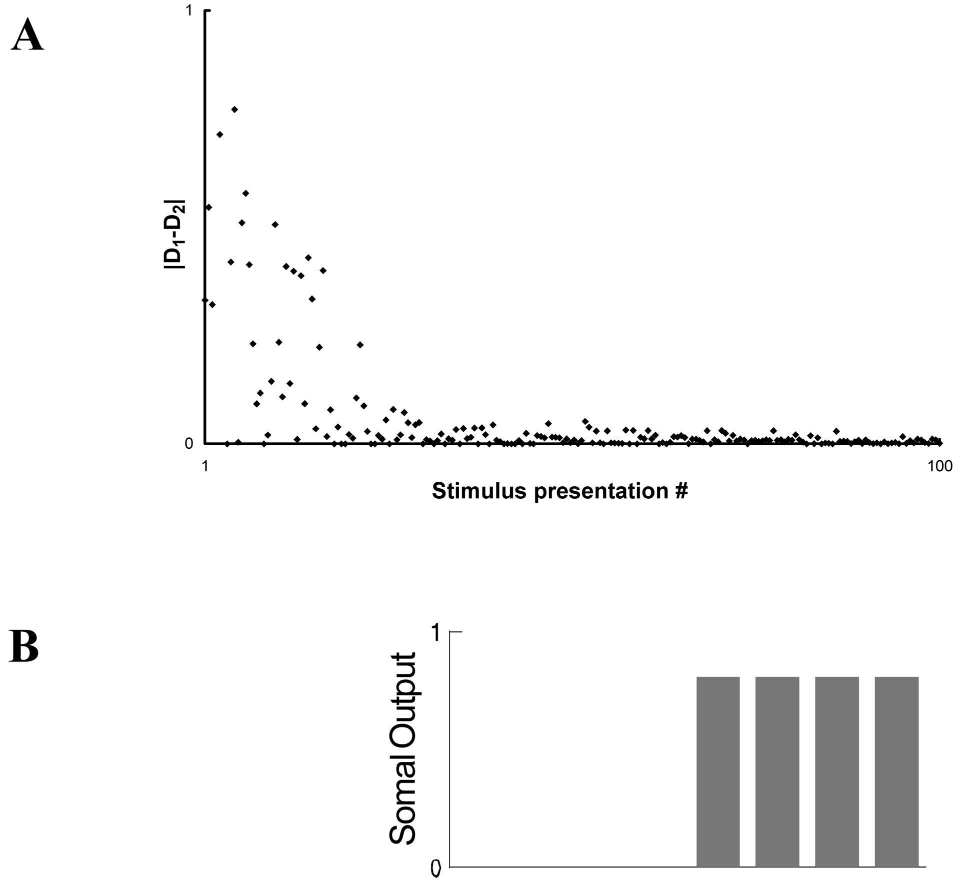
Figure 4 shows the progress and the results of the two dendrites teaching each other how to respond to input patterns. The most pressing question is: Will the two dendrites learn to predict each other's responses to the input patterns? To answer this question, the difference in the output activities of the two dendrites (i. e., $|D_1 - D_2|$) was plotted in Figure 4A as a function of each successive input pattern presentation. This plot shows that the two dendrites discovered very quickly how to respond to their inputs so that they will produce identical outputs. Exposure to a random sequence of fewer than 20 input patterns was sufficient for the two dendrites to discover a way to produce nearly identical responses to their *different* inputs.

The dendrites' learning success raises the following question: What did the dendrites discover about the environment that enabled them to predict each other's responses? To answer this question, we need to examine responses of the whole cell to the entire repertoire of eight possible input patterns. These responses (plotted in Figure 4B) were obtained after the cell was exposed to a random sequence of 100 input patterns, by which time the two dendrites already learned to respond virtually identically (see Figure 4A).

Judging by the cell's responses plotted in Figure 4B, the first dendrite learned to respond to the *exclusive-OR* combination of its input channels $IN_{1,1}$ and $IN_{1,2}$, while the second dendrite learned to respond to the *exclusive-OR* combination of its input channels $IN_{2,1}$ - $IN_{2,2}$. Because of the way we set up the sensory environment (equations 12-13), this pair of transfer functions is the only pair that produces identical outputs, and that is why the dendrites chose them. Thus, Figure 4B shows that the two dendrites successfully identified two nonlinear combinations of their inputs that have an orderly, predictable relationship.

The functional significance of the outcome of the dendrites' learning goes beyond the discovery of two orderly combinations of environmental conditions; it identified the causal source of this order. This causal source is object *X*: the reason why the two combinations of conditions are predictive of each other is because they originate from the same source, object *X*. As Figure 4B shows, the cell's output is accurately indicative of the presence and absence of object *X*. The cell, in effect, discovered the existence in the environment of object *X*.

In conclusion, this modeling exercise shows that by teaching each other, the dendrites in a cell can tune to different combinations of environmental conditions that are predictive of each other and the cell, as a whole, can learn to recognize orderly features in its sensory environment.



	1,1	0	0	1	1	0	0	1	1
Input	**1,2**	0	0	1	1	1	1	0	0
channels	**2,1**	0	1	0	1	0	1	0	1
	2,2	0	1	0	1	1	0	1	0
Causal Source	**X**	0	0	0	0	1	1	1	1

Input Patterns

Figure 4. Learning performance of the SINBAD model. **A.** Magnitude of the difference in outputs of the two dendrites in response to a random sequence of input patterns. In response to each input pattern, the two dendrites adjusted their connections, showing a rapid learning progress, which was essentially completed by the 20th stimulus presentation. **B.** Output of the cell, *OUT*, in response to the 8 training input patterns. The training patterns are shown in the table below, aligned with the plot. Note the match between the state of the causal source *X* and the cell's output.

One necessary condition for cells to be able to discover orderly environmental features is that different dendrites on the same cell should receive connections from different input cells. If all the dendrites in a cell are exposed to the same input sources, then they will not need to discover predictable relations in the environment: having the same inputs, they can easily agree to produce the same outputs, whatever they happen to be. It is only when the dendrites have different input sources that they are forced to rely on regularities in the environment in order to find something common in their different inputs. Thus, for example, if each dendrite in our model had connections from all four input channels, the dendrites would learn transfer functions that produced identical outputs, but most likely they would have no relationship to object *X*; i. e., the dendrites would fail to discover the two combinations of the input channels that are orderly.

There are multiple reasons to believe that real pyramidal cells in the cerebral cortex do indeed satisfy this condition and have their principal dendrites exposed to different input sources. First, it is very unlikely that any given axon making a synaptic contact on one dendrite will also have synapses on *all* the other principal dendrites of the same cell (Schuz, 1992; Thomson and Deuchars, 1994). It can, and frequently does, have synapses on more than one principal dendrite, but not on all of them (Deuchars et al., 1994; Markram et al., 1997a).

Second, connections coming from different sources have prominent tendencies to terminate on different dendrites. For example, neighboring pyramidal cells make synaptic connections preferentially on the basal dendrites (Markram et al., 1997a), whereas more distant cells, including ones several millimeters away in the same cortical area, terminate preferentially on the apical dendrites (McGuire et al., 1991). Another system of connections, the feedback connections from higher cortical areas, terminate preferentially on yet another part of the dendritic tree, i. e., on the terminal tuft of the apical dendrite, located in layer I (see the chapter by Larry Cauller in this volume). This terminal tuft, although not originating directly from the soma, might functionally be considered a principal dendrite, due to its special means of communication with the soma.

The third source of differences in inputs to different principal dendrites of the same cell is cortical topographic organization. Across a cortical area, functional properties of cells change very quickly: even adjacent neurons carry in common less than 20% of stimulus-related information (Gawne et al. 1996; for a review, see Favorov and Kelly, 1996). Basal principal dendrites extend in all directions away from the soma and thus spread into functionally different cortical domains. As a result, these dendrites sample different loci of the cortical topographic map and are exposed to different aspects of the cortical representation of the environment (Malach, 1994). We elaborate this idea further in Section III, which provides a brief review of the fine, minicolumnar structure of cortical topographic maps and relates the SINBAD proposal advanced in this paper to the cortical minicolumnar organization.

Overall, the 5-8 principal dendrites of the same pyramidal cell are exposed to different sources of information about the environment and, in order to have correlated output behaviors, they will have to tune to different but mutually predictive combinations of environmental conditions.

How to divide connections among the dendrites?

So far we have described the principle by which pyramidal cells can discover regularities in their sensory environment. For any practical implementation of this principle we must address the question of how to distribute input channels among the principal dendrites of a cell so that the right combinations of channels will go to the right dendrites. In the previous modeling exercise we skirted this issue by assigning the four input channels to the two dendrites in pairs that we knew were the right ones for the environment we set up there. But if we did not know the orderly organization of the environment in advance, then we would not know how to divide the four channels between the two dendrites. Another concern is that, unlike in the previous exercise, the environment possesses not just one, but many different orderly features and of various levels of complexity. How can the cells discover as many of these orderly features as possible?

To address these issues, we will start with another simple modeling demonstration. For this demonstration we created a more complex sensory environment, which involved regularities of three orders of complexity. This environment is characterized by 32 parameters, or *elementary conditions*. They might be, for example, 32 sensory channels through which some hypothetical animal obtains information about the state of its surroundings. The environment is orderly, which means that only a limited subset of all possible combinations of elementary conditions can occur. Figure 5 shows a number of examples of such orderly patterns, and these examples make it obvious that the orderly features of this environment will not be easy to recognize.

To appreciate the difficulty in making use of the orderly structure of this environment, we can consider a hypothetical animal that inhabits it. Suppose a particular third-order combination of environmental conditions, involving some of the environment's orderly features, has a specific behavioral significance to an animal. In Figure 5, a small sample of the 2^{22} possible environmental states are separated into two panels according to the presence or absence of that combination.

Consider the two panels. How to distinguish between these two sets of patterns? The approach advocated in this paper is to start by learning to recognize first-order regular combinations among elementary conditions, then second-order regular combinations among the first-order ones, and finally the third-order combination.

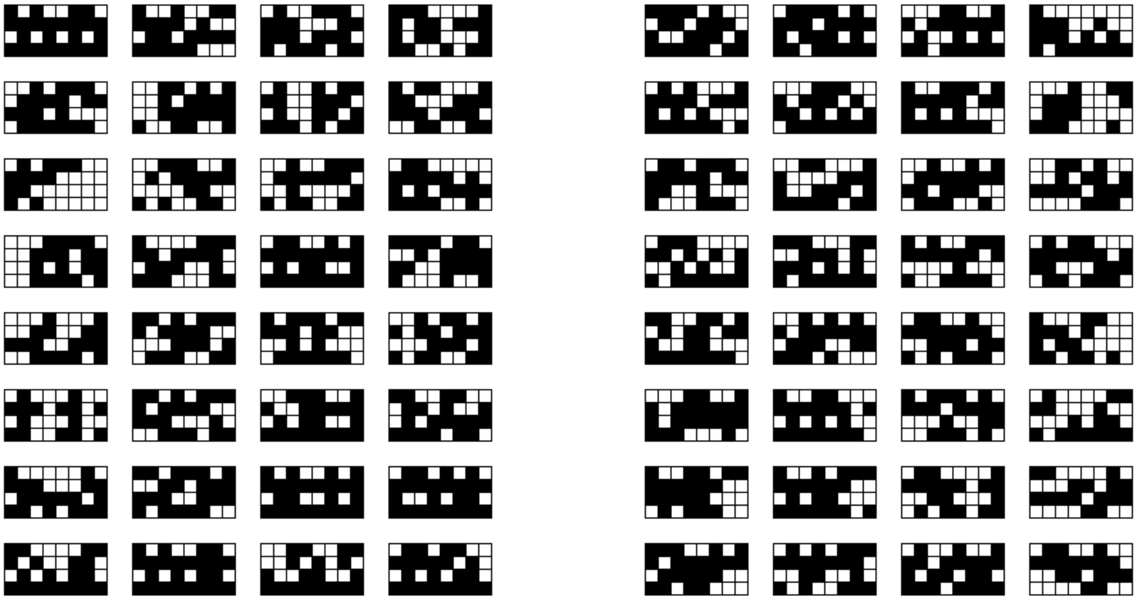


Figure 5. Sixty-four examples (out of 2^{22} possible) of orderly states that can be taken by the model environment. Each state is shown by an 8x4 field of small squares, with each square representing (by its shading) the status of one of the 32 elementary environmental conditions (white = absence, or 0; black = presence, or 1). The environmental states shown in the two panels can be distinguished by the presence (right panel) or absence (left panel) of a particular third-order combination of elementary environmental conditions, described in the text.

The orderly structure of the modeled environment is illustrated in Figure 6. The environment was set up to have two entities, called objects α and β. The animal should act one way when either object α or object β is present, and it should act the other way when neither or both objects are present (this is an *exclusive-OR* logical function).

Object α, turn, is indicated by environmental properties *a* and *b*, let's say α = (*a exclusive-OR b*). That is, object α manifests itself as *a* or *b*, whereas together *a* and *b* do not indicate α, but form an accidental combination.

In addition, α can be reliably predicted by a combination of properties *c* and *d*, let's say α = (*c exclusive-OR d*). It might be, for example, that α has two sides, one side characterized by either *a* or *b* and the other by either *c* or *d* combinations. Or the *cd* combination might describe some other environmental condition that co-occurs with object α.

Object β is organized according to the same plan as object α: β = (*e exclusive-OR f*), and β also can be reliably predicted by a combination (*g exclusive-OR h*).

To add one more layer of complexity, environmental properties *a, b, c, d, e, f, g, h* are in turn defined by elementary environmental conditions *1* through *32*. All of these properties were given the same organization:

a = (*1 exclusive-OR 2*) = (*3 AND 4*)
b = (*5 exclusive-OR 6*) = (*7 AND 8*)
c = (*9 exclusive-OR 10*) = (*11 AND 12*)
d = (*13 exclusive-OR 14*) = (*15 AND 16*)
e = (*17 exclusive-OR 18*) = (*19 AND 20*)
f = (*21 exclusive-OR 22*) = (*23 AND 24*)
g = (*25 exclusive-OR 26*) = (*27 AND 28*)
h = (*29 exclusive-OR 30*) = (*31 AND 32*)

Thus, the modeled environment possesses 32 elementary conditions, 8 first-order predictable combinations of elementary conditions (*a* - *h*), and 2 second-order predictable combinations (α, β).

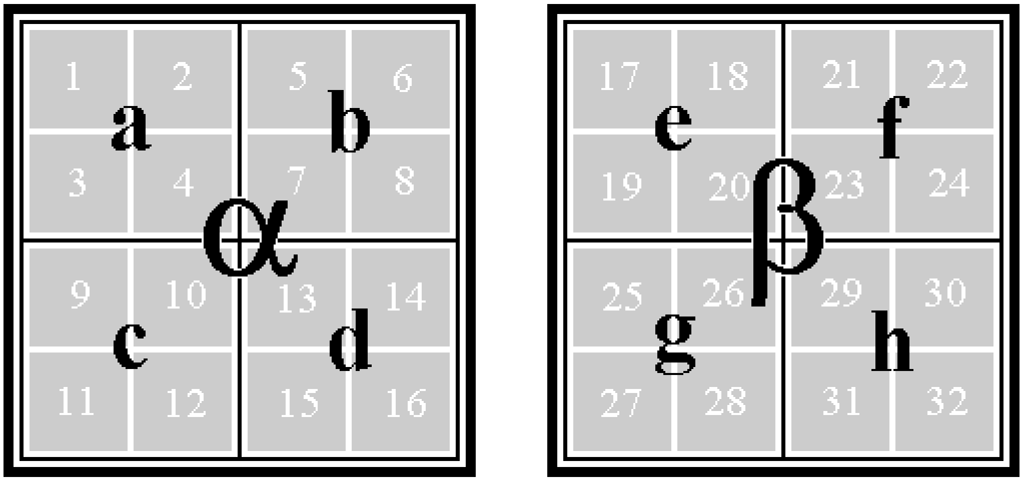


Figure 6. The orderly structure of the model environment. Shown as small gray squares are 32 elementary environmental conditions, organized into 8 sets that define environmental properties *a* - *h*. These 8 properties, in turn, are organized into 2 sets that define objects α and β. The rules by which lower-order conditions define the status of higher-order conditions are specified in the text.

We presented this environment to the model used in the previous demonstration, with a few modifications. The first modification is that this cell now has 32 input channels, each carrying information about one of the 32 elementary environmental conditions. Unlike the previous exercise, these connections are divided randomly between the two dendrites of the cell. Each input channel is assigned at random to either one or the other dendrite, but not to both of them. Also, connection weights of input channels are set initially to zero, except for four randomly chosen channels on each dendrite. For these four connections, their initial connection weights are chosen at random. Thus, unlike the previous exercise, here we do not take advantage of our knowledge of which input channels should go together.

In another modification of the initial design, the number of hidden units in each dendrite is increased to 40. Otherwise, activities of the hidden units $H_{d,h}$ and the dendrite's output unit D_d are computed as described in equations 9 and 10. The cell's output *OUT*, error signal δ_d, error backpropagation $\delta_{d,h}$, and dendritic connection *w* adjustments are computed as before, according to equations 11, 14 - 18.

Figure 7 shows the result of one simulation run during which 10,000 input patterns were presented in a random sequence to the cell. To see whether the two dendrites of the cell learned to produce similar outputs in response to their co-present input patterns, the correlation coefficient between outputs of the two dendrites was computed across 100 successive input pattern presentations. In Figure 7 the value of this correlation coefficient is plotted as a function of time since the start of the training period. This plot shows that in the beginning the two dendrites correlated poorly in their outputs, but after some initially unsuccessful search the dendrites discovered a way to produce very similar outputs. This means that the two dendrites discovered some orderly relationship in the environment, which enabled them to predict each other's responses to input patterns.

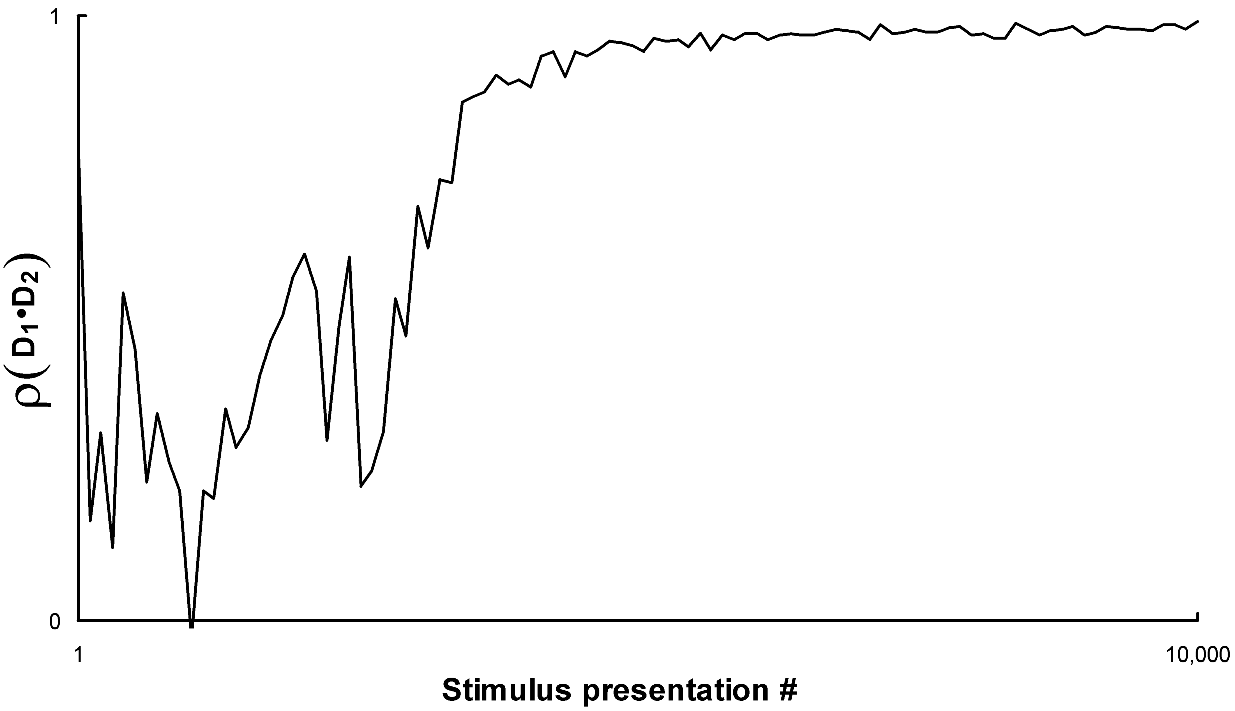


Figure 7. Learning performance of the SINBAD model. The correlation coefficient, ρ, between the outputs of the two dendrites is plotted as a function of time from the start of the training period.

The environment has a number of orderly features, listed above, and which of them were discovered by the cell was determined by the distribution of the input channels on the two dendrites. Because the channels were distributed between the dendrites randomly, there is no guarantee that a set of channels engaged in a given orderly combination (e. g., channels *1* and *2*, as an *exclusive-OR* combination, reporting on the status of property *a*) were placed on the same dendrite, or that predictive pairs of channels were placed on the opposite dendrites (e. g., channels *1* and *2* on one dendrite and channels *3* and *4* on the other dendrite). As a result, the particular distribution of input channels on the dendrites might make it impossible for the cell to discover a particular environmental regularity.

There are a number of ways a cortical network is likely to deal with this limitation. The simplest way is to have many pyramidal cells. Some of them will, by accident, have an appropriate distribution of input channels on their dendrites and thus they will be in a position to discover that orderly environmental feature. With a sufficient number of cells, the network will be able to discover all the orderly features.

To illustrate this idea, we expanded the model to have 32 cells identical to the one used in the last exercise. They differ from each other only in the distribution of input channels on their dendrites, which was assigned randomly. For simplicity, the 32 cells were run in parallel, without interactions among them. To evaluate the learning outcomes of this network, we also set up three additional cells (or more accurately one of the principal dendrites of three additional cells residing in a higher cortical area). The design is shown in Figure 8. Our aim here is to compare the ability of a "test" dendrite to learn a particular behavior, given either raw information, provided by the input channels, or the information provided by the 32 cells (which presumably transformed the raw input patterns into a new form in which some of the orderly environmental features are represented more explicitly).

One test dendrite was trained to respond to the presence of object α, another to object β, and the third to a combination Ω = (α *exclusive-OR* β). (These dendrites might be viewed as belonging to cells that, for some unspecified reasons, are driven by their other dendrites to respond to α, β, or Ω.)

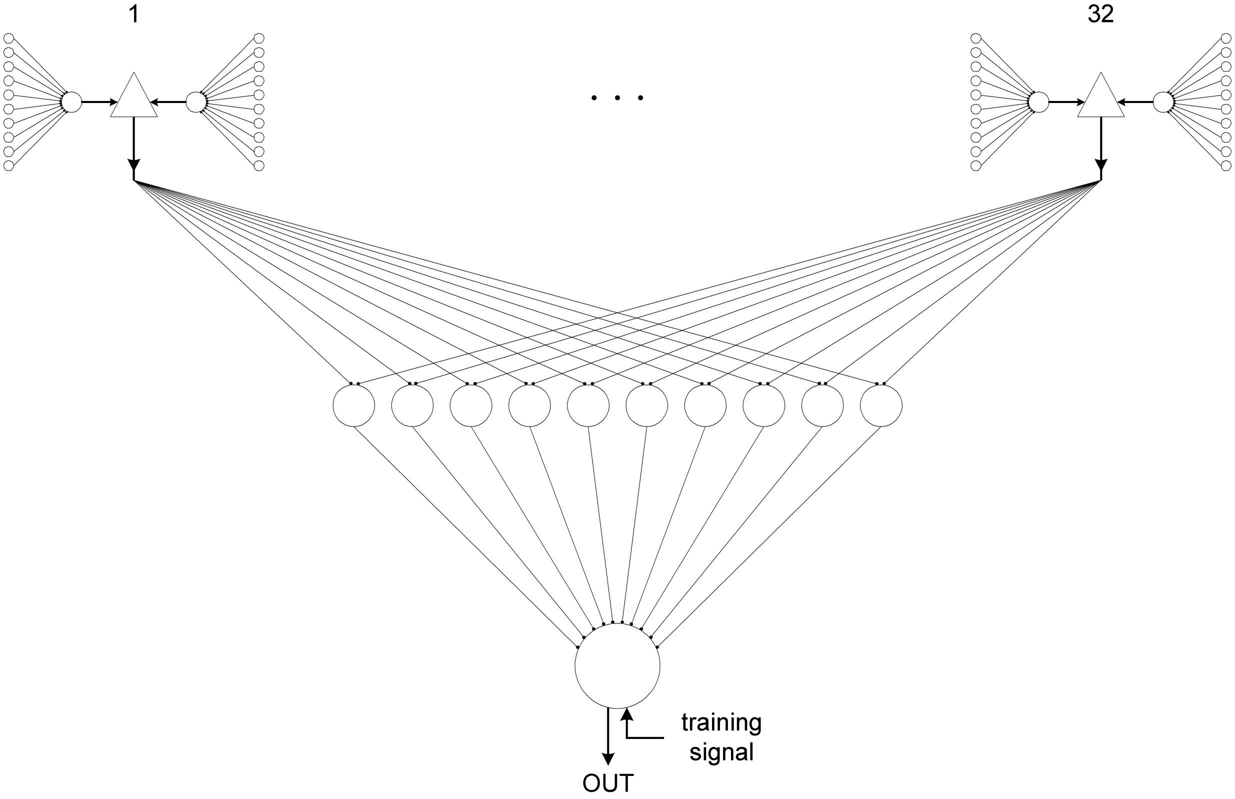


Figure 8. A layer of 32 SINBAD cells whose outputs are connected to a test dendrite, which is modeled as a backpropagation network with one output unit and a single layer of hidden units.

The three test dendrites are modeled the same way as the dendrites of the 32 cells, except that the test dendrite receives connections from all 32 cells (or 32 input channels, in the raw input test) and its output is passed through a sigmoid function:

$$D = \tanh(\sum_{h=1}^{40} w_h \cdot H_h). \tag{19}$$

The error signal is computed as:

$$\delta = (TR - D) \cdot D', \tag{20}$$

where *TR* is the training signal (the status of α, β, or Ω). Effectively, these test dendrites are modeled as standard three-layer backpropagation networks.

To establish the basis for comparison, we first trained the test dendrites on raw sensory information, by using input channels IN_1 - IN_{32} for their inputs. The results are shown in Figure 9A, for the test dendrite trained on object α (see the plot on the left) and for the test dendrite trained on combination Ω (right plot). Each plot shows the *error* (which is computed as |*TR* - *D*|) as a function of time since the start of the training period. As the left plot shows, the α dendrite gradually succeeded in learning to recognize object α. The right plot, on the other hand, shows that the Ω dendrite failed to learn to recognize combination Ω. This failure is not surprising, since Ω is a third-order combination, and backpropagation networks have great difficulties learning such complex functions (Clark and Thorton, 1997).

We next trained the test dendrites on sensory information processed by the 32 cells: outputs of the 32 cells were used as inputs to the test dendrites. The results are shown in Figure 9B, again for the α dendrite on the left and for the Ω dendrite on the right. As these two plots show, the learning performance of the test dendrites improved dramatically. The α dendrite learned to recognize object α orders of magnitude faster than when it learned from the raw input channels.

And the Ω dendrite - which before could not learn at all - now did learn to recognize combination Ω. What is particularly impressive is that the test dendrites learned to respond correctly to the input patterns after being presented with only a small fraction (<0.5%) of all possible input patterns. That is, the test dendrites showed perfect generalization abilities; they discovered the logic underlying the relationship between the input patterns and the behaviors on which they were trained.

The performance of the Ω dendrite is especially impressive as it was able to learn a very challenging, third-order combination of input channels. This task seemed impossible when we discussed it earlier in the example of a hypothetical animal that needed to learn how to distinguish between the two types of input patterns shown in Figure 5. After training, the Ω dendrite became able to distinguish between them; it happens (by our design) that the two sets of patterns in Figure 5 differ in that in one set the Ω combination of environmental conditions is absent (Ω = 0), in the other set it is present (Ω =1).

What is the nature of the information preprocessing, carried out by the layer of the 32 pyramidal cells, that improved so dramatically the learning abilities of the test dendrites? During the period of the network's learning, pairs of dendrites in each cell taught each other to tune to co-occurring combinations of environmental conditions. Significantly, these combinations of conditions were due to environmental properties *a - h*. With the dendrites tuned to individual predictable combinations, the entire cells tuned to recognize the presence and absence of those environmental properties. As we discussed above, due to random assignment of input channels on the dendrites, different cells tuned to different properties, but as a group, the 32 cells were likely to discover all 8 of them, *a* through *h*. In this way, the information about the status of *a - h* was brought to the surface in the 32 cells' outputs.

With environmental properties *a - h* represented directly by individual pyramidal cells, the learning task for the test dendrites was simplified: for the α dendrite, for example, the task was changed from that of learning a second-order nonlinear combination of input channels *1 - 32* to much easier one of learning a first-order combination of properties *a - h*. For the Ω dendrite, its task was simplified from a third-order combination to a second-order one.

If this interpretation of the Figure 9B results is correct, then the improvement in learning by test dendrites should be limited only to *orderly* combinations of environmental conditions. If we were to ask the test dendrites to learn some second- or third-order combinations of input channels that do not involve properties *a - h*, but are simply random in their composition, then information preprocessing by the 32 cells should not be of any help, since the cells will not make explicit the "building blocks" of such accidental combinations. To test this prediction, we trained the test dendrites on the outputs of the 32 cells, but using different training signals. The α and β dendrites were trained to recognize second-order combinations of elementary environmental conditions (just as objects α and β are such second-order combinations), but these new combinations had random compositions, not involving any of environmental properties *a - h*. Analogously, the Ω dendrite was trained to recognize a random third-order combination.

The results are shown in Figure 9C, and they confirm our expectation. Note especially that the learning performance of the dendrite trained on a second-order combination (shown in the left plot) is much worse than when that dendrite was trained to recognize a comparable combination (i. e., object α) from the raw information provided by input channels (see left plot in Figure 9A). This deterioration of learning performance is not surprising, suggesting that the 32 cells made some combinations of environmental conditions - the orderly ones - more explicit, in part, by filtering out accidental combinations.

To conclude this modeling demonstration, we find that the network of cells with backpropagating dendrites discovers high-order regularities in the environment remarkably easily, even when the input connections are distributed among dendrites at random. With their dendrites tuning to co-occurring combinations of environmental conditions, cells learn to recognize the orderly features of the environment that cause these regular combinations. This in turn enables dendrites in the next network to discover even more complex, higher-order co-occurring combinations of environmental conditions, as we showed with the α and Ω test dendrites.

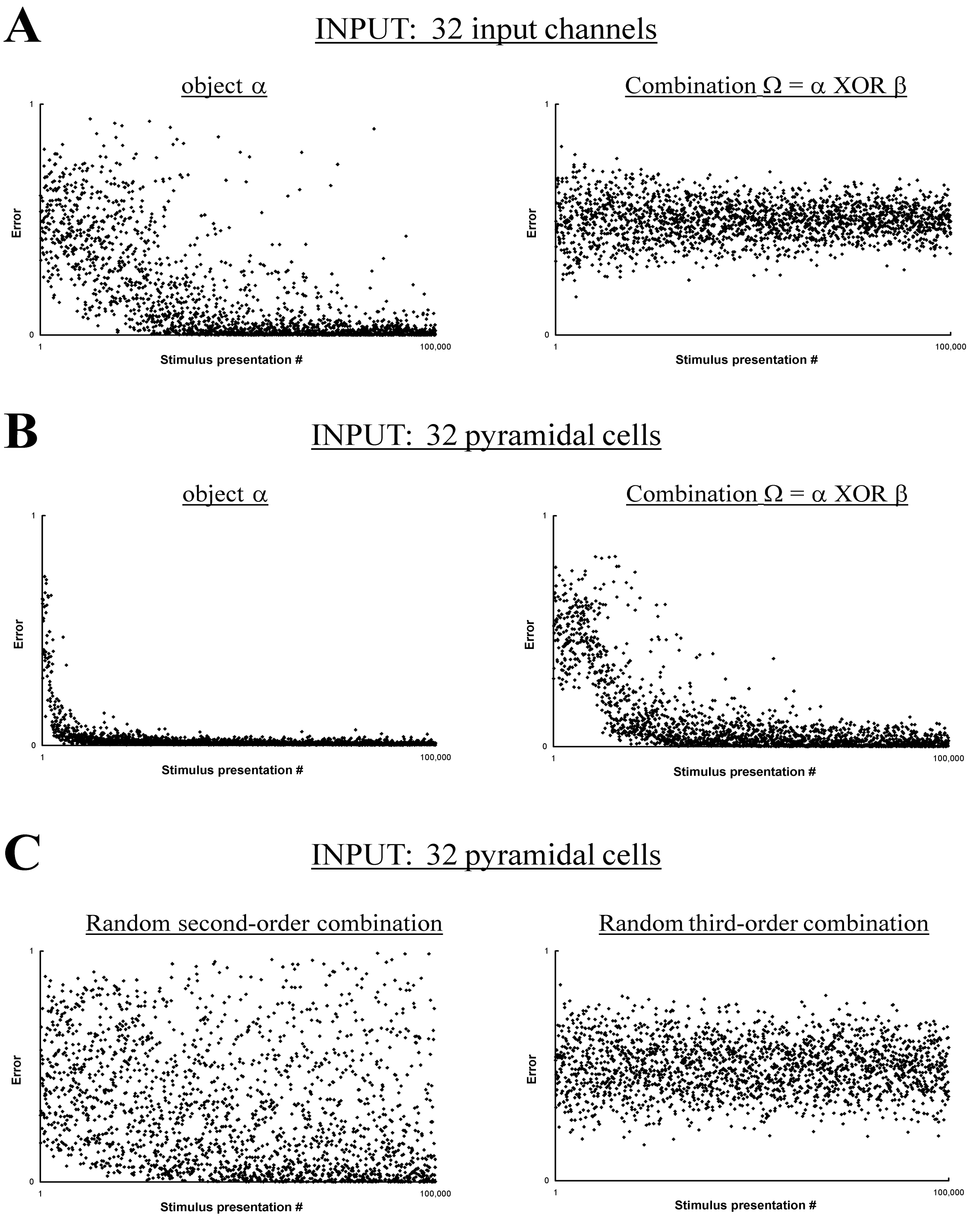


Figure 9. Learning performance of two test dendrites (left and right columns) in response to a random sequence of 100,000 input patterns. The error plotted is the magnitude of the difference between the training signal *TR* and the test dendrite's output *D*, showing how well the dendrite predicted the training signal. Plotted are responses only to every 100th input pattern. **A.** Two test dendrites, trained to recognize object α (left) and combination **Ω** of objects α and β, received their inputs directly from the 32 input channels, rather than from the 32 SINBAD cells. **B.** The α and **Ω** test dendrites received their inputs from the 32 SINBAD cells, rather than directly from the input channels. **C.** Two test dendrites, with their inputs coming from the 32 SINBAD cells, were trained to recognize two randomly defined second- and third-order combinations of elementary environmental conditions.

In our modeling demonstration we did not take advantage of any of a number of readily available, biologically realistic means by which the arrangement of input connections on the dendrites could be optimized. Their detailed demonstration is beyond the scope of this paper, but we ought to briefly mention some of them:

> **Lateral anti-Hebbian inhibitory connections.** In real cortical networks, pyramidal cells inhibit each other's somata (they do not do it directly, but by activating local inhibitory cells, which in turn inhibit other pyramidal cells). If these inhibitory disynaptic connections are anti-Hebbian (the *inhibitory* strength of a connection increases with the correlation in behaviors of the presynaptic channel and the postsynaptic soma), then pyramidal cells that happen to develop similar functional properties will also strengthen their inhibitory connections on each other's somata. But, as our unreported modeling results show, stronger inhibitory interactions will drive these cells to modify their functional properties to make them less similar than before. Thus, anti-Hebbian inhibitory interactions among somata of pyramidal cells will drive these cells to tune to different orderly features of the environment. The network, as a whole, will maximize the number of regularities it will be able to discover in the environment.
>
> **Number of dendritic compartments.** Dendrites might fail to discover predictable combinations of environmental conditions simply because they do not have enough dendritic compartments (or, thinking of dendrites as backpropagation networks, they might have an insufficient number of hidden units). Thus, if a pyramidal cell finds that its dendrites cannot, after a reasonably long period of trying, learn to produce matching outputs, then one possible remedy is to add more compartments to its dendrites, i.e., to add more dendritic branches and/or lengthen the existing ones. This might be a basic strategy: pyramidal cells might start with very small and poorly branched principal dendrites and gradually elaborate their dendritic trees until the principal dendrites succeed in finding something orderly in the environment. (This strategy has been also proposed, on more general grounds, by Quartz and Sejnowski, 1997.) This strategy would most efficiently match the sizes of dendritic trees to the demands of their tasks. Possibly related to this idea is the fact that pyramidal cells in the animals reared in behaviorally more enriched (i.e., more complex) environments have more elaborate dendritic trees (reviewed by Quartz and Sejnowski, 1997).
>
> **Topographic map.** In our last modeling demonstration we connected each cell to *all* the input channels, but in networks with larger numbers of input channels and, of course, in the cortex this approach would be both practically impossible and functionally detrimental. Instead, we can take advantage of the fact that most of orderly relations in natural environments are local in one way or another. For example, lower-order regularities involve environmental conditions in close spatial proximity to each other; consequently, exposing a pyramidal cell in a primary sensory cortex to raw information from distant spatial locations would be useless. In agreement with this observation, afferent connections to cortical areas do not all contact each and every cell but have clear topographic organization (e.g., body maps in somatosensory and motor cortices, retinotopic maps in visual cortex, etc.). These maps are created in middle cortical layers by a host of genetic and epigenetic mechanisms (von der Malsburg and Singer, 1988). From the viewpoint advanced in this paper, we expect that the mechanisms that control perinatal development and adult maintenance of cortical topographic maps are designed to supply each cortical neighborhood with limited but functionally related information in order to improve its chances of discovering orderly environmental features. Some of these mechanisms, in particular the ones that operate very locally, among neighboring cortical minicolumns, are described in detail in Section III,

together with their potential significance for the development of functional properties of SINBAD neurons.

Trial-and-error rearrangement of connections on dendrites. While the initial arrangement of input connections on a cell's dendrites can only be random, it can later be changed, if the dendrites fail to find anything orderly. After a period of unsuccessful learning attempts, few randomly chosen connections might be dropped, while other new connections might be added. This would involve some sprouting of axon collaterals, something that takes place even in the adult cortex (Darian-Smith and Gilbert, 1994; Florence et al., 1998). Dendrites can continue to "experiment" with their input connections until they find co-occurring combinations of environmental conditions.

Usefulness of a cell's output. A pyramidal cell might discover some regularity in the environment, but that regularity might turn out to be inconsequential, of no significance to any other pyramidal cell. (Intuitively, the most useful regularities are those that are most predictive of other regularities, or those that are most relevant to behavior.) In that case it would be functionally desirable for the first pyramidal cell to discard the useless regularity and find another one. This mechanism is easy to implement by monitoring the sum of weights of all the connections a given pyramidal cell has on other cells. If no other cell makes use of a given cell's output, then the sum of its synaptic weights will be zero; in that case, after a reasonable period of time the cell can drop and add some of its own input connections, thus forcing its dendrites to find another regularity in the environment that may prove more useful. In the real cortex, monitoring the sum of connection weights might be carried out by well-known trophic signals that presynaptic cells receive from their postsynaptic counterparts (Purves, 1988; Thoenen, 1995); the net trophic signal reaching a cell's soma is indicative of the number and weights of synapses it has on other cells.

Multiple networks. Obviously, a single layer, or network, of pyramidal cells is limited in the complexity of orderly relations it can discover in the environment. However, if its output is fed to another network of pyramidal cells (e.g., primary visual cortical area V1 projecting to V2, etc.), then the higher network will be able to extract higher-order environmental regularities that have as their building blocks the regularities discovered by the lower network. In this way, a series of networks can discover very complex orderly relations in the environment. Higher cortical areas, in turn, provide feedback connections to the lower areas (Van Essen et al., 1992), where they can participate in shaping cells' tuning properties and enable higher-order regularities to help in discovering lower-order ones.

III. CORTICAL MINICOLUMNAR ORGANIZATION AND SINBAD NEURONS

So far we have focused on implementation of the SINBAD design in individual pyramidal cells. In this section we take a step back and look at how such SINBAD cells might be organized in larger functional cortical structures. The functional structures that we consider in this section are cortical *minicolumns* (Mountcastle, 1978).

Minicolumns are 0.05 mm diameter cords of cells extending radially across all cortical layers; each minicolumn is distinguished from its immediate neighbors by its receptive field and functional properties. A common misconception is that cells located so closely to each other have very similar receptive fields. In fact, most of the experimental literature in somatosensory, visual, auditory, motor, and associative cortical areas is in agreement that while neighboring cortical cells can show a remarkable uniformity in some of their receptive field properties (e.g. stimulus orientation in visual cortex), they can differ prominently in other of those properties (for a review, see Favorov and Kelly, 1996). When receptive fields are considered *in toto*, in all their dimensions, neighboring neurons typically have little in common - a stimulus which is effective in driving one cell will frequently be much less effective in driving its neighbor.

This prominent local receptive field diversity is constrained, however, in the radial cortical dimension. Cells that make up individual radially oriented minicolumns have very similar receptive field properties (Abeles and Goldstein, 1970; Hubel and Wiesel, 1974; Albus, 1975, Merzenich *et al.*, 1981; Favorov and Whitsel, 1988; Favorov and Diamond, 1990; summarized in Favorov and Kelly, 1996). If neighboring neurons have contrasting functional properties, they are likely to belong to different minicolumns.

A number of elements of cortical microarchitecture are responsible for the existence of minicolumnar functional units - among them excitatory *spiny stellate* cells and inhibitory *double bouquet* cells (Jones, 1975, 1981; Lund, 1984; Somogyi and Cowey, 1984). Spiny stellates are excitatory intrinsic cells located in layer 4. They are the major recipients of afferent connections from the thalamus or preceding cortical areas; in turn, they distribute afferent input radially via narrow bundles of axon collaterals to other cells in the same minicolumn (see a connectional diagram in Figure 10), thus imposing on it a uniform set of functional properties.

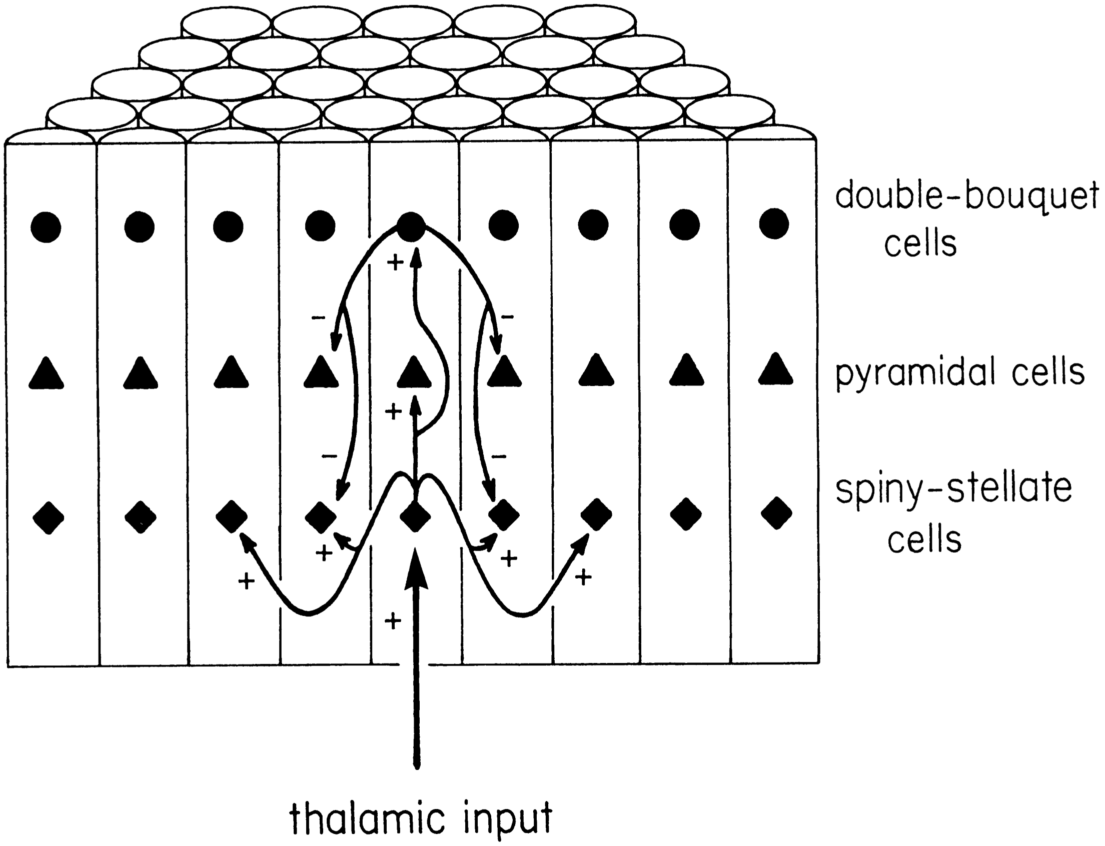


Figure 10. Pattern of minicolumnar connections (modified from Favorov and Kelly, 1994a). Shown here is a section across an idealized cortical region, represented as a tightly packed field of cylinder-shaped minicolumns, each containing one representative spiny stellate, pyramidal, and double bouquet cells. Afferent, intra-minicolumnar, and local inter-minicolumnar connections are shown for one minicolumn.

Double bouquet cells are GABAergic cells with bodies and local dendritic trees in layer 2 and the upper part of layer 3. Their axons descend in tight bundles of collaterals down through layers 3 and 4 and into layer 5, making synapses all along the way on the distal dendrites of pyramidal and spiny stellate cells, but avoiding the main shaft of apical dendrites (DeFelipe *et al.*, 1989; DeFelipe and Farinas, 1992). Due to this arrangement, double bouquet cells are more likely to inhibit cells in adjacent minicolumns rather than in their own (Figure 10), and thus they offer a mechanism by which adjacent minicolumns can inhibit each other.

To explain why adjacent minicolumns have noticeably different receptive fields, Favorov and Kelly (1994a,b; 1996) have proposed that during perinatal development each minicolumn is driven by its inhibitory, double bouquet-mediated interactions with adjacent minicolumns (as shown in Figure 10) to acquire a set of afferent connections that is different from those of its immediate neighbors. On the other hand, each minicolumn is also driven by the excitatory interactions with a larger circle of its neighbors (see Figure 10) to make its set of afferent connections similar to theirs. To satisfy these opposing pressures, minicolumns in local cortical territories arrange their afferent connections in permuted patterns, with shuffled receptive fields. (Shuffling of receptive fields of a local group of minicolumns satisfies the opposing pressures by moving receptive fields of adjacent minicolumns farther apart, while preventing receptive fields of the entire group from diverging too widely).

Because of the prominent differences in functional properties among neighboring minicolumns, peripheral stimuli can be expected to evoke spatially complex minicolumnar patterns of activity in the engaged cortical region, with a mixture of active and inactive minicolumns (Favorov and Kelly, 1994b). This expectation has been experimentally confirmed in studies of stimulus-evoked activity in somatosensory cortex using either 2-deoxyglucose (2-DG) metabolic labeling (McCasland and Woolsey, 1988; Tommerdahl *et al*., 1993) or near-infrared optical imaging of the intrinsic signal (Figure 11). Our modeling studies (Favorov *et al*., 1994b) predict that these minicolumnar activity patterns should be highly stimulus specific and carry detailed information about stimulus features.

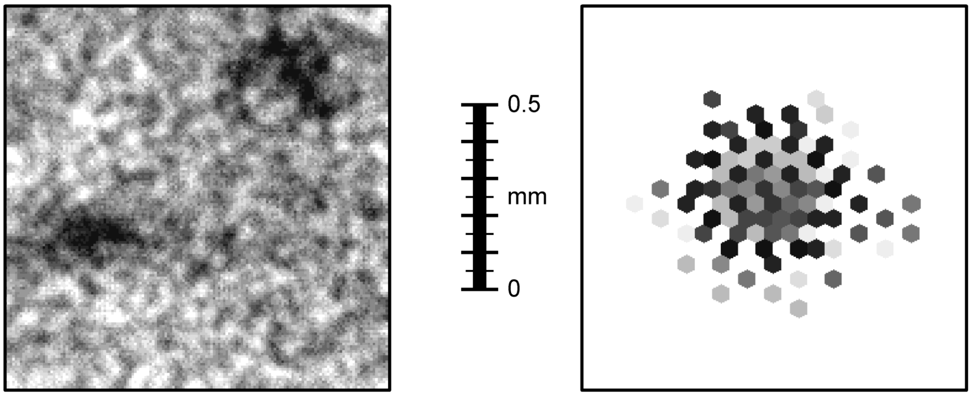


Figure 11. Minicolumnar pattern of activation of cat somatosensory cortex: optical imaging of the stimulus-evoked intrinsic signal (left) and modeling prediction (right). The optical image was obtained in the near-infrared range (830 nm; for details about the method, see Tommerdahl *et al.*, 1999). It shows two activated cortical regions, the bottom one in area 3b, the top one in area 3a. Upon closer inspection, the active regions appear as a patchwork of minicolumn-sized spots. Note that these spots tend to be organized in short parallel strings, and that orientation of these strings varies across the activated region. The image on the right was generated by our minicolumnar model of somatosensory network (Favorov and Kelly, 1994a,b). It shows a spatial pattern of active minicolumns, driven by a punctate stimulus. This pattern is similar to the one on the left in that in it active minicolumns are also organized in short parallel strings that run in different directions in different parts of the activated region.

To conclude this brief description of cortical topographic organization at the minicolumnar level, it appears that local groups of minicolumns bring together a variety of different, but related sensory information concerning a local region of the stimulus space, with adjacent minicolumns tuned to extract only minimally redundant information about what takes place in that stimulus space. Because this information comes from a local region of the stimulus space, it is likely to be rich in orderly relations reflecting the orderly features of the outside world. Thus, it appears that local cortical neighborhoods are designed to create local informational environments enriched in regularity, but low in redundancy.

Such local cortical environments are exactly the right informational environments for SINBAD neurons, to be “mined” by them in their search for orderly relations. Each dendrite of a SINBAD pyramidal cell will be extended through a functionally distinct group of minicolumns, exposing it to a unique combination of afferent information. Principal dendrites of the same cell will thus be forced to find different combinations of environmental conditions that are predictably related, as we have shown in the preceding section, and the cell as a whole will tune to the orderly feature of the environment responsible for that relationship.

Thus, we see the progressive elaboration of functional properties of neurons across cortical areas as a repeated two-step process. The first step takes place among spiny stellate cells in the cortical input layer, layer 4. This step involves the development of afferent connections to spiny stellates. This development is controlled by lateral interactions among minicolumns, operating on a number of different spatial scales (inhibitory among adjacent minicolumns, excitatory among farther neighbors, back to inhibitory among even farther neighbors - this a blend of the classical “Mexican hat” pattern of lateral interactions and the double bouquet pattern we described above; for details, see Favorov and Kelly, 1994b). The task of this development is to organize inputs to the cortical area so that, as we said above, cortical neighborhoods will be supplied with sensory information that is enriched in regularities (by keeping it local but non-redundant).

The second step in the elaboration of neuronal functional properties takes place among pyramidal cells in the upper and deep cortical layers. Spiny stellate cells provide their outputs to pyramidal cells lying above and below them in the same and nearby minicolumns. Pyramidal cells receive from spiny stellates information about some environmental conditions and, in turn, they learn to respond to particular combinations of those conditions, combinations that reflect orderly features of the environment. Pyramidal cells send their outputs to the next cortical area, where the two-step process of building functional properties is repeated, tuning pyramidal cells there to still higher-order combinations of combinations of the environmental conditions, etc.

In conclusion, we view the cortical network as a system that by learning the orderly features of its sensory environment builds an internal model of that environment, along the lines of Craik (1943) and Barlow (1992). This internal model is likely to endow the network with powerful interpretive and predictive capabilities. These capabilities can be used to infer what has happened, is happening, will happen, or what should be done in any particular situation faced by the animal. How such capabilities might be put into practice by a cortical network, and how the entire set of cortical networks composing the cerebral cortex might work together to produce adaptive behaviors are now ready for study.

IV. ASSOCIATIONISM

Mr. Darwin, by a strange inversion of reasoning, seems to think Absolute Ignorance fully qualified to take the place of Absolute Wisdom in all the achievements of creative skill.

Robert MacKenzie, *The Darwinian Theory of the Transmutation of Species Examined*, 1868.

The traditional, fundamental, and decisive objection to association is that it is too stupid a relation to form the basis of a mental life.

Jerry Fodor, *Modularity of Mind*, 1983.

SINBAD as an associationist theory

Our proposal is empiricist and associationist. That is, we think the operations of the mind are founded upon a basic neural mechanism that serves to extract regularities from the environment. This extraction of regularity results in the formation of perceptual, and ultimately conceptual structure. By the same token, it allows an organism to make predictions about its environment, and to act accordingly.

Of course this intuitive idea has a long history. The seeds of associationism can be traced back to Plato and Aristotle (Young, 1968), but it first came into its own in 17th and 18th century Britain. The British Empiricists, starting with Locke, opposed the rationalist view that we are born knowing things – we don't have any innate concepts, we don't innately know that anything is the case, and we don't innately know how to do anything. Locke (1700/1978) hypothesized that we acquire knowledge in accordance with principles of association, principles which may be explained by appeal to a small set of "original qualities of human nature" as Hume put it in his *Treatise* (1739/1978). So, for the British Empiricists, what is innate is not knowledge, but rather a *basic mechanism* for the acquisition of knowledge from experience.

The first philosopher to provide a systematic theory of the principles of association and their mechanism was David Hartley, in his *Observations on Man* (1749/1970). He identified two basic principles of association, *synchronic* and *successive*. Hartley also incorporated into his system associative links between ideas and movements, and mentioned the influence of reinforcement on associative strength. But neither of these are essential to his associationism (nor was it to Locke's or Hume's). It is unfortunate, then, that the last great research tradition in an associationist vein - Skinner's behaviorism - foundered on its exclusive reliance upon the *reinforcement of behavior*, and its eschewal of explanations mentioning internal states.

In modern times associationism has become more quantitatively model-oriented, leading to an appreciation of the ability of associative networks to generalize and do similarity-based computations, but it has so far failed to produce an associationist model of cortical learning that is both powerful enough and clearly biologically realistic (Clark, 1993). We believe that our model makes a significant advance in this direction. It is formulated in biologically explicit terms, and thus is eminently experimentally testable. It is self-organizing and, most importantly, it is capable of discovering nonlinear and higher-order regularities. This is the main advance we offer over previous connectionist theories.

Countering nativist arguments

Locke's opponents were the rationalists, like Descartes, who posited innate ideas and knowledge. The modern day rationalists are Chomsky, Fodor, and other nativists, who posit genetically endowed concepts and knowledge. In contrast to nativists, our research program is to push the empiricist/associationist line as far as possible before admitting the existence of innate concepts, knowledge, or even biases. This seems a sensible strategy: "It's been there all along" should be a last resort in etiological explanation.

Of course, a radical empiricist position is not plausible for the whole brain; one is born, for instance, knowing how to breathe. Some learning is required even in the autonomic nervous system – (e. g., Hamill and LaGamma, 1992). But this is not likely to be *associationist* learning. In many subcortical nuclei, for which the nativist case it strong, learning probably involves a domain specific mechanism, similar to those described by Gallistel (1995). This might be called "nativist learning," and it resembles what Chomsky has in mind with his more recent "principles and parameters" account of language acquisition (e.g., Chomsky, 1988).

So in the current proposal, our empiricism amounts to an insistence upon an associationist learning mechanism in a largely equipotential *cortex*. This is emphatically *not* equivalent to denying the presence of any constraints on learning. *Any* associationist mechanism brings in constraints. There is an "argument" against empiricism that goes like this: "You can't get something from nothing. *Some* mechanism must be innate, which requires that there be *some* constraints on learning, so empiricism must be false." But no empiricist has ever denied that there are innate mechanisms in the mind. In fact, associationism *requires* it - recall Hume's appeal to "original qualities of human nature." The empiricist-rationalist opposition concerns not innate structure generally speaking, but rather innate knowledge and concepts. That is, the empiricist denies, and the rationalist accepts, that we have innate knowledge and/or concepts. Empiricist learning theory, i.e. associationism, dictates that knowledge and concepts are acquired through experience.

Together with others (e. g., Clark, 1993; Quartz and Sejnowski, 1997), we believe our proposal for an associationist, empiricist neural model of the cortex provides a plausible alternative to the “hodgepodge of specialized circuits” (Quartz and Sejnowski, 1997) that nativists advocate. However, there are those who feel that associationism is not a genuine alternative, or at least that it is so highly implausible that it is not worth bothering with. They claim that an associationist mechanism plus the environment are insufficient by themselves to account for knowledge we demonstrably have. Therefore, some knowledge must be innate. For example, when children learn a language, they invariably settle on a number of grammatical rules, rules that are independent of any particular language. This universality cannot be explained by an associationist mechanism since the exemplars children have to learn from are compatible with an indefinite number of *other* rules. Therefore the rules actually used must be innate, and not learned.

Whatever the strength of Chomsky's "poverty of the stimulus" argument in linguistics (see Cowie, 1998 for a trenchant critique), its plausibility is greatly diminished when transferred to the general perceptual domain. First, the argument appeals to the fact that children are not typically reinforced for correct language use, and correction has little effect on their subsequent linguistic behavior - so language acquisition must involve a preset sequence of events that merely unfolds. Though patterns of reinforcement may be relevant to the acquisition of linguistic and other behavior, they are irrelevant to a passive, self-organizing order-extracting mechanism like SINBAD. Second, the poverty of the stimulus argument relies on a premise to the effect that children are exposed to a very partial, degraded, and even flawed data set (since adults often use grammatically incorrect speech.) The equivalent premise simply does not hold for general perceptual mechanisms. The rules that perceptual AI research proposes as innate are, for example, those used to parse a visual scene to arrive at a three dimensional model of it. These rules are there to be discovered in an animal's first few glances at a natural scene, in its first tactile interactions with the world, etc. And the world does not make mistakes. (It is interesting to note here a trend in AI research of rejecting sets of rules or algorithms as *insufficiently general*, leading to models more closely approximating an associationist mechanism [see e.g. Hildreth and Ullman, 1989; Poggio and Hurlbert, 1994].) The evidence does not indicate the presence of innately determined rules that are followed to the exclusion of other rules consistent with the environmental data set. So a poverty of the stimulus argument cannot be given to support complex innate structure in general perceptual mechanisms.

But there is another argument of the same form that is open to the nativist. The argument form, recall, is this: an associationist mechanism plus the environment cannot explain how we have the knowledge we in fact do, so some innate knowledge must be postulated. If the environmental regularities an animal makes use of are "hidden from view," so to speak, because they are higher-order regularities, the animal's knowledge of these regularities cannot be explained by the typical self-organized associationist mechanism. This variation on the standard argument for innateness has been called a version of the poverty of the stimulus argument (Kirsh, 1992). The original version can be conveniently summarized as the underdetermination of a perceptual (or linguistic) system's "theory" of a domain by the evidence available to it. In the variation on the theme, the system's "theory" is not underdetermined. It is simply very difficult to arrive at, since the regularities are higher-order ones. Clearly, SINBAD constitutes a promising response to this argument. As we saw in the final modeling demonstration, SINBAD was capable of discovering combinations of combinations of combination of features. In other words, it discovered higher-order regularities, and is capable of finding "type-2 mappings" (Clark and Thornton, 1997). In conjunction with a mechanism for ensuring that all cells choose "useful" regularities (i.e., ones that other cells take advantage of - see the "usefulness of a cell's output" at the end of part II), we have an extremely powerful extractor of latent order from the environment.

One further note on the linguistic case: Chomsky's argument relies on the premise that grammar and semantics are independent, in the sense that a child learns her grammar without relying upon semantic information. If Chomsky is wrong about this, and semantic information is actually relevant to learning grammar, then providing a plausible empiricist account of language learning may actually involve surmounting the *second* rather than the first version of the poverty of the stimulus argument. SINBAD might be up to this task.

Due to the relatively weak order-extracting capabilities of other self-organizing networks (Clark, 1993), many nativists underestimate the power of association. In the quotation at the beginning of this section, Jerry Fodor underestimates associative learning, just like Mackenzie underestimated evolution. Both underestimate the power of a self-organizing system.

We agree with Fodor (1983) that the problem with traditional associationism – Locke, Hume, and Hartley's associations between ideas – is that it is too impoverished, based as it is on correlations of raw sensory inputs. Fodor conceives of a dressed-up, modern, learning-theoretic, computational associationist. This kind of associationist does not limit himself to Hartley's sympathetic vibrations or Hebb's synaptic modifications. He postulates other associative relations: e.g., the logical functions *OR*, *exclusive-OR* , and more complex functions are defined, making for a small set of associative relations rather than only one.

Fodor's main complaint about computational associationism is as follows. The associationist is forced to admit more complex operations as fundamental in order to account for more complex mental structure. But an acknowledgment of the complexity of mental structure conflicts with the associationist's account of ontogeny (p. 32-34):

> In short, as the operative notion of mental structure gets richer, it becomes increasingly difficult to imagine identifying the ontogeny of such structures with the registration of environmental regularities.... To put the point in a nutshell, the crucial difference between classical and computational associationism is simply that the latter is utterly lacking in any learning theory.

That is, he thinks the classical associationists had a learning theory (association by the constant conjunction of raw stimuli), but that it could not account for the complexity of mental structure. Modern associationism, by contrast, has a better chance of accounting for the complexity of mental structure, but it is mysterious how a system could *acquire* this mental structure by extracting regularities from the environment. We think that SINBAD may remove the mystery by showing how nonlinear and higher-order regularities can be extracted from the environment by a simple associationist mechanism.

Neuroscientists are also looking for such a mechanism. In their response to the commentaries on their recent (1997) Behavioral and Brain Sciences article, Phillips and Singer ask (p. 709):

> *Does unsupervised learning in the cortex discover higher-order variables?*
> In section 6.4 we asked whether there is any evidence that self-organization in the cortex can discover nonlinear variables such as XOR. No such evidence was offered in the commentaries, nor have we yet found any from other sources. The continued failure of such evidence to appear suggests that reliable discovery of such nonlinear variables may not be a fundamental capability of cortex.

We hope that Phillips and Singer will take heart. If our model is correct, the discovery of nonlinearly separable functions and higher order relations is what the cortex is best at.

ACKNOWLEDGEMENTS: This work was supported in part by ONR grant N00014-95-1-0113 and NIH grant NS35222. DR was supported by the Social Sciences and Humanities Research Council of Canada.

REFERENCES

Abeles, M. and Goldstein, M. H., Jr. (1970) Functional architecture in cat primary auditory cortex: columnar organization and organization according to depth. *J. Neurophysiol.* **33**: 172-187.

Albus, K. (1975) A quantitative study of the projection area of the central and the paracentral visual field in area 17 of the cat: I. The spatial organization of the orientation domain. *Exp. Brain Res.* **24**:181-202.

Artola, A., Brocher, S. and Singer, W. (1990) Different voltage dependent thresholds for the induction of long-term depression and long-term potentiation in slices of rat visual cortex. *Nature* **347**: 69-72.

Barlow, H. B. (1992) The biological role of neocortex. In *Information Processing in the Cortex*, Aertsen, A. and Braitenberg, V. (eds), Springer, Berlin, pp. 53-80.

Brown,T.H., Kairiss, E.W. and Keenan, C.L. (1990) Hebbian synapses: biophysical mechanisms and algorithms. *Annu. Rev. Neurosci.* **13**: 475-511.

Chomsky, N. (1988) *Language and Problems of Knowledge*. MIT Press, Cambridge, MA.

Clark, A. (1993) *Associative Engines: Connectionism, Concepts, and Representational Change.* MIT Press, Cambridge, MA.

Clark, A. and Thornton, C. (1997) Trading places: computation, representation, and the limits of uninformed learning. *Behav. Brain Sci.* **20**: 57-90.

Cowie, F. (1998) *What's Within.* Oxford University Press, Oxford.

Craik, K. J. W. (1943) *The Nature of Explanation*. Cambridge University Press, London.

Darian-Smith, C. and Gilbert, C. D. (1994) Axonal sprouting accompanies functional reorganization in adult cat striate cortex. *Nature* **368**: 737-740.

DeFelipe, J., Hendry, M.C. and Jones, E.G. (1989) Synapses of double bouquet cells in monkey cerebral cortex visualized by calbindin immunoreactivity. *Brain Res*. **503**: 49-54.

DeFelipe, J. and Farinas, I. (1992) The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs. *Prog. Neurobiol.* **39**: 563-607.

Deuchars, J., West, D. C. and Thomson, A. M. (1994) Relationships between morphology and physiology of pyramid-pyramid single axon connections in rat neocortex in vitro. *J. Physiol.* (*Lond.*) **478**: 423-435.

Favorov, O.V. and Whitsel, B.L. (1988) Spatial organization of the peripheral input to area 1 cell columns: I. The detection of "segregates." *Brain Res. Revs* **13**: 25-42.

Favorov, O.V. and Diamond, M.E. (1990) Demonstration of discrete place-defined columns - segregates - in the cat SI. *J. Comp. Neurol.* **298**: 97-112.

Favorov, O.V. and Kelly, D.G. (1994a) Minicolumnar organization within somatosensory cortical segregates: I. Development of afferent connections. *Cereb. Cortex* **4**: 408-427.

Favorov, O.V. and Kelly, D.G. (1994b) Minicolumnar organization within somatosensory cortical segregates: II. Emergent functional properties. *Cereb. Cortex* **4**: 428-442.

Favorov, O. V. and Kelly, D. G. (1996) Local receptive field diversity within cortical neuronal populations. In *Somesthesis and the Neurobiology of the Somatosensory Cortex*, Franzen, O., Johansson, R. and Terenius, L. (eds), Birkhauser, Basel, pp. 395-408.

Feldman, M. L. (1984) Morphology of the neocortical pyramidal neuron. In *Cerebral Cortex*, Peters, A. and Jones, E. G. (eds.), Plenum Press, New York, vol. 1, pp. 123-200.

Florence, S. L., Taub, H. B. and Kaas, J. H. (1998) Large-scale sprouting of cortical connections after peripheral injury in adult macaque monkeys. *Science* **282**: 1117-1121.

Fodor, J. (1983) *Modularity of Mind.* MIT Press, Cambridge, MA.

Gallistel, C. R. (1995) The replacement of general-purpose theories with adaptive specializations. In *The Cognitive Neurosciences*, Gazzaniga, M. S. (ed.), MIT Press, Cambridge, Mass., pp. 1255-1267.

Gawne, T. J., Kjaer, T. W., Hertz, J. A. and Richmond, B. J. (1996) Adjacent visual cortical complex cells share about 20% of their stimulus-related information. *Cereb. Cortex* **6**: 482-489.

Hamill, R. W. and LaGamma, E. F. (1992) Autonomic nervous system development. In *Autonomic Failure: A Textbook of Clinical Disorders of the Autonomic Nervous System,* Bannister, R. and Mathias, C. J. (eds.), Oxford University Press, Oxford.

Hartley, D. (1749/1970) *Observations on Man*, selections in Robert Brown (ed.) *Between Hume and Mill: An Anthology of British Philosophy 1749-1843.* Random House, New York.

Hebb, D. O. (1949) *The Organization of Behavior: a Neuropsychological Theory*. John Wiley and Sons, New York.

Hildreth, E. C. and Ullman, S. (1989) The Computational Study of Vision. In *Foundations of Cognitive Science*, Posner, M. I. (ed.), MIT Press, Cambridge, MA, pp. 581-630.

Hubel, D. H. and Wiesel, T. N. (1974) Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J. Comp. Neurol.* **158**: 267-294.

Hume, D. (1740/1978) *A Treatise of Human Nature,* Selby-Bigge, L. A. (ed.), Oxford University Press, Oxford.

Jones, E. G. (1975) Varieties and distribution of non-pyramidal cells in the somatic sensory cortex of the squirrel monkey. *J. Comp. Neurol.* **160**: 205-267.

Jones, E. G. (1981) Anatomy of cerebral cortex: columnar input-output organization. In *The Organization of the Cerebral Cortex,* edited by F. O. Schmitt, Cambridge: MIT Press, pp.199-235.

Kirkwood, A., Gold, S. M. D. J. T., Aizenman, C. and Bear, M. F. (1993) Common forms of synaptic plasticity in hippocampus and neocortex *in vitro*. *Science* **260**: 1518-1521.

Kirsh, D. (1992) PDP Learnability and Innate Knowledge of Language. In *Connectionism: Theory and Practice*, Davis, S. (ed.), Oxford University Press, Oxford.

Komatsu, Y. (1994) Age-dependent long-term potentiation of inhibitory synaptic transmission in rat visual cortex. *J. Neurosci*. **14**: 6488-6499.

Komatsu, Y. and Iwakiri, M. (1993) Long-term modification of inhibitory synaptic transmission in developing visual cortex. *Neuroreport* **4**: 907-910.

Locke (1700/1978) *An Essay Concerning Human Understanding,* (4th ed.) P. H. Nidditch (ed.) Oxford: Oxford University Press.

Lund, J. S. (1984) Spiny stellate neurons. In *Cerebral Cortex,* edited by A. Peters and E. G. Jones, Plenum Press, pp. 255-308.

Malach, R. (1994) Cortical columns as devices for maximizing neuronal diversity. *TINS* **17**: 101-104.

Markram, H., Lubke, J., Frotscher, M., Roth, A. and Sakmann, B. (1997a) Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol.* (*Lond.*) **500**: 409-440.

Markram, H., Lubke, J., Frotscher, M. and Sakmann, B. (1997b) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**: 213-215.

McCasland, J. S. and Woolsey, T. A. (1988) High-resolution 2-deoxyglucose mapping of functional cortical columns in mouse barrel cortex. *J. Comp. Neurol.* **278**: 555-569.

McGuire, B., Gilbert, C. D., Wiesel, T. N. and Rivlin, P. K. (1991) Targets of horizontal connections in macaque primary visual cortex. *J. Comp. Neurol*. **305**: 370-392.

Mel, B. W. (1994) Information processing in dendritic trees. *Neural Comp* **6**: 1031-1085.

Merzenich, M. M., Sur, M., Nelson, R. J. and Kaas, J.H. (1981) Organization of the SI cortex: multiple cutaneous representations in areas 3b and 1 of the owl monkey. In *Cortical Sensory Organization*, edited by C. N. Woolsey, vol. 1, Humana Press, Clifton, N.J., pp. 47-66.

Mountcastle, V. B. (1978) An organizing principle for cerebral function. In *The Mindful Brain,* Edelman, G. M. and Mountcastle, V. B. (eds.), MIT Press, Cambridge, MA, pp. 7-50.

Pelletier, M. C. and Hablitz, J. J. (1996) Tetraethylammonium-induced synaptic plasticity in rat neocortex. *Cereb. Cortex* **6**: 771-780.

Phillips, W. A. and Singer, W. (1997) In search of common foundations for cortical computation. *Behav. Brain Sci*. **20**: 657-722.

Poggio, T and Hurlbert, A. (1994) Observations on Cortical Mechanisms for Object Recognition and Learning. In *Large-Scale Neuronal Theories of the Brain.* MIT Press, Cambridge, Mass., pp. 153-182.

Purves, D. (1988) *Body and Brain: a Trophic Theory of Neural Connections*. Harvard University Press.

Quartz, S. R. and Sejnowski, T. J. (1997) The neural basis of cognitive development: a constructivist manifesto. *Behav. Brain Sci.* **20**: 537-596.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of*

Cognition, Rumelhart, D. E., McClelland, J. L. and PDP Research Group (eds), MIT Press, Cambridge, Mass., vol. 1, pp. 318-362.

Schuz, A. (1992) Randomness and constraints in the cortical neuropil. In *Information Processing in the Cortex*, Aertsen, A. and Braitenberg, V. (eds), Springer, Berlin, pp. 3-21.

Segev, I., Fleshman, J. W., and Burke, R. E. (1989) Compartmental models of complex neurons. In Methods in Neuronal Modeling, Koch, C. and Segev, I. (eds), MIT Press, Cambridge, MA, pp. 63-96.

Singer, W. (1995) Development and plasticity of cortical processing architectures. *Science* **270**, 758-764.

Somogyi, P. and Cowey, A. (1984) Double bouquet cells. In *Cerebral Cortex*, Peters, A. and Jones, E. G. (eds.), Plenum Press, New York, vol. 1, pp. 337-360.

Stuart, G., Spruston, N., Sakmann, B. and Hausser, M. (1997) Action potential initiation and backpropagation in neurons of the mammalian CNS. *TINS* **20**: 125-131.

Thoenen, H. (1995) Neurotrophins and neuronal plasticity. *Science* **270**: 593-598.

Thomson, A. M. and Deuchars, J. (1994) Temporal and spatial properties of local circuits in neocortex. *TINS* **17**: 119-126.

Tommerdahl, M., Favorov, O.V., Whitsel, B.L., Nakhle, B. and Gonchar, Y.A. (1993) Minicolumnar activation patterns in cat and monkey SI cortex. *Cereb. Cortex* **3**: 399-411.

Tommerdahl, M., Delemos, A. K., Whitsel, B. L., Favorov, O. V. and Metz, C. B. (1999) Response of anterior parietal cortex to cutaneous flutter versus vibration. *J. Neurophysiol.* (in press).

Van Essen, D. C., Anderson, C. H. and Felleman, D. J. (1992) Information processing in the primate visual system: an integrated systems perspective. *Science* **255**: 419-423.

von der Malsburg, C. and Singer, W. (1988) Principles of cortical network organization. In *Neurobiology of Neocortex*, Rakic, P. and Singer, W. (eds), John Wiley and Sons, New York, pp. 69-99.

Widrow, B. and Hoff, M. E. (1960) Adaptive switching circuits. *1960 IRE WESCON Convention Record*, Part 4, IRE, New York, pp. 96-104.

Young, R. M. (1968) Association of ideas. In *Dictionary of the History of Ideas*, Wiener, P. P. (ed.), Charles Scribner's Sons, New York, vol. 1, pp. 111-118.