Bulletin of Mathematical Biology Vol. 51, No. 2, pp. 195–205, 1989.
printed in Great Britain.

0092–8240/89$3.00+0.00
Pergamon Press plc

CONTEXT-DEPENDENT ASSOCIATIONS IN LINEAR DISTRIBUTED MEMORIES*

■ EDUARDO MIZRAJI
Sección Biofísica,
Facultad de Humanidades y Ciencias
and Departmento de Biofísica, Facultad de Medicina,
Montevideo, Uruguay

In this article we present a method that allows conditioning of the response of a linear distributed memory to a variable context. This method requires a system of two neural networks. The first net constructs the Kronecker product between the vector input and the vector context, and the second net supports a linear associative memory.This system is easily adaptable for different goals. We analyse here its capacity for the conditional extraction of features from a complex perceptual input, its capacity to perform quasi-logical operations (for instance, of the kind of "exclusive-or"), and its capacity to structurate a memory for temporal sequences which access is conditioned by the context. Finally, we evaluate the potential importance of the capacity to establish arbitrary contexts, for the evolution of biological cognitive systems.

1. Introduction. The establishment of the physiological basis of complex cognitive processes is a precise frontier in contemporary research. Some promising paths are being opened in this field. Surely, these complex cognitive processes are a sequel of the high dimensionality of some biological neural networks. The linear models of distributed memories (Anderson, 1970; 1972; Cooper, 1973; Kohonen, 1977) are powerful instruments to explore the consequences of the high dimensionality. Furthermore, they allow a natural contact with biochemistry and physiology, by focusing on synaptic connectivity. Even some of the most simple linear models display "near-psychobiological" abilities like tolerance to errors, construction of concepts and capacity of losing units without destroying the function (Anderson, 1972; 1983; Cooper, 1973).

The central nervous system of the humans is the most versatile adaptive control system we know. In fact, this extraordinary versatility depends strongly on the capacity to recognize contexts. In a human being, the moving image of a tiger provokes very different responses whether it is seen on the screen of a TV set or in the middle of the jungle. The finding of context-dependent responses is a basic goal in some processes of learning. For instance, during an exercise of language, given an adjective, the teacher imposes a context asking for

* Part of this study has been presented in a preliminary version at the XVI Reunión Científica de la Sociedad Argentina de Biofísica, Tigre, Argentina, December 1987.

antonyms or synonyms, and the brain selects only one of two opposite responses.

The purpose of this communication is to describe a procedure that, within the framework of the linear formalism, allows for the existence of context dependent responses. In this formalism, inputs, contexts and outputs are vectors. The procedure that I present here requires a first neural net capable of constructing, using the input and the context vectors, the Kronecker product operation. The output of this net converges to a second neural net that sustains a linear associative memory. This system of nets operates in an extremely simple way, and it is adaptable to perform a variety of tasks. We will show its capacity to selectively extract features from a complex perception, its capacity to perform some quasi-logical operations, and its ability to construct "mnemonical gates" able to trigger the activation of associative paths.

2. Linear Distributed Memories: An Overview. The basic processing device in linear theories (Anderson, 1970; 1972; 1983; Cooper, 1973) is a unit α which response $\mathbf{g}_\alpha$ in the time $T+1$ is a linear combination of its n inputs $\mathbf{f}_\beta$ ($\beta = 1, \ldots, n$) in the time T:

$$\mathbf{g}_\alpha = \sum_\beta m_{\alpha\beta} \mathbf{f}_\beta, \tag{1}$$

where the coefficients $m_{\alpha\beta}$ represent the properties of the connections.

A neural net is built up of a set of N units, and their collective behaviour can be described by the equation:

$$\mathbf{g} = \mathbf{M}\mathbf{f}, \tag{2}$$

where the column vectors $\mathbf{f}$ and $\mathbf{g}$ represent the network input and outputs, respectively, and the matrix $\mathbf{M} = [m_{\alpha\beta}]$ represents the pattern of connectivity. The vectors $\mathbf{f}$ and $\mathbf{g}$ do not necessarily have the same dimension.

Under certain conditions, biological neural networks can be described by equation (2) (for a discussion, see Nass and Cooper, 1975). In these cases, the elements of vectors $\mathbf{f} = [\mathrm{f}_\beta]$ and $\mathbf{g} = [\mathrm{g}_\alpha]$ are deviations of the frequencies of action potentials with respect to an average basal value. The coefficients of the matrix $\mathbf{M}$ are directly linked to the properties of the synaptic connections.

Several important properties depend on the pattern of connectivity represented by $\mathbf{M}$, particularly the capacity to implement a distributed memory. Consider the problem of the associated pairs (Anderson, 1970; 1972; Kohonen, 1977): given the arbitrary pairs of vectors $(\mathbf{f}_i, \mathbf{g}_i)$, $i = 1, \ldots, K$, construct the matrix $\mathbf{M}$ such that $\mathbf{g}_i = \mathbf{M}\mathbf{f}_i$. In the case in which the K input vectors are orthonormal, the following correlation matrix is a solution:

$$\mathbf{M} = \sum_{i=1}^{K} \mathbf{g}_i \mathbf{f}_i^{\mathrm{T}}, \tag{3}$$

$\mathbf{f}^{\mathrm{T}}$ being the transpose of $\mathbf{f}$.

In order to simplify the arguments, we frequently assume in what follows that the vectors are normalized. Consequently, the scalar product between $\mathbf{f}_i$ and $\mathbf{f}_j$ is $\langle \mathbf{f}_i, \mathbf{f}_j \rangle = \mathbf{f}_i^{\mathrm{T}} \mathbf{f}_j = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Given an input $\mathbf{f}_k$, the network produces the following output:

$$\mathbf{M}\mathbf{f}_k = \sum_{i=1}^{K} \mathbf{g}_i \langle \mathbf{f}_i, \mathbf{f}_k \rangle = \mathbf{g}_k. \tag{4}$$

Hence, a correlation matrix $\mathbf{M}$ like equation (3) represents a distributed memory able to associate K arbitrary patterns $\mathbf{g}_i$ with K orthogonal vectors $\mathbf{f}_i$.

The capacity for pattern recognition displayed by a linear memory in the presence of random vectors $\mathbf{f}$ can be measured by means of a sort of signal/noise ratio that is of the order of N/K (Anderson, 1972). Therefore, high dimensionalities can generate an efficient filter.

3. Context-Dependent Recognition. Within the linear formalism, the context problem can be stated as follows: given two pairs of vectors $(\mathbf{f}, \mathbf{p}_1)$ and $(\mathbf{f}, \mathbf{p}_2)$, associate respectively the arbitrary vectors $\mathbf{g}_1$ and $\mathbf{g}_2$, with $\mathbf{f} \in \mathbf{V}m$, $\mathbf{p}_i \in \mathbf{V}q$, $\mathbf{g}_i \in \mathbf{V}n$, $\mathbf{V}r$ being a r-dimensional vector space.

Suppose the existence of a neural net apt to compose the input vector $\mathbf{f}$ and the context $\mathbf{p}$, generating a new vector $\mathbf{f} \times \mathbf{p}$, mq-dimensional, defined by:

$$\mathbf{f} \times \mathbf{p} \equiv [\mathbf{f}_\beta \mathbf{p}]. \tag{5}$$

The operation $\mathbf{f} \times \mathbf{p}$ is the Kronecker product (Bellman, 1960) defined for rectangular matrices and its properties are shown in the Appendix.

For orthonormal vectors $\mathbf{p}_1$ and $\mathbf{p}_2$, an exact and simple solution to the context problem is given by the matrix:

$$\mathbf{M} = \mathbf{g}_1 (\mathbf{f} \times \mathbf{p}_1)^{\mathrm{T}} + \mathbf{g}_2 (\mathbf{f} \times \mathbf{p}_2)^{\mathrm{T}}. \tag{6}$$

An input represented by the vector $\mathbf{f} \times \mathbf{p}_2$, is processed in the following way:

$$\mathbf{M}(\mathbf{f} \times \mathbf{p}_2) = \mathbf{g}_1 \langle \mathbf{f}, \mathbf{f} \rangle \langle \mathbf{p}_1, \mathbf{p}_2 \rangle + \mathbf{g}_2 \langle \mathbf{f}, \mathbf{f} \rangle \langle \mathbf{p}_2, \mathbf{p}_2 \rangle = \mathbf{g}_2. \tag{7}$$

The preceding example can describe a translation process: vector $\mathbf{f}$ could represent the pattern of neural activity associated to the English word "cat". Vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ could represent the patterns of neural activity on which the contexts "Spanish translation" and "French translation" respectively map, with $\mathbf{g}_1$ representing "gato" and $\mathbf{g}_2$, "chat" (Fig. 1).

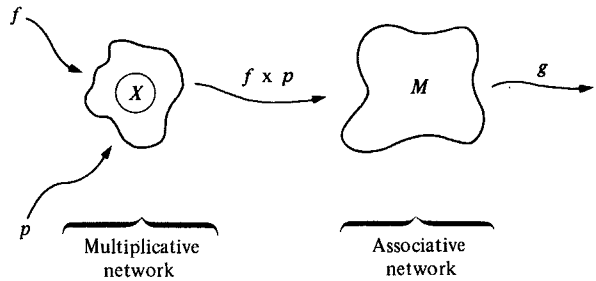

Figure 1. Context sensitive associative memory.

By means of this procedure, the existence of a multiplicative network capable of generating the composite vectors $\mathbf{f} \times \mathbf{p}$, allows construction of associative memories sensible to contexts with the structure:

$$\mathbf{M} = \sum_i \sum_j \mathbf{g}_{ij}(\mathbf{f}_i \times \mathbf{p}_{ij})^{\mathrm{T}} \tag{8}$$

In these memories, the pattern recognition capacities depend on the double filtering imposed by the two scalar products:

$$\mathbf{M}(\mathbf{f}_1 \times \mathbf{p}_{1k}) = \sum_i \sum_j \mathbf{g}_{ij} \langle \mathbf{f}_i, \mathbf{f}_1 \rangle \langle \mathbf{p}_{ij}, \mathbf{p}_{1k} \rangle. \tag{9}$$

4. Conditional Feature Selection. The system of nets presented in the last section, permits interesting insights concerning possible ways of analysing complex perceptual inputs. Imagine a perceptual input that maps on a high dimensional vector **f**. Let us assume that the previous experiences of the system have led it to develop a distributed memory structured around a "conceptual basis" $\{\mathbf{f}_r\}$. Each element belonging to a perceptual space can usually be expressed as a linear combination of vectors $\mathbf{f}_r$:

$$\mathbf{f} = \sum_r \gamma_r \mathbf{f}_r. \tag{10}$$

Therefore, given a conceptual basis $\{\mathbf{f}_r\}$ and an input **f**, one objective can be to construct a system capable of selecting some conceptual features $\mathbf{f}_r$ from **f**.

A fundamental solution to this issue was obtained by Anderson *et al.* (1977). They have assumed an analyser conformed by a neural network with feedback, such that the output in time T is the input in time $T+1$, having the structure:

$$\mathbf{M} = \sum_r \mu_r \mathbf{f}_r \mathbf{f}_r^{\mathrm{T}}. \tag{11}$$

Let $\{\mathbf{f}_r\}$ be an orthonormal basis. Then, if $\mathbf{f}_k \in \{\mathbf{f}_r\}$:

$$\mathbf{M}\mathbf{f}_k = \mu_k \mathbf{f}_k, \tag{12}$$

and the $\mathbf{f}_r$ are the eigenvectors associated to the eigenvalues μ_r.

A n-times iteration shows that:

$$\mathbf{M}^n \mathbf{f}_k = \mu_k^n \mathbf{f}_k. \tag{13}$$

Consequently, if $\mu_k < 1$, $\mathbf{f}_k$ tends to vanish. In this way, the eigenvalues of the matrix $\mathbf{M}$ determine which components from the input $\mathbf{f}$ are retained.

Imagine now that we are looking for a network able to select a pattern as a function of a variable context. Let $\mathbf{p}_{ki}$ represent the different contexts associable to vectors $\mathbf{f}_k$. Then, a conditional selector could be implemented by a multiplicative net that generates the vector $\mathbf{f}_k \times \mathbf{p}_{ki}$, and by an associative memory with the structure:

$$\mathbf{M} = \sum_r \mathbf{f}_r \left[\sum_i (\mathbf{f}_r \times \mathbf{p}_{ri})^{\mathrm{T}} \right], \tag{14}$$

where the outputs are reinjected on the multiplicative net. Notice that in equation (14) different contexts $\mathbf{p}_{ri}$ are associated with the same pattern $\mathbf{f}_r$.

As a simple illustration imagine that there are two orthogonal basic patterns $\mathbf{f}_1$ and $\mathbf{f}_2$, and two contexts $\mathbf{p}_{10}$ and $\mathbf{p}_{20}$, with:

$$\mathbf{M} = \mathbf{f}_1(\mathbf{f}_1 \times \mathbf{p}_{10})^{\mathrm{T}} + \mathbf{f}_2(\mathbf{f}_2 \times \mathbf{p}_{20})^{\mathrm{T}}. \tag{15}$$

Imagine a perception $\mathbf{f}$ that can be expressed as a linear combination of $\mathbf{f}_1$ and $\mathbf{f}_2$:

$$\mathbf{f} = \gamma_1 \mathbf{f}_1 + \gamma_2 \mathbf{f}_2, \tag{16}$$

γ_1 and γ_2 being constants.

If the initial input to the memory is:

$$\mathbf{f}^{(1)} = \mathbf{f} \times \mathbf{p}_{20}, \tag{17}$$

then:

$$\mathbf{M}\mathbf{f}^{(1)} = \gamma_1 \langle \mathbf{p}_{10}, \mathbf{p}_{20} \rangle \mathbf{f}_1 + \gamma_2 \langle \mathbf{p}_{20}, \mathbf{p}_{20} \rangle \mathbf{f}_2. \tag{18}$$

Defining $\mathbf{f}^{(u)} = \mathbf{M}\mathbf{f}^{(u-1)} \times \mathbf{p}_{20}$, after n iterations we have:

$$\mathbf{M}\mathbf{f}^{(n)} = \gamma_1 \langle \mathbf{p}_{10}, \mathbf{p}_{20} \rangle^n \mathbf{f}_1 + \gamma_2 \langle \mathbf{p}_{20}, \mathbf{p}_{20} \rangle^n \mathbf{f}_2. \tag{19}$$

Consequently, if the context vectors are normalized, $|\langle \mathbf{p}_{10}, \mathbf{p}_{20} \rangle| < 1$, $\langle \mathbf{p}_{20}, \mathbf{p}_{20} \rangle = 1$, the process can converge quickly:

$$\mathbf{Mf}^{(n)} \sim \gamma_2 \mathbf{f}_2. \tag{20}$$

Notice that orthogonal contexts give this result in one step.

This device allows, for the same perceptual input, the existence of many different ways of enhancing features. In general, for a complex perceptual pattern **f** whose input remains constant during a certain time, to explore a region of a context space $\{\mathbf{p}_{ki}\}$ implies to explore the basic composition of the pattern. This kind of adaptive emphasis can permit the design of exploratory strategies that could be useful in the domain of problem solving.

5. *Quasi-Logical Behaviour.* Some quasi-logical performances can be easily executed employing the Kronecker product. Suppose that two opposite concepts map on two orthogonal vectors **a** and **b**. Assume that two opposite decisions (eg "yes" and "not") map on two orthogonal vectors **s** and **n**. We can imagine a system of nets with two entries that responds with a positive decision only when both inputs are opposite [(**a**, **b**) or (**b**, **a**)], and that responds with a negative decision when both inputs are identical [(**a**, **a**) or (**b**, **b**)]. In order to construct this system it is first necessary to have a multiplicative network that, starting from inputs **u** and **v** (with equal dimensions to those of **a** and **b**), produces a vector $\mathbf{u} \times \mathbf{v}$. Secondly, that vector $\mathbf{u} \times \mathbf{v}$ must be sent to an associative memory with the following structure:

$$\mathbf{X} = \mathbf{s}(\mathbf{a} \times \mathbf{b})^{\mathrm{T}} + \mathbf{s}(\mathbf{b} \times \mathbf{a})^{\mathrm{T}} + \mathbf{n}(\mathbf{a} \times \mathbf{a})^{\mathrm{T}} + \mathbf{n}(\mathbf{b} \times \mathbf{b})^{\mathrm{T}}. \tag{21}$$

Then:

$$\mathbf{X}(\mathbf{a} \times \mathbf{b}) = \mathbf{X}(\mathbf{b} \times \mathbf{a}) = \mathbf{s}, \tag{22}$$

$$\mathbf{X}(\mathbf{a} \times \mathbf{a}) = \mathbf{X}(\mathbf{b} \times \mathbf{b}) = \mathbf{n}. \tag{23}$$

This behaviour has some similarity with the behaviour of logical (binary) nets that display the operation exclusive-or (XOR problem, see Rumelhart *et al.*, 1986).

6. *Gated Associative Pathways.* An interesting property of distributed memories is their capacity to store temporal sequences. Imagine an arbitrary sequence of orthonormal vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. A distributed memory **A** "learns" this arbitrary sequence if, given $\mathbf{a}_j$, the memory evokes its successor $\mathbf{a}_{j+1}$:

$$\mathbf{a}_{j+1} = \mathbf{A}\mathbf{a}_j. \tag{24}$$

It is easy to see that this happens when A is of the form:

$$\mathbf{A}=\mathbf{a}_1\mathbf{a}_0^{\mathrm{T}}+\mathbf{a}_2\mathbf{a}_1^{\mathrm{T}}+\cdots+\mathbf{a}_n\mathbf{a}_{n-1}^{\mathrm{T}}. \tag{25}$$

Provided that a feedback loop that reinjects the output into the same system (a reverberatory net) exists, the input $\mathbf{a}_0$ generates the sequence $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. In this highly simplified situation, the strongest imposition is orthogonality. In all other senses, the sequence is arbitrary. Remark that this kind of memory stores a temporal sequence superimposing the successive patterns in the same space (the matrix hardware). We will call a net with the structure of equation (25) a time net, or a time matrix (see also Kohonen, 1977, pp. 6–10).

The activation of a time matrix can be determined by a context vector. As usual, we need two networks: a multiplicative net that associates the pattern $\mathbf{a}_i$ to the context $\mathbf{p}$, generating $\mathbf{a}_i \times \mathbf{p}$, and an associative memory with matrix:

$$\mathbf{Q}=\sum_{j=1}^{M} \mathbf{a}_j(\mathbf{a}_{j-1}\times\mathbf{p})^{\mathrm{T}}. \tag{26}$$

From the properties of the Kronecker product, it follows that:

$$\mathbf{Q}=\mathbf{A}\times\mathbf{p}^{\mathrm{T}}, \tag{27}$$

where $\mathbf{A}$ is the time matrix [equation (25)].

Many temporal sequences can be stored in a distributed memory with the structure:

$$\mathbf{M}=\mathbf{A}_1\times\mathbf{p}_1^{\mathrm{T}}+\mathbf{A}_2\times\mathbf{p}_2^{\mathrm{T}}+\mathbf{A}_3\times\mathbf{p}_3^{\mathrm{T}}+\cdots, \tag{28}$$

where $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots,$ are time matrices. In these systems, a particular context $\mathbf{p}_i$ is a kind of gate able to control the access to an associative pathway represented by matrix $\mathbf{A}_i$.

A complex temporal memory like equation (28) could be a basis for the existence of some interesting complex behaviours. Suppose that to each vector $\mathbf{a}_i$ it corresponds another pattern of neural activity $\mathbf{b}_i$, being associated by a linear net described by matrix $\mathbf{B}$:

$$\mathbf{b}_i=\mathbf{B}\mathbf{a}_i. \tag{29}$$

If the outputs from $\mathbf{M}$ project on $\mathbf{B}$, then a context $\mathbf{p}$ and an initial state $\mathbf{a}$ generate two sequences: $(\mathbf{a}_{t1}, \ldots, \mathbf{a}_{tn})$ and $(\mathbf{b}_{t1}, \ldots, \mathbf{b}_{tn})$.

The sequence $(\mathbf{b}_{t1}, \ldots, \mathbf{b}_{tn})$ could, for instance, be associated with some kind of motor behaviour (Fig. 2).

In this way, many contexts $\mathbf{p}$, under different conditions $\mathbf{a}$, could generate a variety of complex motor behaviours. In spite of the great potential variability of the final result, these kinds of complex behaviours are structured using elemental actions, governed by the pattern associator $\mathbf{B}$.

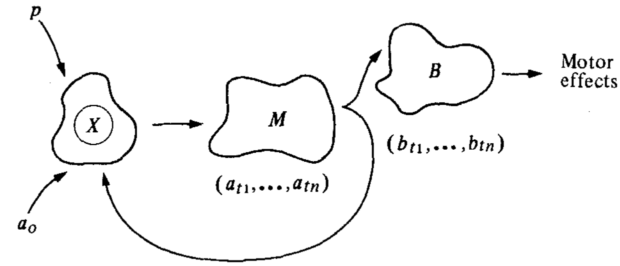

Figure 2. Context dependent time net with a motor effector.

7. Anatomy of the Multiplicative Nets. The nucleus of the method presented here is the existence of a neural network able to construct the Kronecker product. In Fig. 3, a very simple multiplicative net is shown.

The biological implementation of these nets needs a precise anatomical design, and synapses apt to multiply frequencies.

The building up of a multiplicative network could also be achieved via the interposition of many interneurons which axons ramify a few times. In this way, an extensive parallel iteration of an input frequency could be obtained without an enormous ramification of individual axons.

8. Some Perspectives. The Kronecker product provides a simple and efficient procedure to compose inputs and contexts. Nevertheless, it is stringent with regard to its anatomical and physiological implementations.

Real neural nets have, in fact, precise and complex patterns of connectivity. A multiplicative network could be the result of processes of "synaptic selection" arising from an initially redundantly connected net (following, for instance, the mechanism proposed by Changeux and Danchin, 1976; see also Ribchester, 1986). Another question concerns the existence of multiplicative synapses for inputs belonging to the physiological range of frequencies. As far as I know, the neuroanatomical and neurophysiological knowledge about these matters is still lacking.

The main strength of the multiplicative method is the following. Once the multiplicative net established, it sends a composed pattern $\mathbf{f} \times \mathbf{p}$ towards a distributed memory. The procedures for the self-organization of this memory could be of the kind of the powerful learning algorithms currently investigated, e.g. the δ-rule (Kohonen, 1977; Stone, 1986) or the generalized δ-rule (Rumelhart *et al.*, 1986).

On the other hand, the associations between inputs, outputs and contexts are independent and arbitrary, except for the imposition of orthogonalities. That is, given an input vector, an arbitrary set of responses $\{\mathbf{g}_i\}$, and an orthonormal set of contexts $\{\mathbf{p}_k\}$, the associations between the $\mathbf{g}_i$'s and the $\mathbf{p}_k$'s are entirely arbitrary (or "gratuitous" using Monod's terminology (Monod,

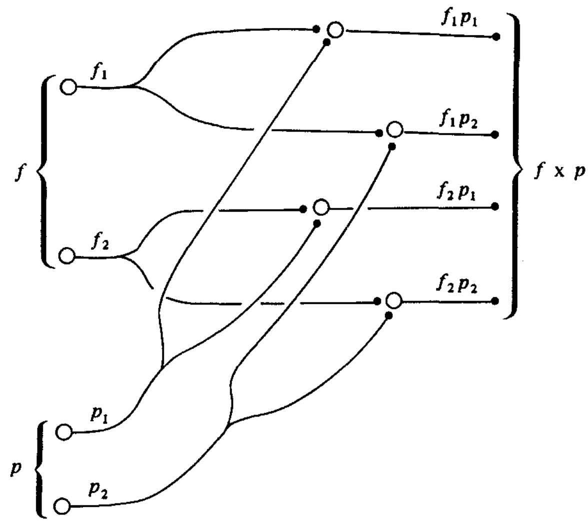

Figure 3. Connectivity of a multiplicative net.

1967)). In fact, the learning algorithms, like the δ-rule, allow to relax the exigencies of orthogonality (Kohonen, 1977; Stone, 1986).

Finally, let me make a comment about the evolution of cognitive systems. Perhaps, pattern recognition in biological (or prebiological) systems began when certain macromolecules (ancestors of the present enzymes and receptors) recognized other molecules in a more or less specific way, binding them. Perhaps, the context-dependent recognition also began at the biochemical level of primitive organisms with the emergence of allosteric proteins. The activity of these allosteric proteins (catalysis or signal transduction) can be delicately tuned up by a variety of molecules, the allosteric ligands (Monod *et al.*, 1963). These allosteric ligands represent chemically "gratuitous" molecular contexts (Monod, 1967), that is, there is no obligatory chemical similarity between the usual input to the allosteric protein (a substrate if the protein is an enzyme, or a chemical signal if it is a receptor), and an allosteric ligand. The existence of molecular control systems at the cellular level depends strongly on the allosteric proteins, that submit the complex metabolic behaviours to some strategic chemical contexts. Plausibly, the appearance of allosteric proteins was a critical point of departure in the evolution of living organisms in our planet.

On the other extreme of the complexity scale, the emergence of neural systems able to process high dimensional patterns, opened the possibility to develop very performant cognitive systems. The appearance of neural networks able to subject their behaviour to "gratuitous" contexts could have been a promoting factor for the construction of a cognitive system with a powerful capacity of adaptation to changing environments. With this device, old perceptions and new contexts, constitute a totally new experience, and can be stored without destroying the memory of previous experiences. The result can be a dramatic extension of the variety (see Ashby, 1958) of the neural regulator.

Hence, neural nets apt to modulate associations by means of gratuitous contexts could have been points of departure for the complex, astonishing and contradictory kind of cultural evolution travelled by the human species.

I would like to thank Dr. Julio Hernández and Dr. Ricardo Ehrlich for encouragement and helpful discussions and Verónica Etchart for technical help in the preparation of the manuscript.

APPENDIX

Kronecker Product (Bellman, 1960). Given two rectangular matrices $\mathbf{A}=[a_{ij}]_{mn}$ and $\mathbf{B}=[b_{ij}]_{pq}$, the Kronecker product (or direct product) is defined by:

$$\mathbf{A}\times\mathbf{B}\equiv[a_{ij}\mathbf{B}],$$

and has the following properties:

(a) $$\mathbf{A}\times\mathbf{B}\times\mathbf{C}=(\mathbf{A}\times\mathbf{B})\times\mathbf{C}=\mathbf{A}\times(\mathbf{B}\times\mathbf{C}),$$

(b) $$(\mathbf{A}+\mathbf{B})\times(\mathbf{C}+\mathbf{D})=\mathbf{A}\times\mathbf{C}+\mathbf{A}\times\mathbf{D}+\mathbf{B}\times\mathbf{C}+\mathbf{B}\times\mathbf{D},$$

(c) $$(\mathbf{A}\times\mathbf{B})(\mathbf{C}\times\mathbf{D})=(\mathbf{AC})\times(\mathbf{BD}).$$

From property (c), it is easy to see that for column vectors **a** and **c** of dimension m, and **b** and **d** of dimension n, we have:

$$(\mathbf{a}\times\mathbf{b})^T(\mathbf{c}\times\mathbf{d})=(\mathbf{a}^T\times\mathbf{b}^T)(\mathbf{c}\times\mathbf{d})=(\mathbf{a}^T\mathbf{c})\times(\mathbf{b}^T\mathbf{d})=\langle\mathbf{a},\mathbf{c}\rangle\langle\mathbf{b},\mathbf{d}\rangle.$$

LITERATURE

Anderson, J. A. 1970. "Two Models for Memory Organization Using Interacting Traces." *Math. Biosci.* **8**, 137–160.

——— 1972. "A Simple Neural Network Generating an Interactive Memory." *Math. Biosci.* **14**, 197–220.

——— 1983. "Cognitive and Psychological Computation with Neural Models." *IEEE Trans. Systems. Man Cybernetics* **13**, 799–815.

———, J. W. Silverstein, S. A. Ritz and R. S. Jones. 1977. "Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model." *Psychol. Rev.* **84**, 413–451.

Ashby, W. R. 1958. "Requisite Variety and its Implications for the Control of Complex Systems." *Cybernetica* **1**, 83–99.

Bellman, R. 1960. *Introduction to Matrix Analysis*. New York: McGraw-Hill.

Changeux, J. P. and A. Danchin. 1976. "Selective Stabilization of Developing Synapses as a Mechanism for the Specification of Neuronal Networks." *Nature* **264**, 705–712.

Cooper, L. N. 1973. "A Possible Organization of Animal Memory and Learning." In Proceedings of the Nobel Symposium on Collective Properties of Physical Systems, B. Lundquist and S. Lundquist (Eds). New York: Academic Press.

Kohonen, T. 1977. *Associative Memory: A System Theoretical Approach*. New York: Springer.

Monod, J. 1967. Leçon inaugurale, Collège de France.

———, J. P. Changeux and F. Jacob. 1963. "Allosteric Proteins and Cellular Control Systems." *J. Molec. Biol.* **6**, 306–329.

Nass, M. M. and L. N. Cooper. 1975. "A Theory for the Development of Feature Detecting Cells in Visual Cortex." *Biol. Cybernetics* **19**, 1–18.

Ribchester, R. R. 1986. *Molecule, Nerve and Embryo*. Glasgow: Blackie.

Rumelhart, D. E., G. E. Hinton and J. L. McClelland. 1986. "A General Framework for Parallel Distributing Processing." In *Parallel Distributing Processing*, D. E. Rumelhart and J. L. McClelland (Eds). Massachusetts: MIT Press.

———, ——— and R. J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* **323**, 533–536.

Stone, G. O. 1986. "An Analysis of the Delta Rule and the Learning of Statistical Associations." In *Parallel Distributing Processing*, D. E. Rumelhart and J. L. McClelland (Eds). Massachusetts: MIT Press.

Received 17 April 1988