



HAL
open science

Prosopagnosia in high capacity neural networks storing uncorrelated classes

S. Franz, D.J. Amit, M. A. Virasoro

► **To cite this version:**

S. Franz, D.J. Amit, M. A. Virasoro. Prosopagnosia in high capacity neural networks storing uncorrelated classes. *Journal de Physique*, 1990, 51 (5), pp.387-408. 10.1051/jphys:01990005105038700 . jpa-00212375

HAL Id: jpa-00212375

<https://hal.science/jpa-00212375>

Submitted on 4 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification

Physics Abstracts

87.30 — 64.60 — 75.10

Prosopagnosia in high capacity neural networks storing uncorrelated classes

S. Franz ^(1, 3), D. J. Amit ^(2, *) and M. A. Virasoro ⁽³⁾

⁽¹⁾ Istituto di Psicologia del C.N.R., Viale K. Marx 14, Roma, Italy

⁽²⁾ I.N.F.N., Dipart. di Fisica, Università di Roma « La Sapienza », Roma, Italy

⁽³⁾ Dipartimento di Fisica dell' Università di Roma « La Sapienza », Piazzale Aldo Moro 4, Roma, Italy

(Reçu le 7 août 1989, accepté le 20 octobre 1989)

Résumé. — On met en évidence une matrice synaptique qui stocke efficacement les patterns organisés en catégories non corrélées dans les réseaux neuronaux à attracteurs et les perceptrons. La capacité de stockage limite augmente avec le recouvrement m d'un *pattern* avec sa catégorie ancestrale, et diverge lorsque m tend vers 1. La distribution de probabilité des paramètres de stabilité locaux est étudiée, et conduit à une analyse complète des performances d'un perceptron en fonction de sa matrice synaptique, ainsi qu'à une compréhension qualitative du comportement du réseau neuronal correspondant. L'analyse de l'attracteur du réseau est complétée à l'aide de la mécanique statistique. La motivation d'une telle construction est de rendre possible l'étude d'un modèle de prosopagnosie : le passage du rappel individuel à celui de catégories lors de lésions, c'est-à-dire d'une détérioration aléatoire des efficacités synaptiques. Les propriétés de rappel du modèle en fonction de la matrice synaptique proposée sont étudiées en détail. Enfin nous comparons notre matrice synaptique à une matrice générique dont tous les paramètres de stabilité sont positifs.

Abstract. — We display a synaptic matrix that can efficiently store, in attractor neural networks (ANN) and perceptrons, patterns organized in uncorrelated classes. We find a storage capacity limit increasing with m , the overlap of a pattern with its class ancestor, and diverging as $m \rightarrow 1$. The probability distribution of the local stability parameters is studied, leading to a complete analysis of the performance of a perceptron with this synaptic matrix, and to a qualitative understanding of the behavior of the corresponding ANN. The analysis of the retrieval attractor of the ANN is completed *via* statistical mechanics. The motivation for the construction of this matrix was to make possible a study of a model for prosopagnosia, i.e. the shift from individual to class recall, under lesion, i.e. a random deterioration of the synaptic efficacies. The retrieval properties of the model with the proposed synaptic matrix, affected by random synaptic dilution are studied in detail. Finally we compare our synaptic matrix with a generic matrix which has all positive stability parameters.

1. Introduction.

Following the discovery by Gardner [1] that there exist synaptic matrices which would store sparsely coded (magnetized) patterns in neural networks with capacity that exceeds that of

(*) On leave from the Racah Institute of Physics, Hebrew University, Jerusalem.

uncorrelated patterns, and diverges as the correlation between the patterns tends to one [2], several proposals for explicit realization of such matrices have been put forth. The aim has been to find a modified synaptic prescription of the original Hopfield model [3] for which the Gaussian noise, that affects the local stability of the stored patterns, decreases as the correlation between patterns increases [4, 5], while the signal is controlled by an appropriate threshold [6]. Such constructions filled the gap between the Gardner results and previous work on the storage of magnetized patterns [7-9] which predicted a storage capacity decreasing with the magnetization. In an independent development it has been shown [10], using Gardner's method, that the limit of capacity of a network storing patterns organized in a finite number of uncorrelated classes, is the same as in the case of a single class of sparsely coded (biased) patterns.

On the other hand, the attempt [11] to extend the high storage prescription [4-6] to multiple classes has turned out to be limited to classes of highly correlated ancestors. Here we generalize the work of reference [6] to find a synaptic prescription which can store uncorrelated classes of patterns. Each of the classes is represented by an N -bit ancestor or prototype. These ancestors are uncorrelated. The individuals in a given class are generated by a random branching process in which they are chosen randomly with a fixed level of correlation with the ancestor. This is carried out in section 2 where we reformulate the synaptic matrix of reference [6] so as to make it gauge invariant. The gauge invariant single class dynamics is easily extended to multiple uncorrelated classes, giving a Gaussian distribution of the local stability parameters identical to that of a single class [6]. Consequently, the capacity diverges as in the optimal storage of Gardner, when the difference between individuals and their corresponding prototypes tends to zero (limit of full correlation). It is shown that while the natural extension of the synaptic matrix is asymmetrical, a symmetrical version can also be constructed. Section 2 concludes showing that the proposed synaptic prescription applies to the storage of patterns in an ANN (attractor neural network), as well as to the implementation of an associative rule for a perceptron. In the latter case the network is expected to associate correctly a set of input patterns to a set of output patterns. Both the input and the output patterns are grouped in corresponding classes, i.e. if two inputs belong to the same class, the corresponding outputs belong to the same class. Inside the corresponding classes the association is unspecified. The task of the perceptron is to associate with all individuals in an input class the representative of one output class.

In section 3 and 4 we analyze the probability distribution of the local stability parameters both for a perceptron and for an ANN. For perceptrons, the probability distribution of the local stability parameters fully characterizes the properties of the output state when a given pattern is presented as input [12].

It is interesting to analyze the behavior of networks with this synaptic prescription when they undergo deterioration, for instance by random destruction of a fixed fraction of the synapses. For one individual pattern belonging to a particular class it is useful to distinguish those sites that are common with the class prototype (which we call plus sites) and those in which the individual pattern differs from the class (minus sites). It was pointed out [14] that in Gardner's measure almost every network that stores correlated patterns will spontaneously show a level of robustness which is higher in the plus sites than in the minus sites. This is a probabilistic statement and as such it depends on the *a priori* probability distribution assumed. But a particular synaptic matrix, the pseudo-inverse for example, may lead to a network which is highly untypical with respect to the probability measure. In section 5 we show that when the number of plus sites incorrectly retrieved is of the same order of that of the minus sites, then following a random lesion the probability of making an error at a minus site is greater than the corresponding probability for a plus site.

For ANN's this probability distribution determines the state of the network after one step of parallel iteration, and gives a rough estimate on the attractors of the dynamics [12, 13]. Other methods must be employed to produce a full description of the attractors. In the case of symmetrical synapses, statistical mechanics can be used. This is done in section 6 where the analysis confirms and clarifies the picture that emerges from the analysis of the stability parameters.

In section 7, we study the effect of symmetric dilution of the synaptic matrix on the attractors of an ANN. Just like in the perceptron, or for the first step parallel dynamics of an ANN, the lesion affects the attractor asymmetrically at the plus and the minus sites. For low levels of dilution the individual patterns can be still retrieved, although, the probability of making an error at a plus site increases more rapidly than the corresponding probability at a minus site. As the dilution level increases, the attractors of individuals are completely destabilized, while those of class prototypes remain fixed points of the dynamics.

Finally in section 8 we compare our synaptic matrix to a generic one in the Gardner volume. For that matrix the stability parameters at every site are positive when the network is in a stored pattern, i.e. there are no retrieval errors in absence of fast noise. We find that this overlap does not depend on the parameter which is introduced to reduce the slow noise. We show that in the limit of full correlation our synaptic matrix is the axis of the cone in which the Gardner volume is embedded (1).

2. The model.

2.1 REFORMULATION OF HIGH STORAGE IN A SINGLE CLASS. — In a model storing a single class $p = \alpha N$ magnetized patterns $\{\sigma_i^\mu\}$, the patterns are chosen with the probability law

$$P(\sigma_i^\mu = \pm 1) = \frac{1 \pm m}{2} \quad i = 1, \dots, N; \quad \mu = 1, \dots, p \quad (2.1)$$

m is the magnetization, or access activity, of each pattern in the limit $N \rightarrow \infty$. The model is formulated in terms of a Hamiltonian

$$H\{S_i\} = -\frac{1}{2} \sum_{i \neq j} J_{ij} (S_i - c)(S_j - c) - \frac{U}{\sqrt{\alpha}} \sum_i S_i \quad (2.2)$$

with $S_i = \pm 1$ and

$$J_{ij} = \frac{1}{N(1 - m^2)\sqrt{\alpha}} \sum_{\mu=1}^p (\sigma_i^\mu - m)(\sigma_j^\mu - m). \quad (2.3)$$

The parameters U and c , are then varied to maximize the storage capacity which is an increasing function of m , and in the limit $m \rightarrow 1$, the capacity diverges as the theoretical limit found by Gardner [1].

As usual, the trend of the result can be detected from a signal to noise calculation for the stability of the stored patterns. Details are filled in by a statistical mechanics computation, which is possible due to the symmetry of the J_{ij} 's. To generalize the model we focus our attention on the signal and the noise parts of the stability parameters at each site of the stored patterns

$$\Delta_i^\mu \equiv \sigma_i^\mu h_i(\{\sigma_j^\mu\}) = \sum_{j \neq i} \sigma_i^\mu J_{ij} (\sigma_j^\mu - c) + \frac{\sigma_i^\mu U}{\sqrt{\alpha}} \quad (2.4)$$

(1) When this work was completed we received a brief report of Ioffe, Kuhn and Van Hemmen, proposing a similar synaptic structure.

where

$$h_i(\{\sigma_j^\mu\}) = \sum_{j \neq i} J_{ij}(\sigma_j^\mu - c) + \frac{U}{\sqrt{\alpha}} \quad (2.5)$$

is the local field.

The first step in the generalization of the model to uncorrelated classes is to store a class of patterns whose ancestor is different from the state of all 1's i.e. to store patterns $\bar{\sigma}_i^\mu = \sigma_i^\mu \sigma_i$ with an arbitrary ancestor $\sigma_i = \pm 1$. Note that due to the presence of c and U in (2.5) the stability parameters Δ_i^μ in formula (2.4) are not invariant under the gauge transformation (Mattis transformation) [15]

$$\begin{aligned} \sigma_j^\mu &\rightarrow \sigma_j^\mu \sigma_j \\ J_{ij} &\rightarrow J_{ij} \sigma_j \sigma_i \end{aligned} \quad (2.6)$$

where $\sigma_j = \pm 1$ are arbitrary. Hence it is not possible to pass from the storage of the ferromagnetic patterns $\{\sigma_j^\mu\}$ to the storage of a single class of patterns $\{\bar{\sigma}_j^\mu\} = \{\sigma_j^\mu \sigma_j\}$ by gauge transforming the synaptic matrix. This is also the main obstacle to the storage of uncorrelated classes.

The model is therefore reformulated to ensure that Δ_i^μ and its probability distribution be gauge invariant. We can rewrite (2.4) in the form

$$\Delta_i^\mu = \sigma_i^\mu \sum_{j \neq i} J'_{ij} \sigma_j^\mu \quad (2.7)$$

if we take

$$J'_{ij} = J_{ij} + \frac{U}{\sqrt{\alpha} mN} - \frac{c}{mN} \sum_k J_{ik} \quad (2.8)$$

In other words, U and c are absorbed in the definition of J'_{ij} : from PSPs or thresholds they become synaptic efficacies. The non-local term that has appeared in (2.8) could not be avoided when storing multiple classes. It will play an essential role in reducing the slow noise due to the storage of an extensive number of patterns, taking advantage of the correlations between the patterns with a class. The distribution of the Δ_i^μ is the same as in the original model, as can be seen by substituting (2.8) in (2.7) and using the fact that $\frac{1}{N} \sum_j \sigma_j^\mu = m$. If in

(2.6) one substitutes J'_{ij} for J_{ij} the stability parameters are now gauge invariant. We can, therefore, pass from one class to another by a mere gauge transformation [15].

Note that while the distribution of the stability parameters in the model of reference [6] and that of the reformulated model are the same, the dynamical properties of the two models can be different. This can be seen observing that the original model is formulated in terms of a Hamiltonian, which disappears in the reformulation because $J'_{ij} \neq J_{ji}$.

2.2 HIGH STORAGE IN MULTIPLE UNCORRELATED CLASSES. — Next we proceed to formulate a synaptic matrix that stores $p = \alpha N$ patterns $\{\xi_j^{\mu\nu}\}$ organized in classes. There are p_c classes, defined by the uncorrelated prototype patterns or ancestors $\{\xi_j^\mu\}$, chosen with probability

$$P(\xi_j^\mu = \pm 1) = 1/2 ; j = 1, \dots, N ; \mu = 1, \dots, p_c \quad (2.9)$$

A pattern in class μ is defined in relation to its ancestor $\{\xi_j^\mu\}$ by a stochastic branching process [8] :

$$P(\xi_i^{\mu\nu} = \pm \xi_i^\mu) = \frac{1 \pm m}{2}; \quad j = 1, \dots, N; \quad \mu = 1, \dots, p_c; \quad \nu = 1, \dots, p/p_c. \quad (2.10)$$

The number of classes $p_c = (\alpha N)^\gamma$ with $0 \leq \gamma < 1$ and the number of patterns per class $(\alpha N)^{1-\gamma}$. To store the $\{\xi_i^{\mu\nu}\}$'s we use

$$J_{ij} = \frac{U}{\sqrt{\alpha} N} \sum_{\mu} \xi_i^\mu \xi_j^\mu + \frac{1}{\sqrt{\alpha} (1 - m^2) N} \sum_{\mu\nu} (\xi_i^{\mu\nu} - m\xi_i^\mu)(\xi_j^{\mu\nu} - m\xi_j^\mu) - \frac{c}{\sqrt{\alpha} m(1 - m^2) N^2} \sum_{\mu\nu\lambda k} \xi_j^\lambda \xi_k^\lambda (\xi_i^{\mu\nu} - m\xi_i^\mu)(\xi_k^{\mu\nu} - m\xi_k^\mu). \quad (2.11)$$

In the sums, the indices μ and λ , each takes the value, $1, \dots, p_c$, ν takes the values $1, \dots, p/p_c$, and k , the values $1, \dots, N$.

The first two terms on the r.h.s. of (2.11) are just those that appear in reference [8, 9]. To understand the effect of the third term, let us note that if the number of classes $p_c \ll N$, the matrix $P_{ij} = 1/N \sum_{\lambda} \xi_i^\lambda \xi_j^\lambda$ acts as the projector in the space spanned by the $\{\xi_j^\lambda\}$. The third term in (2.11) can, therefore, be written as

$$- \frac{c}{m} \frac{1}{\sqrt{\alpha} (1 - m^2) N^2} \sum_{\mu\nu} (\xi_i^{\mu\nu} - m\xi_i^\mu) \left[\sum_k P_{jk} (\xi_k^{\mu\nu} - m\xi_k^\mu) \right].$$

Which subtracts from the J_{ij} that projection of the vectors $\{(\xi_j^{\mu\nu} - m\xi_j^\mu)\}$ in the space spanned by the class prototypes. As such, as we will see, it is very effective in subtracting from the stability parameters of the individual patterns that part of noise that comes from this projection.

Admittedly, this term gives rise to a non-locality of the synaptic matrix. Yet this may not be such a serious drawback, since it may be possible to generate it by a local iterative procedure.

It should be pointed out that a local algorithm, such as the «perceptron learning algorithm», gives rise to matrices which can not be expressed as local expression in terms of the patterns.

The matrix (2.11) is not symmetric. It is, however, possible to symmetrize it without changing in a relevant way the distribution of the stability parameters. A symmetrical version of (2.11) is

$$J_{ij} = \frac{U}{\sqrt{\alpha} N} \sum_{\mu} \xi_i^\mu \xi_j^\mu + \frac{1}{\sqrt{\alpha} (1 - m^2) N} \sum_{\mu\nu} (\xi_i^{\mu\nu} - m\xi_i^\mu)(\xi_j^{\mu\nu} - m\xi_j^\mu) - \frac{c}{\sqrt{\alpha} m(1 - m^2) N^2} \sum_{\mu\nu\lambda\kappa} \xi_k^\lambda (\xi_k^{\mu\nu} - m\xi_k^\mu) [\xi_j^\lambda (\xi_i^{\mu\nu} - m\xi_i^\mu) + \xi_i^\lambda (\xi_j^{\mu\nu} - m\xi_j^\mu)]. \quad (2.12)$$

Now the model is again Hamiltonian, with $H = - \sum_{i < j} J_{ij} S_i S_j$ and it can be investigated by statical mechanics.

2.3 THE HETERO-ASSOCIATION PERCEPTRON VIEW. — Before proceeding to show that this synaptic prescription actually stores the extensive set of $\{\xi_j^{\mu\nu}\}$, let us point out that (2.11) can be considered as a weight arrangement for an auto-association rule in a perceptron with N

input units and N output units. It is natural to extend the prescription (2.11) to the case in which the output patterns are not necessarily the same as the input patterns. We associate a set of input patterns $\{\sigma_j^{\mu\nu}\}$ $j = 1, \dots, N$, organized in uncorrelated classes, to a set of output patterns $\{\xi_i^{\mu\nu}\}$ $i = 1, \dots, N'$ belonging to corresponding classes. N and N' are, respectively, the numbers of input and output units.

The generation of input and output patterns is done by independent branching processes similar to (2.10), from input and output ancestors $\{\sigma_j^\mu\}$ and $\{\xi_i^\mu\}$, respectively. For a perceptron, in contrast to an ANN, the overlaps of the input patterns with their input prototypes, m_{in} , and the corresponding overlaps of the output patterns with the output prototypes, m_{out} , are independent variables [10].

The association is performed through the usual noiseless response function

$$S_i^{\text{output}} = \text{sgn} \left(\sum_{j=1}^N J_{ij} S_j^{\text{input}} \right); \quad i = 1, \dots, N'. \quad (2.13)$$

In the case of auto-association $N = N'$, $m_{\text{in}} = m_{\text{out}}$, $\sigma_j^{\mu\nu} = \xi_j^{\mu\nu}$ for all i, μ, ν . Equation (2.13) can be viewed as the first step of the parallel dynamics of an ANN.

The extension of (2.11) to the hetero-association perceptron can be written as

$$\begin{aligned} J_{ij} = & \frac{Um_{\text{out}}}{\sqrt{\alpha} m_{\text{in}}} \sqrt{\frac{1 - m_{\text{in}}^2}{1 - m_{\text{out}}^2}} \frac{1}{N} \sum_{\mu} \xi_i^{\mu} \sigma_j^{\mu} + \\ & + \frac{1}{\sqrt{\alpha} (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)} \sum_{\mu\nu} (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^{\mu})(\sigma_j^{\mu\nu} - m_{\text{in}} \sigma_j^{\mu}) \\ & - \frac{c}{m_{\text{in}} \sqrt{\alpha} (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)} \sum_{\mu\nu\lambda k} \sigma_j^{\lambda} \sigma_k^{\lambda} (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^{\mu})(\sigma_k^{\mu\nu} - m_{\text{in}} \sigma_k^{\mu}). \quad (2.14) \end{aligned}$$

The overall factors in equations (2.11), (2.12) and (2.14) have been chosen so that

$$\sum_j J_{ij}^2 = 1. \quad (2.15)$$

This will turn out to be useful in section 8. The prescriptions (2.11), (2.12) and (2.14) can be generalized to a full ultrametric hierarchy of patterns. The presentation of this more involved case will be given elsewhere.

3. The distribution of the local stability parameters.

In this section we analyze separately the probability distribution of the stability parameters when the network is in a state which is an individual pattern or in a state which is an ancestor, with the synaptic matrix (2.14). The perceptron will have N input units and N' output units, and will store $p = \alpha N$ patterns equally divided in a number $p_c = (\alpha N)^\gamma$, of classes ($\gamma < 1$). The corresponding analysis for the first step parallel dynamics in an ANN is recovered by setting $m_{\text{in}} = m_{\text{out}} = m$ in all formulae. We do not analyze in detail the case of the symmetric matrix (2.12), which is very similar, except to indicate when formulae undergo essential modifications.

As we stressed in the introduction, it is useful to distinguish in a pattern $\{\xi_i^{\mu\nu}\}$ the sites in which $\xi = \xi_i^{\mu\nu} \xi_i^{\mu} = 1$ (plus sites) and the sites in which $\xi = \xi_i^{\mu\nu} \xi_i^{\mu} = -1$ (minus sites). The

probability distribution of ξ is (from (2.10))

$$P(\xi = \pm 1) = \frac{1 \pm m_{out}}{2} \tag{3.1}$$

i.e. the fraction of ξ sites is $(1 + \xi m_{out})/2$ of the total.

Let us consider the stability parameter Δ_ξ of one of the stored patterns (e.g. pattern $\{\xi_i^{11}\}$) at a ξ site.

Using (2.14) and $\frac{1}{N} \sum_j \sigma_j^1 \sigma_j^{11} = m_{in}$ we obtain

$$\begin{aligned} \Delta_\xi \equiv \xi_i^{11} \sum_j J_{ij} \sigma_j^{11} &= \sqrt{\frac{1 - m_{in}^2}{\alpha (1 - m_{out}^2)}} (1 + m_{out} \xi (U - 1)) + \\ &+ \frac{\xi_i^{11}}{N \sqrt{\alpha (1 - m_{in}^2)(1 - m_{out}^2)}} \sum_{(\mu, \nu) \neq (1, 1)} (\xi_i^{\mu\nu} - m_{out} \xi_i^\mu) (\sigma_j^{\mu\nu} - m_{in} \sigma_j^\mu) (\sigma_j^{11} - c \sigma_j^1) \\ &+ \frac{Um_{out}}{m_{in} N} \sqrt{\frac{1 - m_{in}^2}{\alpha (1 - m_{out}^2)}} \xi_i^{11} \sum_{j \neq 1} \xi_i^\mu \sigma_j^\mu \sigma_j^{11} - \frac{c \xi_i^{11}}{m_{in} \sqrt{\alpha (1 - m_{in}^2)(1 - m_{out}^2)}} \frac{1}{N^2} \\ &\times \sum_{\mu\nu, \lambda \neq 1, kj} \sigma_k^\lambda \sigma_j^\lambda \sigma_j^{11} (\xi_i^{\mu\nu} - m_{out} \xi_i^\mu) (\sigma_k^{\mu\nu} - m_{in} \sigma_k^\mu). \end{aligned} \tag{3.2}$$

The first term in (3.2) is the expected value of Δ_ξ . The second term is a Gaussian noise with zero mean and variance $(1 - 2cm_{in} + m_{in}^2)$ which has a minimum equal to $1 - m_{in}^2$ for $c = m_{in}$ [6]. The third and fourth terms in (3.2) sum terms in J_{ij} different from $\{\sigma_j^{11}, \xi_i^{11}\}$. They do not contribute at all. They have zero mean and do not fluctuate in the limit $N \rightarrow \infty$. This can be seen, for example, from the expectation

$$\begin{aligned} E \left(\frac{1}{N} \sum_{j, \mu \neq 1} \xi_i^{11} \xi_i^\mu \sigma_j^\mu \sigma_j^{11} \right)^2 &= \frac{1}{N^2} \sum_{jk, \mu \neq 1, \mu' \neq 1} \xi_i^\mu \xi_i^{\mu'} \sigma_j^\mu \sigma_k^{\mu'} \sigma_j^{11} \sigma_k^{11} = \\ &= \frac{1}{N^2} \sum_{jk, \mu \neq 1, \mu' \neq 1} \delta_{jk} \delta_{\mu\mu'} = \frac{\alpha^\gamma}{N^{1-\gamma}} \rightarrow 0 \end{aligned} \tag{3.3}$$

since $p_c = (\alpha N)^\gamma$.

In the case of the symmetrical prescription (2.12) for ANN (i.e. $m_{in} = m_{out} = m$, $\sigma_i^{\mu\nu} = \xi_i^{\mu\nu}$) an extra term appears in Δ_ξ , i.e.

$$- \frac{c \xi_i^{11}}{m(1 - m^2) \sqrt{\alpha}} \frac{1}{N^2} \sum_{\mu\nu, \lambda \neq 1, kj} \xi_k^\lambda (\xi_k^{\mu\nu} - m \xi_k^\mu) [\xi_i^\lambda (\xi_j^{\mu\nu} - m \xi_j^\mu) + \xi_j^\lambda (\xi_i^{\mu\nu} - m \xi_i^\mu)] \xi_j^{11}.$$

In the limit $N \rightarrow \infty$ this term becomes

$$- \sqrt{\alpha} m \xi$$

namely it is effectively a shift of U by a quantity $-\alpha$. Such a shift is irrelevant, since U is a parameter to be optimized. It should be stressed that even if the asymmetrical and the symmetrical prescriptions (2.11) and (2.12) lead to the same distribution of the stability parameters, the dynamical properties of the two models may be different.

The ratio between the expected value of Δ_ξ (signal) and its standard deviation (noise) is

$$\frac{1 + m_{\text{out}} \xi (U - 1)}{\sqrt{\alpha (1 - m_{\text{out}}^2)}}. \tag{3.4}$$

This shows that in the limit $m_{\text{out}} \rightarrow 1$ α can diverge roughly as $1/(1 - m_{\text{out}})$ and that unless $U = 1$ the plus and the minus sites have different degrees of stability. We will see in the next section, studying the tail of the distribution of the Δ_ξ [15], that the maximization of the storage capacity will be obtained for $U > 1$. This implies that the plus sites will be more stable than the minus sites. Further insight can be gained by studying the tail of the probability distribution [14], as we will see in the next section.

The stability parameter at a site i of a prototype (ancestor) pattern is given by

$$\begin{aligned} \Delta_c \equiv \xi_i^1 \sum_j J_{ij} \sigma_j^1 &= \frac{Um_{\text{out}}}{m_{\text{in}}} \sqrt{\frac{1 - m_{\text{in}}^2}{\alpha (1 - m_{\text{out}}^2)}} + \\ &+ \frac{\xi_i^1 (1 - c/m_{\text{in}})}{N \sqrt{\alpha (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)}} \sum_{j, \mu, \nu} (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^\mu) (\sigma_j^{\mu\nu} - m_{\text{in}} \xi_j^\mu) \sigma_j^1 \\ &+ \frac{\xi_i^1 Um_{\text{out}}}{Nm_{\text{in}}} \sqrt{\frac{1 - m_{\text{in}}^2}{\alpha (1 - m_{\text{out}}^2)}} \sum_{j, \mu \neq 1} \xi_i^\mu \sigma_j^\mu \sigma_j^1 \\ &- \frac{c\xi}{N^2 m_{\text{in}} \sqrt{\alpha (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)}} \sum_{\mu, \nu, k, j, \alpha \neq 1} (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^\mu) (\sigma_k^{\mu\nu} - m_{\text{in}} \sigma_k^\mu) \sigma_k^\alpha \sigma_j^\alpha \sigma_j^1. \end{aligned} \tag{3.5}$$

As in the case of Δ_ξ the third and the fourth terms in Δ_c have zero mean and do not fluctuate in the limit $N \rightarrow \infty$. The second term is a Gaussian variable of zero mean, which vanishes if $c = m_{\text{in}}$.

The choice $c = m_{\text{in}}$ is optimal because it minimizes the noise both in Δ_ξ and in Δ_c [6, 11]. In particular, Δ_c does not fluctuate at all and is

$$\Delta_c = \frac{Um_{\text{out}}}{m_{\text{in}}} \sqrt{\frac{1 - m_{\text{in}}^2}{\alpha (1 - m_{\text{out}}^2)}}. \tag{3.6}$$

In fact, Δ_c plays the role of the parameter M introduced by Gardner [1]. We will see below that optimizing U for $\alpha = \alpha_G$, where α_G is the Gardner bound on storage capacity [1], implies $\Delta_c \simeq M$ for $m_{\text{out}} \rightarrow 1$. With the synaptic matrix (2.14), the r.h.s. of (3.6) becomes

$$\frac{U}{\sqrt{\alpha}} - \sqrt{\alpha}, \tag{3.7}$$

where we have set $m_{\text{in}} = m_{\text{out}} = m$, as is required for an ANN. This corresponds to shifting U by $-\alpha$ and does not have any effect when U is optimized.

The above analysis does not hold in the limit cases $m_{\text{in}} = 0$, i.e. when all input pattern are independent and there are no classes. It also does not hold when $m_{\text{in}} = 1$, i.e. when all individuals are equal to their class prototype and there are no individuals.

4. Performance.

From the analysis of the above section, the probability distribution of Δ_ξ is :

$$P_\xi(\Delta_\xi) = \frac{1}{\sqrt{2\pi}(1-m_{in}^2)} \exp \left[- \frac{\left(\Delta_\xi - \sqrt{\frac{1-m_{in}^2}{\alpha(1-m_{out}^2)}} (1-m_{out}\xi(1-U)) \right)^2}{2(1-m_{in}^2)} \right], \quad (4.1)$$

where we have chosen $c = m_{in}$ as in references [6, 11].

To study the performance of the network we introduce the probability E_ξ of an error in a ξ site, i.e. a site of descendent with relative sign ξ to the same site in its ancestor. From (4.1)

$$E_\xi \equiv P(\Delta_\xi | \xi) = H \left(\frac{1+m_{out}\xi(U-1)}{\sqrt{\alpha(1-m_{out}^2)}} \right) \quad (4.2)$$

where $H(x) = \int_x^\infty \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} = 1/2[1 - \text{erf}(x)]$. Note that E_ξ is independent of m_{in} (except for $m_{in} = 0$ and $m_{in} = 1$ where (4.2) does not hold). Consequently, the discussion in this section can be carried out for fixed (arbitrary) m_{in} . E_ξ represents the number of errors in the ξ sites divided by the number of ξ sites. The contribution of these sites to the total average error is

$$\frac{1 + \xi m_{out}}{2} E_\xi, \quad (4.3)$$

and the total average error E is given by

$$E = \frac{1 + m_{out}}{2} E_+ + \frac{1 - m_{out}}{2} E_- . \quad (4.4)$$

Clearly, as E decreases the network retrieves the stored information more effectively. The performance of the network is optimized if U is chosen to minimize E . Differentiating (4.2) respect to U one finds :

$$U_{opt} = 1 + \frac{\alpha(1-m_{out}^2)}{2m_{out}} \log \left(\frac{1+m_{out}}{1-m_{out}} \right) . \quad (4.5)$$

For this value of U the error in the plus sites and the error in the minus sites are of the same order of magnitude. In fact,

$$\frac{1 + m_{out}}{2} E_+ \approx \frac{1 - m_{out}}{2} E_- . \quad (4.6)$$

Since $U_{opt} > 1$, it can be seen from equation (4.2) that

$$E_- > E_+ , \quad (4.7)$$

which is a consequence of the fact that the expected value of the stability parameter is greater at the plus sites than at the minus sites, the noise at both types of sites being equal.

In figures 1 and 2 we plot E_+ , E_- and E vs. m_{out} for $\alpha = \beta \alpha_G$ with $\beta = 0.9$ and $\beta = 0.1$, respectively. E_+ and E are decreasing functions of m_{out} . E_- is a concave function,

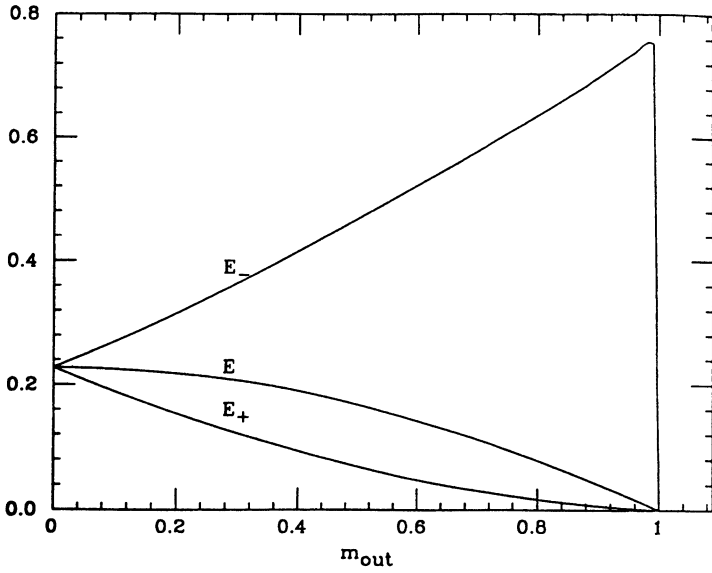


Fig. 1. — E_+ , E_- and E as functions of m_{out} for $\alpha = 0.9 \times \alpha_G$. E_+ and E are decreasing functions of m_{out} . E_- has a maximum which can lie very close to the value $m_{out} = 1$.

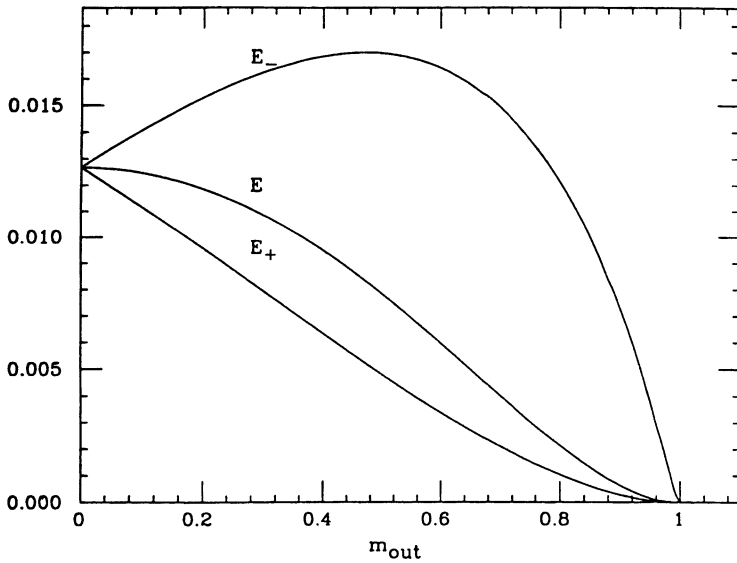


Fig. 2. — E_+ , E_- and E as functions of m_{out} for $\alpha = 0.1 \times \alpha_G$. The maximum of E_- is lower than in figure 1 and lies at a lower value of m_{out} . In this case, the range of m_{out} for which the model can retrieve correctly the individual pattern is wider.

equal to E for $m_{out} = 0$ and tend to zero for $m_{out} \rightarrow 1$. It has a maximum, that shifts to the right for increasing values of β . Note that for $\beta = 0.9$, the value of the maximum is very high, it lies very close to $m_{out} = 1$. Clearly, a good discrimination of an individual pattern from its class prototype is obtained if, in addition to $E \ll 1/2$, also $E_- \ll 1/2$. It is, therefore, necessary that the argument of the error function in E_- , be positive, which is satisfied if $U < 2$. When $m_{out} \rightarrow 1$, it is possible to determine the asymptotic behavior of E_ξ , equation (4.2), and of the storage capacity. In this limit, as all the output patterns tend to their output prototypes, the task for the network reduces to the classification of input patterns, and it is possible to reduce the error to zero, since the number of classes is not extensive.

To see this note that since E is given by (4.4), to have $E \rightarrow 0$ it is sufficient that $E_+ \rightarrow 0$, regardless to the value of E_- . E_+ will vanish when

$$\frac{1 + m_{out}(U - 1)}{\sqrt{\alpha(1 - m_{out}^2)}} \rightarrow \infty . \tag{4.8}$$

At each input pattern in a given class, has to be associated the prototype of the corresponding output class. This is an easy task [10], and can be performed for an arbitrary large number of patterns. But it is more relevant, to require that even in the limit $m_{out} \rightarrow 1$ the individuals be distinguished from the class prototypes ; i.e. to require that $E_- \rightarrow 0$. Then, using (4.2), we must have

$$\frac{1 - m_{out}(U - 1)}{\sqrt{\alpha(1 - m_{out}^2)}} \rightarrow \infty . \tag{4.9}$$

This is particularly relevant in an ANN, where $E_\xi \rightarrow 0$ guarantees that the p stored patterns ξ_j^{mv} be fixed points of the dynamics.

Using the value (4.5) for U , and writing α as

$$\alpha = \beta \frac{1}{(1 - m_{out}) \ln \left(\frac{1}{1 - m_{out}} \right)} \tag{4.10}$$

we find that condition (4.9) is fulfilled for any β in the interval $0 < \beta < 1$. Substituting (4.10) in (4.5) we find

$$U = 1 + \beta . \tag{4.11}$$

It is interesting to compare Δ_c , given by (3.6), with Gardner's parameter $M = \sum_j \xi_i^\mu J_{ij}^* \sigma_j^\mu$

where J_{ij}^* is a matrix in the Gardner volume. One obtains, from (4.11) and the saddle point equation in [1],

$$\frac{\Delta_c}{M} \simeq \frac{1 + \beta}{2} . \tag{4.12}$$

This shows that, in the limit $m_{out} \rightarrow 1$, in which α approaches the Gardner bound, the stability parameters of the class representatives approach the value taken by the optimal matrices that store the given patterns. In section 8 we will see that the distance between the matrix (2.14) and a generic matrix in the Gardner volume tends to zero in this limit.

The asymptotic form of E_ξ , equation (4.2) is given, as $m_{\text{out}} \rightarrow 1$, by

$$E_\xi = \frac{(1 - m_{\text{out}})^{\frac{(1 + \xi\beta)^2}{4\beta}}}{\frac{(1 + \xi\beta)}{\sqrt{\beta}} \sqrt{\pi \log \left(\frac{1}{1 - m_{\text{out}}} \right)}} \quad (4.13)$$

which vanishes with a β -dependent power as $m_{\text{out}} \rightarrow 1$, apart from logarithmic corrections. For high values of β (≈ 1), the decay of E_- to zero is very slow. Formula (4.13) applies in a very narrow range below $m_{\text{out}} = 1$, because the power in the exponent becomes very small. This can be seen in figure 1, for $\beta = 0.9$, where E_- has a maximum for $m_{\text{out}} = 0.98$, and only then decreases to zero.

When $\beta \geq 1$, equation (4.13) is still valid for E_+ , while $E_- = 1/2$ for $\beta = 1$ and $E_- = 1$ for $\beta > 1$. We see that for $\beta = 1$ ($\alpha = \alpha_G$ and $\Delta_c = M$) there is a transition, in which the network loses its ability to recall individuals, but can still recall classes.

5. Noisy patterns and lesions.

Next we study the performance of the network when the input is a pattern with random errors or when the synaptic matrix is lesioned by a random dilution of synapses. If the input pattern $\{\sigma_j\}$ has an overlap a with one of the stored pattern, e.g. $\{\sigma_j^{11}\}$ (i.e. $1/N \sum_j \sigma_j^{11} \sigma_j = a$), then the same arguments that led to formula (4.2) give

$$E_\xi = H \left(\delta \frac{1 + m_{\text{out}} \xi (U - 1)}{\sqrt{\alpha (1 - m_{\text{out}}^2)}} \right) \quad (5.1)$$

with

$$\delta = a \sqrt{\frac{1 - m_{\text{in}}^2}{1 - a^2 m_{\text{in}}^2}}. \quad (5.2)$$

The error E_ξ is defined with respect to the desired output ξ^{11} . The resulting overlap of the output with ξ^{11} is [12]

$$a' = 1 - 2E = \frac{1 + m_{\text{out}}}{2} \operatorname{erf} \left(\delta \frac{1 + m_{\text{out}}(U - 1)}{\sqrt{2\alpha(1 - m_{\text{out}}^2)}} \right) + \frac{1 - m_{\text{out}}}{2} \operatorname{erf} \left(\delta \frac{1 + m_{\text{out}}(U - 1)}{\sqrt{2\alpha(1 - m_{\text{out}}^2)}} \right). \quad (5.3)$$

The analogous equation was derived in references [12, 13] to determine the « basins of attraction » of the stored patterns. Note that for a in the interval $(0, 1)$, δ is a decreasing function of m_{in} . This means that for greater m_{in} the basins of attraction of the stored patterns are smaller.

Expressions (5.1)-(5.3) apply also to the case of random dilution of synapses, with a^2 replaced by p , the probability of survival of a synapse. In other words, if

$$J_{ij} \rightarrow J_{ij} c_{ij} \quad (5.4)$$

with

$$P(c_{ij}) = p\delta(c_{ij} - 1) + (1 - p)\delta(c_{ij}), \quad (5.5)$$

then (5.1) holds with

$$\delta = \sqrt{\frac{p(1 - m_{in}^2)}{1 - pm_{in}^2}}. \quad (5.6)$$

The presentation of a noisy pattern as well as the dilution of the synapses are, therefore, equivalent to an increase of α by a factor $1/\sqrt{\delta}$ (keeping U fixed).

The derivative of E_ξ with respect to δ at $\delta = 1$ gives the increment of the error due to an infinitesimal lesion

$$E'_\xi = \left. \frac{\partial E}{\partial \delta} \right|_{\delta=1} = - \frac{1 + \xi m_{out}(U - 1)}{\sqrt{2 \pi \alpha (1 - m_{out}^2)}} \exp\left(\frac{[1 + m_{out}(U - 1)]^2}{2 \alpha (1 - m_{out}^2)}\right). \quad (5.7)$$

As long as $U < 2$, where it should be to allow recall of individual patterns, the error increment in the minus sites is greater than that in the plus sites. As was mentioned in the Introduction, this difference in the increments occurs for almost all synaptic matrices in the volume which optimizes the storage of the patterns [14]. There, it is a consequence of the fact that the expected value of the stability parameters at the plus sites is greater than at the minus sites. In other words, more information is lost in those sites that code the difference of individuals with respect to their class (i.e. sites at which a pattern differs from its class representative prototype) than at those sites that confirm the class. This effect can be interpreted as prosopagnosia once a quantitative criterion is given to establish what is the maximum amount of error in the minus sites for which one can distinguish individuals, and what is the maximum amount of error in the plus sites for which one can have class recall. It is clear that a very small value of δ has a marginal effect on the performance of the network, while a large δ destroys all the informations stored (agnosia). Somewhere in between there is the region of prosopagnosia, in which the classes are recalled, but it is impossible to distinguish the individuals inside a class.

On the other hand, if the input is a pattern which has an overlap a with one of the ancestors and is otherwise uncorrelated with any of the individuals, then the error fraction of the output, with respect to the corresponding output prototype, is given by

$$E_c = H\left(\frac{a}{\sqrt{1 - a^2}} \Delta_c\right) = H\left(\frac{a}{\sqrt{1 - a^2}} \frac{\sqrt{1 - m_{in}^2}}{m_{in}} \frac{Um_{out}}{\sqrt{\alpha(1 - m_{out}^2)}}\right). \quad (5.8)$$

Consequently, for any set of parameters (α , m_{in} , m_{out} , a) we can make E_c arbitrarily small by taking

$$U \gg \frac{\sqrt{1 - a^2}}{a} \frac{m_{in}}{\sqrt{1 - m_{in}^2}} \frac{\sqrt{\alpha(1 - m_{out}^2)}}{m_{out}}. \quad (5.9)$$

Of course, if U is optimized to have the minimum E , the classes are efficiently retrieved only for $\alpha(1 - m_{out}^2) \ll 1$.

6. Statistical mechanics of the symmetrical model.

We have seen that up to a redefinition of U , the distribution of the stability parameters does not change if the symmetrical or asymmetrical version of the model is chosen. The analysis of the distribution of the Δ 's, that is complete in itself for a perceptron, can only give a tentative

[12, 13] picture of what happens in an ANN. For the symmetrical matrix (2.12) the model is governed by the Hamiltonian $H = - \sum_{i < j} J_{ij} S_i S_j$. A complete analysis of the retrieval states of

the dynamics can be obtained studying the statistical mechanics of the system [17]. In the following we restrict the discussion to the cases in which the overlap with a single individual pattern (e.g. $\{\xi_j^{11}\}$) and/or the overlap with a single class prototype ($\{\xi_j^1\}$) is condensed.

The free energy density of the system is found by a mean field theory, using the replica method to average over the quenched patterns [17]. If replica symmetry holds, then

$$f = \frac{\hat{m}^2}{2} \left(\frac{U}{\sqrt{\alpha}} - \sqrt{\alpha} \left(1 + \frac{1}{1+C} \right) \right) + \frac{\hat{M}^2}{2\sqrt{\alpha}} - \frac{\sqrt{\alpha} r C}{2} + \frac{\alpha}{2\beta} \log(1-C) - \frac{\sqrt{\alpha}}{2} \frac{q}{1-C} - \frac{1}{\beta} \left\langle \left\langle \int Dy \log 2 \cosh \beta \left(\xi \frac{\hat{M}}{\sqrt{\alpha}} + \left[\frac{U}{\sqrt{\alpha}} - \sqrt{\alpha} \left(1 + \frac{1}{1-C} \right) \right] \hat{m} - \frac{m\hat{M}}{\sqrt{\alpha}} \right) + y \sqrt{r} \right) \right\rangle \right\rangle \quad (6.1)$$

where β is the inverse temperature and the double angular brackets imply an average over the distribution of ξ . The meaning of the parameters in (6.1) is the following : \hat{m} is the mean temporal overlap with the retrieved class prototype,

$$\hat{m} = \frac{1}{N} \sum_j \langle S_j \xi_j^1 \rangle ; \quad (6.2)$$

\hat{M} is the normalized mean temporal overlap with $\{\xi_j^{\mu\nu} - m\xi_j^\mu\}$, i.e.

$$\hat{M} = \frac{1}{N} \frac{1}{1-m^2} \sum_j \langle S_j (\xi_j^{11} - m\xi_j^1) \rangle ; \quad (6.3)$$

r is the normalized square of the uncondensed overlaps,

$$r = \frac{1}{\alpha(1-m^2)} \sum_{(\mu\nu) \neq (11)} \left(\sum_j \frac{1}{N} (\xi_j^{\mu\nu} - m\xi_j^\mu) (\xi_j^{\mu\nu} - m\xi_j^\mu) (\langle S_j \rangle - \hat{m}\xi_j^1) \right) ; \quad (6.4)$$

q is the Edwards-Anderson [18] order parameter,

$$q = \frac{1}{N} \sum_j \langle S_j \rangle^2 \quad (6.5)$$

and C is the susceptibility,

$$C = 1 - \frac{\beta}{\sqrt{\alpha}} (1-q) . \quad (6.6)$$

The angular brackets stand for thermal, or temporal, average, and Dy is a normalized Gaussian measure for the slow noise.

The parameters (6.2)-(6.6) satisfy a set of saddle point equations. In the limit $\beta \rightarrow \infty$ these equations can be written as :

$$\hat{m} = \langle \langle \xi \operatorname{erf}(\Lambda_\xi) \rangle \rangle \quad (6.7)$$

$$\hat{M} = \frac{1}{2} [\operatorname{erf}(\Lambda_+) + \operatorname{erf}(\Lambda_-)] \quad (6.8)$$

$$C = \sqrt{\frac{2}{\pi \alpha r}} \langle e^{-\Lambda_\xi^2} \rangle \tag{6.9}$$

$$r = \frac{1 - \hat{m}}{(1 - C)^2} \tag{6.10}$$

where

$$\Lambda_\xi = \frac{\hat{M} + m\xi \left\{ \frac{\hat{m}}{m} \left[U - \alpha \left(1 + \frac{1}{1 + C} \right) \right] - \hat{M} \right\}}{\sqrt{2 \alpha r}} . \tag{6.11}$$

From the definitions of \hat{m} and \hat{M} , equations (6.2) and (6.3), and from their saddle point values (6.7) and (6.8), respectively, the probability E_ξ of making an error in a ξ site is :

$$E_\xi = \frac{1 + [\hat{M} + \xi(\hat{m} + \hat{M})]}{2} = H(\sqrt{2} \Lambda_\xi) . \tag{6.12}$$

When $m \rightarrow 1$ below α_c , $\hat{M} \rightarrow 1$, $C \rightarrow 0$, $\hat{m} \rightarrow m$ and equation (6.12) reduces to equation (4.2) for the first step, as expected.

The numerical solutions of the saddle point equations can be used to find the storage capacity and the corresponding value of the retrieval quality for classes and for individuals, \hat{m} and \hat{M} , respectively. In figure 3 we plot α_c vs. m , which is an increasing function of m , and diverges for $m \rightarrow 1$, as expected. In figure 4 \hat{M} and \hat{m} are plotted. The shape of \hat{M} as a function of m can be understood observing that

$$\hat{M} = 1 - 2(E_+ + E_-)$$

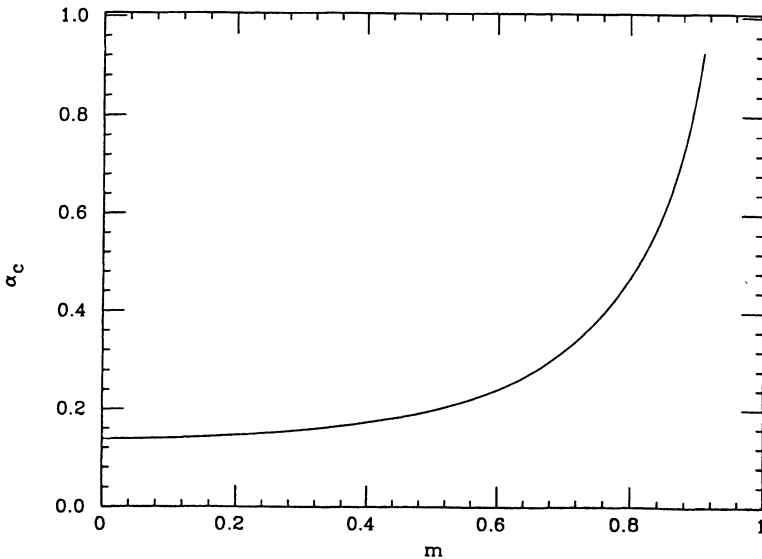


Fig. 3. — The storage capacity of the model vs. m . Qualitatively the shape of the curve is similar to that of optimal storage, reference [1].

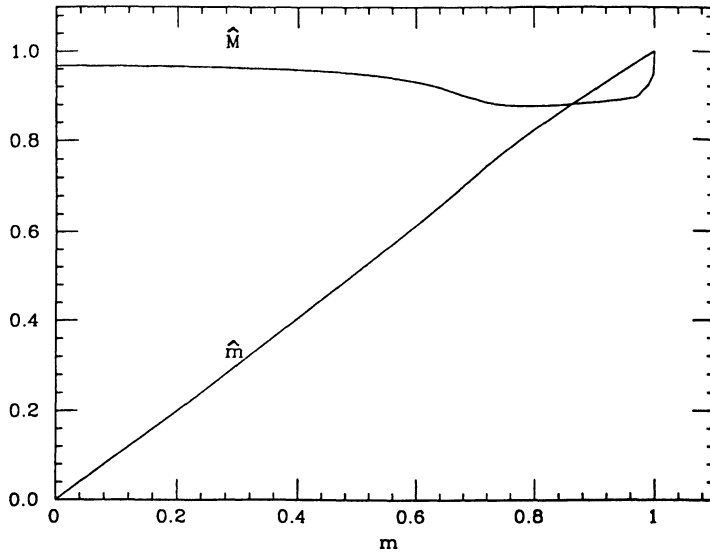


Fig. 4. — \hat{M} and \hat{m} vs. m for $\alpha = \alpha_c$. Note that \hat{m} grows almost linearly and for all value of m one has $\hat{m} \approx m$. \hat{M} has a minimum at $m = 0.78$, as a consequence of the optimization of U .

and that E_- has a maximum for $m > 0$, as can be seen from figure 5. The maximum of E_- in figure 5 has the same origin as those of figures 1 and 2, it is due to the optimization of U . It is convenient, in order to maximize the storage capacity, to have relatively high values of E_- , and correspondingly, low values of \hat{M} . It must be noted also that unlike the first step case, E also has a maximum as a function of m . Note the qualitative accord of figure 5 with figures 1 and 2.

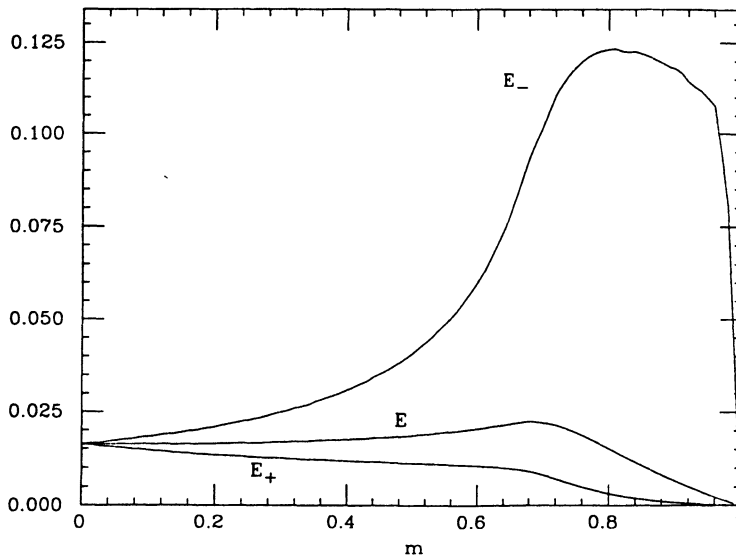


Fig. 5. — E_+ , E_- and E vs. m for $\alpha = \alpha_c$. The maximum in E_- corresponds to the minimum in \hat{M} .

Of course the states in which a class, but no individual, is retrieved ($\hat{m} \neq 0, \hat{M} = 0$) are also relevant fixed points. In this case, for any value of α and m ,

$$C = r = \hat{M} = 0$$

while

$$\hat{m} = 1 ,$$

confirming that the classes are stored with no errors. This can be understood since we have constructed synaptic matrix so as to have no noise on the class prototypes.

7. Symmetrical dilution of synaptic matrix.

We now proceed to study the effect on the attractors of a lesion of the synaptic matrix. In order to keep the system Hamiltonian we consider a symmetric dilution of the synaptic matrix according (4.4), (4.5).

It has been proven that such a kind of dilution is equivalent to the introduction in the Hamiltonian of an SK spin-glass interaction among neurons [19],

$$-\frac{1}{2} \sum_{i \neq j} \bar{J}_{ij} S_i S_j , \tag{7.1}$$

with \bar{J}_{ij} independent Gaussian variables with zero mean and variance $1/N\alpha(1-p)/p$. In the replica symmetric theory, this effective interaction adds to the free energy a term of the form [20]

$$\frac{\beta}{4} \sqrt{\alpha} \frac{(1-p)}{p} q^2 . \tag{7.2}$$

The effect of this term in the saddle point equations, is to modify equation (6.10) to read :

$$r = \frac{1 - \hat{m}}{(1 - C)^2} + \frac{1 - p}{p} . \tag{7.3}$$

A first effect of the dilution on the attractor is analogous to the effect on the perceptron. The noise induced by the lesion affects asymmetrically the plus sites and the minus sites of the individual attractors. Repeating the argument of section 5, one finds that the increment in E_- is greater then the corresponding increment in E_+ . A more important effect, regards the disappearance of the individual attractors. The capacity limit for the recall of individuals is lowered by the new term is r , as can be realized observing that α_c is a decreasing function of the noise r . Furthermore, due to the dependence of the additional noise (7.1) on α , another critical value of α appears, to be denoted by α_{Cl} , at which the class prototypes cease to be attractors. This is a point of ferromagnet-spin glass phase transition. Nevertheless, for a small noise level, $\alpha_{Cl} \gg \alpha_c$. There is a whole region $\alpha_c < \alpha < \alpha_{Cl}$ in which the network is in a state of prosopagnosia. The individual information is completely lost, while the information about the classes is still stored in the network. Only for very high damage, the network loses completely all information, falling into a state of agnosia.

8. Comparison with optimal storage.

In section 4, we have found that in the limit $m_{out} \rightarrow 1$ the storage capacity of a network with the matrix (2.14) approaches the Gardner bound and in addition, the error goes to zero and Δ_c tends to the Gardner saddle point value of M . Hence in this limit, the matrix (2.14) shares many feature with the matrices in the Gardner volume. In this section we compare explicitly

the matrix J_{ij} (2.14) with a generic matrix that stores the given patterns with no errors, i.e. a matrix within the volume of J_{ij}^* 's that satisfy [1]

$$\begin{aligned} \xi_i^{\mu\nu} \sum_j J_{ij}^* \sigma_j^{\mu\nu} &> 0 \\ \sum_j J_{ij}^{*2} &= 1. \end{aligned} \quad (8.1)$$

Specifically, we compute the overlap

$$Q = \sum_j J_{ij} J_{ij}^* \quad (8.2)$$

averaged over the Gardner volume. Q can also be interpreted as the angle cosine between the matrix (2.14) and a generic matrix that satisfies (8.1).

The probability distribution corresponding to the constraints (8.1) is

$$P \{J_{ij}^*\} \prod_j dJ_{ij} = \frac{\delta \left(\sum_j J_{ij}^{*2} - 1 \right) \prod_{\mu\nu} \theta \left(\xi_i^{\mu\nu} \sum_j J_{ij}^* \sigma_j^{\mu\nu} \right) \prod_j dJ_{ij}^*}{\int \delta \left(\sum_j J_{ij}^{*2} - 1 \right) \prod_{\mu\nu} \theta \left(\xi_i^{\mu\nu} \sum_j J_{ij}^* \sigma_j^{\mu\nu} \right) \prod_j dJ_{ij}^*}. \quad (8.3)$$

The quantity Q is self-averaging with respect to the measure (8.3) and does not depend on the particular sample of patterns stored. Substituting (2.14) in (8.3) we obtain

$$\begin{aligned} Q &= \frac{1}{N \sqrt{\alpha (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)}} \sum_{\mu\nu j} (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^\mu) (\sigma_j^{\mu\nu} - m_{\text{in}} \sigma_j^\mu) J_{ij}^* + \\ &+ \frac{Um_{\text{out}}}{m_{\text{in}} \sqrt{\alpha}} \sqrt{\frac{1 - m_{\text{in}}^2}{1 - m_{\text{out}}^2}} \frac{1}{N} \sum_{\mu j} \xi_i^\mu \sigma_j^\mu J_{ij}^* \\ &- \frac{c}{m_{\text{in}} \sqrt{\alpha (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)}} \frac{1}{N^2} \sum_{\mu\nu\alpha k j} \sigma_j^\alpha \sigma_k^\alpha (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^\mu) (\sigma_k^{\mu\nu} - m_{\text{in}} \sigma_k^\mu) J_{ij}^*. \end{aligned} \quad (8.4)$$

Since $M = \sum_j \xi_i^\mu J_{ij}^* \sigma_j^\mu$ is a quantity that does not fluctuate we can write the second and the third terms in (8.4), respectively, as

$$\frac{Um_{\text{out}}}{m_{\text{in}} \sqrt{\alpha}} \sqrt{\frac{1 - m_{\text{in}}^2}{1 - m_{\text{out}}^2}} \frac{p_c}{N} M \quad (8.5)$$

and

$$-\frac{cM}{m_{\text{in}} \sqrt{\alpha (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)}} \frac{1}{N^2} \sum_{\mu\nu\alpha k} \xi_i^\alpha \sigma_k^\alpha (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^\mu) (\sigma_k^{\mu\nu} - m_{\text{in}} \sigma_k^\mu), \quad (8.6)$$

respectively. These two terms are of order p_c/N and do not contribute to Q when $N \rightarrow \infty$. So the terms proportional to U and c in the J_{ij} 's, that are essential for the storage, are here totally irrelevant. This fact become less surprising when one observes that the terms with U and c in J_{ij} 's are of order $\sqrt{p_c/N}$, relative to the leading term

$$J_{ij}^{\text{lead}} = \frac{1}{\sqrt{\alpha (1 - m_{\text{in}}^2)(1 - m_{\text{out}}^2)}} \sum_{\mu\nu} (\xi_i^{\mu\nu} - m_{\text{out}} \xi_i^\mu) (\sigma_j^{\mu\nu} - m_{\text{in}} \sigma_j^\mu), \quad (8.7)$$

which is of order $1/\sqrt{N}$. Developing the leading term of (8.4) we obtain

$$Q = \frac{1}{N \sqrt{\alpha (1 - m_{in}^2)(1 - m_{out}^2)}} \times \sum_{\mu\nu} \left[(1 - m_{out} \xi_i^{\mu\nu} \xi_i^\mu) \sum_j \xi_i^{\mu\nu} J_{ij}^* \sigma_j^{\mu\nu} + (\xi_i^{\mu\nu} \xi_i^\mu - m_{out}) M \right]. \quad (8.8)$$

Hence, in order to compute Q , all that is needed is the probability distribution of the stability parameters

$$\Delta_\xi^* = \sum_j \xi_i^{\mu\nu} J_{ij}^* \sigma_j^{\mu\nu} \quad (8.9)$$

at the plus and the minus sites. In fact, if we denote this function in the ξ sites by $F_\xi(\Delta_\xi^*)$ then

$$Q = \sqrt{\frac{\alpha}{(1 - m_{in}^2)(1 - m_{out}^2)}} (1 - m_{out} \xi) \int F_\xi(\Delta_\xi^*) \Delta_\xi^* d\Delta_\xi^*, \quad (8.10)$$

where the angular brackets stand for the average over the distribution (3.1) of ξ i.e. expliciting the average over ξ

$$Q = \sqrt{\frac{\alpha (1 - m_{out}^2)}{(1 - m_{in}^2)}} \sum_{\xi = \pm 1} \int F_\xi(\Delta_\xi^*) \Delta_\xi^* d\Delta_\xi^*. \quad (8.11)$$

The distribution $F_\xi(\Delta_\xi^*)$ was computed in reference [14] as a function of the order parameter M and the parameter $q = \sum_j J_{ij}^{*a} J_{ij}^{*b}$, where the distinct matrices J_{ij}^{*a}, J_{ij}^{*b} satisfy equation

(8.1). It is :

$$F_\xi(\Delta_\xi^*) = - \int \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \theta(\Delta_\xi^*) \frac{\frac{d}{d\Delta_\xi^*} H\left(\frac{\Delta_\xi^* - \xi m_{in} M + \sqrt{q(1 - m_{in}^2)} t}{\sqrt{(1 - q)(1 - m_{in}^2)}}\right)}{H\left(\frac{-\xi m_{in} M + \sqrt{q(1 - m_{in}^2)} t}{\sqrt{(1 - q)(1 - m_{in}^2)}}\right)} \quad (8.12)$$

where, for a given $\alpha < \alpha_c$, q and M satisfy the Gardner saddle point equations [1]. One therefore finds :

$$Q = \frac{\sqrt{\alpha (1 - m_{out}^2)(1 - q)}}{2} \times \sum_{\xi = \pm 1} \int \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \int_0^\infty \frac{d\Delta_\xi^*}{\sqrt{2\pi}} \frac{\exp\left(-\Delta_\xi^* + \frac{\sqrt{q} t - \xi \frac{m_{in}}{\sqrt{1 - m_{in}^2}} M}{\sqrt{1 - q}}\right)^2}{H\left(-\Delta_\xi^* + \frac{\sqrt{q} t - \xi \frac{m_{in}}{\sqrt{1 - m_{in}^2}} M}{\sqrt{1 - q}}\right)} \quad (8.13)$$

Q is consequently, independent of m_{in} because, as was shown in [10], $\frac{m_{in}}{\sqrt{1 - m_{in}^2}} M$ does not depend on m_{in} . For $q = 1$, i.e. $\alpha = \alpha_c$ (8.13) reduces to

$$Q = \frac{\frac{m_{in}}{\sqrt{1 - m_{in}^2}} M}{2 m_{out}} \sqrt{\alpha (1 - m_{out}^2)} \tag{8.14}$$

which tends to 1 in the limit $m_{out} \rightarrow 1$. From (8.14) one can see that for all $m_{out} \neq 1$, one has $Q < 1$, i.e. the angle formed by matrix (2.14) and the Gardner volume is different from zero. In figure 6, Q is plotted for fixed values of q . Q is an increasing function of m_{out} and one sees that as m_{out} increases, the distance of matrix (2.14) from the Gardner volume decreases. This can be intuitively understood observing that the retrieval properties of the matrix (2.14) are more and more similar to that of the optimal matrices as m_{out} approaches the 1.

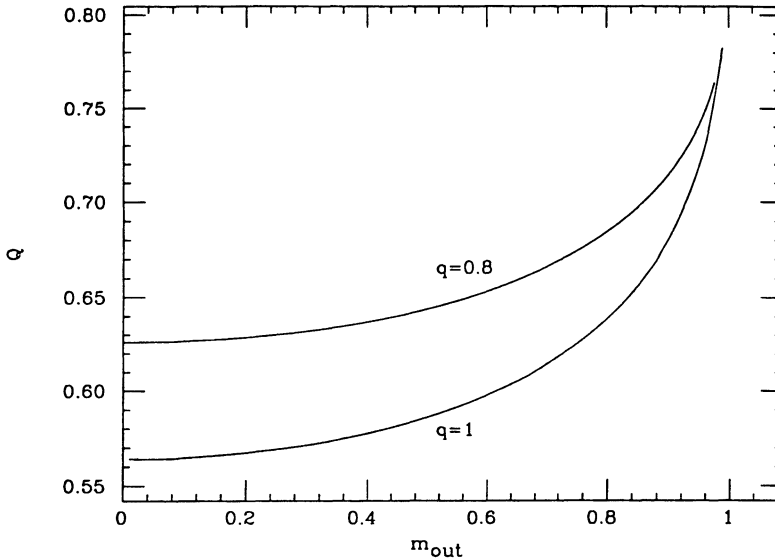


Fig. 6. — Q vs. m_{out} for $q = 1$ and $q = 0.8$. For fixed q , Q is an increasing function of m_{out} showing that the distance of matrix (1.14) from the Gardner volume decreases with m_{out} .

Actually, for a generic q , as $m_{out} \rightarrow 1$ (8.13) reduces to

$$Q = \sqrt{q} . \tag{8.15}$$

This implies that

$$\int P(\{J_{ij}^*\}) J_{ij}^* \prod_j J_{ij}^* = \sqrt{q} J_{ij} , \tag{8.16}$$

i.e. J_{ij} is the axis of a cone in which the J_{ij}^* are uniformly distributed.

Conclusion.

In this paper we have considered a synaptic matrix that stores efficiently uncorrelated classes of pattern. The storage capacity limit behaves qualitatively as in the case of optimal storage : it is an increasing function of the correlation of the individuals with their class prototypes, and diverges in the limit of full correlation. We have obtained this result by modifying the synaptic matrix of reference [8, 9].

In the framework of this model the learning of a new pattern, after $O(N)$ have already been memorized, can be imagined as occurring in two steps. The new pattern is compared with the memorized classes, and the difference with the representative of the class with which it overlaps class is stored in a Hebbian-like outer product. In addition the projection of this difference on the space spanned by the classes prototypes is subtracted. This reduces the Gaussian noise on the stability parameters of the patterns stored.

Then, together with the individual patterns, the classes prototypes are stored so as to have greater stability than the individual pattern. This is necessary to minimize the number of errors when an individual pattern is retrieved. It then follows that the plus sites are more stable than the minus sites in accord with what happens in the Gardner case. Prosopagnosia appears at this point. It is modeled by the fact that a lesion in the network structure destroys first the information about individuals then the information about classes. Only at much higher level of disruption the information about classes is destroyed, mimicking agnosia.

We have studied the matrix both in the context of ANN's and perceptrons, and have argued that as far as memory errors are concerned the two become equivalent when one identifies the errors in the perceptron output with those present in the configuration of the attractor.

The retrieval of information in perceptron, is fully characterised by the probability distribution of the local stability parameters of individuals and class prototypes. For ANN's we have considered a symmetric matrix and an asymmetric one. At the level of the probability distribution of the stability parameters, the two matrices have the same properties. However, there would normally be dynamical differences. The symmetric matrix has been studied in detail by statistical mechanics, which confirms the picture obtained from the analysis of the stability parameters.

We did not simulate extensively the dynamics of the ANN. We performed simulations on the zero temperature dynamics of neural networks up to 100 neurons with these synaptic matrices. Both the symmetric and the asymmetric model have a storage capacity consistent with the results of section 6. We observed that, as expected, beyond the limit of storage capacity, initial states highly correlated with one individual pattern evolves to the corresponding class prototype.

Acknowledgments.

S.F. gratefully acknowledges a scholarship of Selenia S.P.A. D.J.A and S.F. are grateful for the hospitality extended to them by Imperial College London, where the work was completed.

References

- [1] GARDNER E., *J. Phys. A* **21** (1988) 257 ; *Europhys. Lett.* **4** (1987) 481.
- [2] WILLSHAW D. J., BUNEMAN O. P. and LONGNET-HIGGINS H. C., *Nature* **222** (1969) 960.
- [3] HOPFIELD J. J., *Proc. Nac. Acad. Sci. USA* **79** (1982) 2554.
- [4] BUHMANN J., DIVKO R. and SCHULTEN K., Associative memory with high informational content, Munich Technical University preprint (1987).

- [5] TSODYSK M. V. and FEIGELMANN M. V., *Europhys. Lett.* **6** (1988) 101.
- [6] PEREZ VICENTE C. J. and AMIT D. J., *J. Phys. A : Math. Gen.* **22** (1989) 559.
- [7] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. A* **35** (1987) 2293.
- [8] PARGA N. and VIRASORO M. A., *J. Phys. France* **47** (1986) 1857.
- [9] FEIGELMAN M. V. and IOFFE L., Chernogolovka preprint (1988).
- [10] DELGIUDICE P., FRANZ S. and VIRASORO M. A., *J. Phys. France* **50** (1989) 121.
- [11] PEREZ VICENTE C. J., *preprint* (1988).
- [12] KRAUTH W., MEZARD M. and NADAL J. P., *Complex Syst.* **2** (1988) 387.
- [13] KEPLER T. B. and ABBOT L. F., *J. Phys. France* **49** (1988) 1657.
- [14] VIRASORO M. A., *Europhys. Lett.* **7** (1988) 299.
- [15] MATTIS D. C., *Phys. Lett.* **56A** (1976) 421.
- [16] WEISBUCH G. and FOLGEMAN-SOULIÉ F., *J. Phys. Lett. France* **46** (1985) L-623 ;
MCÉLIECE R. J., POSNER E. C., RODEMICH E. R. and VENKATESH S. S., *Caltech preprint* (1986).
- [17] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. Lett.* **55** (1985) 1530 ; *Ann. Phys.*
173 (1987) 30.
- [18] EDWARDS S. F., ANDERSON P. W., *J. Phys. France* **5** (1975) 956.
- [19] SHERRINGTON D., KIRKPATRICK S., *Phys. Rev. Lett.* **32** (1975) 1792.
- [20] SOMPOLINSKY H., In *Heidelberg Colloquium on Glassy Dynamics* (1986) p. 486.