Compensatory mechanisms in an attractor neural network model of Schizophrenia

D. Horn
School of Physics and Astronomy,
Raymond and Beverly Sackler Faculty of Exact Sciences,
Tel Aviv University, Tel Aviv 69978, Israel

E. Ruppin
Department of Computer Science,
Raymond and Beverly Sackler Faculty of Exact Sciences,
Tel Aviv University, Tel Aviv 69978, Israel

March 31, 1994

Abstract

We investigate the effect of synaptic compensation on the dynamic behavior of an attractor neural network receiving its input stimuli as external fields projecting on the network. It is shown how, in face of weakened inputs, memory performance may be preserved by strengthening internal synaptic connections and increasing the noise level. Yet, these compensatory changes necessarily have adverse side effects, leading to spontaneous, stimulus-independent retrieval of stored patterns. These results can support Stevens' recent hypothesis that the onset of Schizophrenia is associated with frontal synaptic regeneration, occurring subsequent to the degeneration of temporal neurons projecting on these areas.

1 Introduction

A prominent feature of attractor neural networks (ANN) as models for associative memory is their robustness, i.e. their ability to maintain performance in face of damage to their neurons and synapses. Robustness of biological systems is due, however, not just to their distributed structure, but depends also on compensatory mechanisms which they employ. In a recent paper [Horn *et al.*, 1993], we have shown that while some of the synapses are deleted, compensatory strengthening of all the remaining ones can rehabilitate the system, and that different compensation strategies can account for the observed variation in the progression of Alzheimer's disease. The ANN we examined represented an isolated cortical module, receiving its input as an initial state into which the network is 'clamped', after which it evolves in an autonomous manner.

In this work, we study an ANN representing a cortical module receiving input patterns as persistent external fields projecting on the network (as, for example, in [Parisi and Nicolis, 1990]), presumably arising from other cortical modules. We examine the network's potential to compensate for the weakening of the external input field. It is shown that, to a certain limit, memory performance may be preserved by strengthening the internal synaptic connections and by increasing the noise which stands for other, non-specific external connections. However, these compensatory changes necessarily have adverse side effects, leading to spontaneous, stimulus-independent, retrieval of stored patterns.

Our interest in studying synaptic deletion and compensation in an external-input driven model is motivated by Stevens' recent hypothesis concerning the possible role of such synaptic changes in the pathogenesis of Schizophrenia [Stevens, 1992]. Schizophrenia is a devastating psychiatric disease, whose broad clinical picture ranges from 'negative', deficit symptoms including pervasive blunting of affect, thought, and socialization, to 'positive' symptoms such as florid hallucinations and delusions. Its worldwide prevalence is approximately 1 percent, and even with the most up-to-date treatment the majority of patients suffer from chronic deterioration. While the introduction of relatively objective criteria has improved diagnostic uniformity, and dopamine-blocking neuroleptic drugs have enhanced symptomatic relief, the diagnosis still remains phenomenologic, and the treatment palliative. Our goal in this paper is to provide a computational account of Stevens' theory of the pathogenesis of Schizophrenia, in a framework of an ANN model.

Only a few neural network models of Schizophrenia have been proposed. Hoffman [Hoffman, 1987, Hoffman and Dobscha, 1989] has previously presented Hopfield-like ANN models of schizophrenic disturbances. He has demonstrated on small networks that when, due to synaptic deletion the network's memory capacity becomes overloaded, the memories' basins of attraction are distorted and 'parasitic foci' emerge, which he suggested could underlie some schizophrenic symptoms such as hallucinations and delusions. This scenario implies, however, that a considerable deterioration of memory function should accompany the appearance of psychotic symptomatology already in the early stages of the disease process, in contrast with the clinical data [Mesulam, 1990, Kaplan and Sadock, 1991]. We shall show that when the broad spectrum of synaptic changes that occur in accordance with Stevens' theory are considered, memory functions may remain preserved while spontaneous retrieval rises. The latter may be an important mechanism participating in the generation of some psychotic symptoms.

Cohen and Servan-Schreiber have presented connectionist feed-forward back-propagation networks that were able to simulate normal and schizophrenic performance in several attention and language-related tasks [Cohen and Servan-Schreiber, 1992]. In the framework of a model corresponding to the assumed function of the prefrontal cortex, they demonstrate that some schizophrenic functional deficits can arise from neuromodulatory effects of dopamine which may take place in schizophrenia. Their model obtains an impressive quantitative fit with human performance in a broad spectrum of cognitive phenomena. They also provide a thorough review of previous neural models of schizophrenia to which the interested reader is referred.

In this work the discussion is restricted to memory retrieval, assuming that the bulk of long-term memory has already been stored. The next section defines the network and its dynamics. Section 3 describes its role as a functional model of Stevens' hypothesis. In section 4 we derive an analytic approximation of the network performance, and study the relation between stimuli-driven retrieval and spontaneous retrieval following synaptic deletion and compensation. The results of this approximation and of corresponding simulations are presented in section 5. Finally, the relevance of our findings to Stevens' hypothesis concerning the pathogenesis of Schizophrenia is discussed.

2 The model

We build upon a biologically-motivated variant of Hopfield's ANN model, proposed by Tsodyks & Feigel'man (TF) [Tsodyks and Feigel'man, 1988]. Each neuron i is described by a binary variable $S_i = \{1, 0\}$ denoting an active (firing) or passive (quiescent) state, respectively. $M = \alpha N$ distributed memory patterns ξ^μ are stored in the network. The elements of each memory pattern are chosen to be 1 (0) with probability p $(1-p)$ respectively, with $p \ll 1$. All N neurons in the network have a fixed uniform threshold θ.

In its initial, undamaged state, the weights of the internal synaptic connections are

$$W_{ij} = \frac{c_0}{N} \sum_{\mu=1}^{M} ({\xi^\mu}_i - p)({\xi^\mu}_j - p) \ , \tag{1}$$

where $c_0 = 1$. The post-synaptic potential (input field) h_i of neuron i is the sum of internal contributions from other neurons in the network and external projections F_i^e

$$h_i(t) = \sum_j W_{ij} S_j(t-1) + F_i^e \ . \tag{2}$$

The updating rule for neuron i at time t is given by

$$S_i(t) = \begin{cases} 1, & \text{with prob. } G(h_i(t) - \theta) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where G is the sigmoid function $G(x) = 1/(1 + \exp(-x/T))$, and T denotes the noise level. The activation level of the stored memories is measured by their *overlaps* m^μ with the current state of the network, defined by

$$m^\mu(t) = \frac{1}{p(1-p)N} \sum_{i=1}^{N} (\xi_i^\mu - p) S_i(t) \ . \tag{4}$$

The initial state of the network $S(0)$ is random, with average activity level $q < p$, reflecting the notion that the network's spontaneous level of activity is lower than its activity in the persistent attractor states. *Stimulus-dependent retrieval* is modeled by orienting the field F^e with one of the memorized patterns (the *cued* pattern, say ξ^1), such that

$$F_i^e = e \cdot {\xi^1}_i \ \ , \quad (e > 0) \ . \tag{5}$$

Following the dynamics defined in (2) and (3), the network state evolves until it converges to a stable state. Performance is then measured by the final overlap m^1.

In addition to investigating the network's activity in response to the presentation of an external input, we also examine its behavior in the absence of any specific stimulus. In this case, the network may either continue to wander around in a state of random low baseline activity, or it may converge onto a stored memory state. We refer to the latter process as *spontaneous retrieval.*

In its undamaged state, the noise level $T = T_0$ is low and the strength $e = e_0$ of the external projections is chosen such that the network will flow dynamically into the cued memory pattern (ξ^1 in our example) with high probability. The stored memories are hence attractors of the network's dynamics. We find the optimal threshold which ensures high performance in the initial undamaged state, and investigate the effects of synaptic changes on the network's behavior, as described in the next section.

3 Stevens' hypothesis in an ANN framework

The wealth of data gathered concerning the pathophysiology of Schizophrenia supports the involvement of two major cortical areas, the frontal and the temporal lobes [Kaplan and Sadock, 1991]. On one hand, there are atrophic changes in the hippocampus and parahippocampal areas in the brains of a significant number of schizophrenic patients, including neural loss and gliosis. On the other hand, neurochemical and morphometric studies testify to an expansion of various receptor binding sites and increased dendritic branching in the frontal cortex of schizophrenics (reviewed in [Stevens, 1992]). Unifying these temporal and frontal findings, Stevens [1992] has claimed that the onset of Schizophrenia is associated with reactive anomalous sprouting and synaptic reorganization taking place at the frontal lobes, subsequent to the degeneration of temporal neurons projecting at these areas.

To study the possible functional aspects of Stevens' hypothesis, we model a frontal module as an ANN, receiving its inputs from degenerating temporal projections and undergoing reactive synaptic regeneration. It is conventionally believed that the hippocampus and its anatomically related cortical medial temporal structures (MTL) have an important role in establishing long-term memory for facts and events in the neocortex [Squire, 1992]. Widespread damage to MTL structures may result in both severe anterograde and retrograde amnesia, so that the hippocampus may have an important role not only in storage but also in retrieval of memory. The work of Heit et al. [Heit *et al.*, 1988] suggests that the MTL contributes specific information rather than diffuse modulation to the encoding and retrieval of memories. We hence assume that memory retrieval from the frontal cortex is invoked by the firing of external incoming MTL projections.

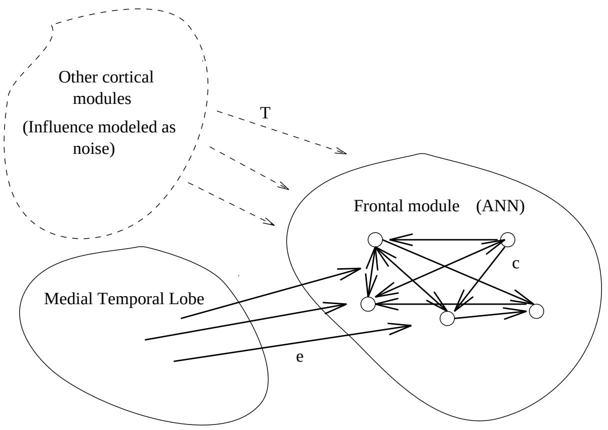

Figure 1: A schematic illustration of the model. Each frontal module is modeled as an ANN whose neurons receive inputs via three kinds of connections: internal connections from other neurons in the module, external connections from MTL neurons, and diffuse external connections from other cortical modules.

The basic analogy is straightforward and is described in figure 1. When schizophrenic pathological processes occur, the degeneration of temporal projections is modeled by decreasing the strength e of the incoming input fibers. [1] Reactive frontal synaptic sprouting and reorganization are modeled by increasing the strength c of the internal connections, such that

$$W_{ij_{new}} = c \cdot W_{ij_{old}} \ , \ (c > 1) \ . \tag{6}$$

The expansion of diffuse external projections is modeled as increased noise T, reflecting the assumption that during stimulus-dependent retrieval performed via the temporal-frontal pathway, the contribution of other cortical areas is nonspecific.

[1] Alternatively one can choose random deletion of the inputs to the neurons of the ANN, with a fraction e/e_0 surviving intact. All the results described in this work remain qualitatively similar. The only notable difference is that increased noise has a weaker compensatory effect.

4 Analysis of the model

4.1 Threshold determination and the overlap equation

The TF model has been studied quantitavely [Tsodyks and Feigel'man, 1988, Tsodyks, 1988] by deriving a set of mean-field equations that describe the state of several macroscopic variables as the network state evolves. In the limit of low activity rate p and low memory load α the only equation needed is the macroscopic overlap equation which describes the level of overlap m with the cued memory pattern [Tsodyks and Feigel'man, 1988].

Given an overlap $m^1(t)$ at iteration t, we estimate the overlap $m^1(t+1)$. By (4),

$$m^1(t+1) = \tag{7}$$

$$\frac{1}{p(1-p)}\Big[(1-p)P(\xi_i^1=1)P(S_i(t+1)=1|\xi_i^1=1) - pP(\xi_i^1=0)P(S_i(t+1)=1|\xi_i^1=0)\Big] =$$

$$= P(S_i(t+1)=1|\xi_i^1=1) - P(S_i(t+1)=1|\xi_i^1=0)\ .$$

Using (2), (3), (5) and (6) we find

$$S_i(t+1) = G\left[\frac{c}{N}\left(\sum_{j=1}^{N}(\xi_i^1-p)(\xi_j^1-p)S_j(t) + \sum_{\mu=2}^{M}\sum_{j=1}^{N}(\xi_i^\mu-p)(\xi_j^\mu-p)S_j(t)\right) + e{\xi_i}^1 - \theta\right] \tag{8}$$

where we have separated the signal from the noise term. Hence, the input field of neuron i is conditionally normally distributed (with respect to its correct value ${\xi_i}^1$) with means $\mu_1 = cp(1-p)^2m^1 + e - \theta$, $\mu_0 = -cp^2(1-p)m^1 - \theta$, where the indices refer to the two possible choices of the value of ξ. The variance can be approximated by $\sigma^2 \approx \alpha q p^2 c^2 \approx \alpha p^3 c^2$ in leading order in p.

The probability of firing of a sigmoidal neuron i, with noise level T, receiving a normally distributed input field h_i with mean μ and standard deviation σ is

$$P(S_i=1|h_i) = \int_{-\infty}^{+\infty}\frac{1}{1+e^{-(\mu+\sigma z)/T}}\phi(z)dz \approx \int \Phi\left(\frac{\mu+\sigma z}{1.702T}\right)\phi(z)dz = \tag{9}$$

$$\int P\left(Z_2 \le \frac{\mu+\sigma z}{1.702T}\right)\phi(z)dz = P\left(Z_2 \le \frac{\mu+\sigma Z_1}{1.702T}\right) =$$

$$P\left(\frac{Z_2 - \frac{\sigma}{1.702T}Z_1}{\sqrt{1+\frac{\sigma^2}{(1.702T)^2}}} \le \frac{\frac{\mu}{1.702T}}{\sqrt{1+\frac{\sigma^2}{(1.702T)^2}}}\right) = \Phi\left(\frac{\mu}{\sqrt{(1.702T)^2+\sigma^2}}\right)$$

where Φ is the standard normal distribution function, ϕ is the standard density function, and where we have used a sigmoidal approximation to the Gaussian cumulative distribution function (see [Haley, 1952])

$$Sup_x|\Phi(x) - \frac{1}{1+\exp(-1.702x)}| < 0.01 \ . \tag{10}$$

Hence, by (7) and (9), the overlap equation is

$$m^1(t+1) = \Phi\left(\frac{cp(1-p)^2 m^1(t) + e - \theta}{\sqrt{(1.702T)^2 + \alpha p^3 c^2}}\right) - \Phi\left(\frac{-cp^2(1-p)m^1(t) - \theta}{\sqrt{(1.702T)^2 + \alpha p^3 c^2}}\right) \ . \tag{11}$$

Maximizing the value of (11) we find that the optimal threshold is

$$\theta = 0.5[(1-2p)p(1-p)cm^1(t) + e] \ . \tag{12}$$

It is of interest to note that the optimal threshold is a function of the current overlap $m(t)$. Best results would have therefore been achieved with a dynamically varying threshold that constantly increases in every iteration until convergence is achieved, testifying to the possible computational efficiency of neural adaptation, a characteristic of many cortical neurons [Connors and Gutnick, 1990]. However, for simplicity we used a fixed threshold whose value in the baseline, undamaged phase ($c = c_0 = 1 \ , \ e = e_0$)

$$\theta = 0.45[(1-2p)p(1-p) + e_0] \tag{13}$$

was found to generate the best results in our simulations.

We see that the optimal threshold is determined by the initial baseline values of c_0 and e_0. In our simulations, described in the following, we have used $c_0 = 1, e_0 = 0.035$ and $T_0 = 0.005$. When e is decreased, the threshold is no more optimal. This decrease may be compensated for by an increase in c as well as in T, as will be shown below. From expressions 12 and 13, it is obvious that from a computational standpoint, when the external synapses degenerate, the performance of the ANN may be improved by threshold adjustment instead of internal synaptic strengthening. Indeed, from a biological point of view, the strengthening of external diffuse projections may have a net excitatory effect on frontal neurons (following the common belief that cortical long-range connections are excitatory), and therefore may lead to a combination of both increased noise levels and effective threshold decrease. In our analysis we have assumed the former, in the form of the temperature T, but neglected the latter. For clarity of exposition, we have chosen to separate between compensatory synaptic modifications which primarily change the mean of the neuron's input field, and those modifications which only change its variance. An effective decrease of the threshold would simply call for a different choice of the compensation parameter c.

4.2 The effects of synaptic changes and noise level

Let $\bar{\mu_1} = p(1-p)^2$,$\bar{\mu_0} = p^2(1-p)$ and $\bar{\sigma} = \sqrt{(1.702T)^2 + \alpha p^3 c^2}$, and rewrite (11) as

$$m^1(t+1) = \Phi\left(\frac{cm^1(t)\bar{\mu_1} + e - \theta}{\bar{\sigma}}\right) + \Phi\left(\frac{cm^1(t)\bar{\mu_0} + \theta}{\bar{\sigma}}\right) - 1 \; . \tag{14}$$

Then, by (14), the following observations are straightforward:

1. The fixed point solution of (14) is monotonically increasing in c.

2. Since $\bar{\mu_1} \gg \bar{\mu_0}$, the first term dominates this equation.

3. In the initial undamaged state, $e_0 < \theta$, ensuring that the network does not follow non-stored patterns if the latter are presented as inputs. Therefore, at $t = 0$ and $m^1(0) \approx 0$, the argument of the first term is negative, and increased noise increases the magnitude of the overlap $m^1(1)$ (the overlap of the input pattern). However, as the dynamics evolve, the argument of the first term becomes positive, and the increased noise reduces the final overlap.

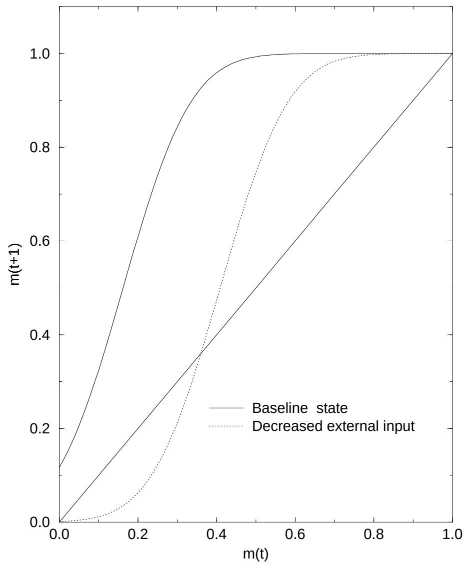

Figure 2: The map $(m(t+1) \mid m(t))$ generated by iterating the overlap equation. $\alpha = 0.05$, $p = 0.1$, $e_0 = 0.035$, $c = 1$ and $T = T_0 = 0.005$. In the initial undamaged state (full curve) $e = 0.035$ and in the decreased input case $e = 0.015$. Recall that θ remains fixed and its value is determined by (13), with $e_0 = 0.035$, $c_0 = 1$ and $p = 0.1$. All figures are based on this choice of initial synaptic strengths and threshold, as well as on $\alpha = 0.05$.

Figure 2 displays the map $(m^1(t+1) \mid m^1(t))$ defined by (11), describing stimulus-dependent retrieval. In the baseline, undamaged, state there is only one (stable) fixed point solution, with $m^1 \approx 1$ (full curve). After the weakening of external projections, two additional fixed points may appear (dotted curve). The lowest fixed point is not visible on the scale of figure 2, but it always coexists with the middle fixed point since, by (11), $m^1(1) > 0$ when $m^1(0) = 0$. It is easy to see that the middle fixed point is unstable, while the two extreme ones are stable. The middle, unstable, fixed point denotes the critical value C_r that divides the map; only input patterns with overlap $m^1 > C_r$ will converge to the higher fixed point, signifying successful retrieval. As the initial overlap of an input pattern is essentially zero, equation (11) will converge to its lower fixed point whenever it exists, resulting in stimulus-retrieval failure.

The compensatory potentials of internal synaptic strengthening and increased noise level are illustrated in figure 3. In both cases we see that with sufficient compensatory increase the original situation (where only a single, high overlap, fixed point exists) is restored. Note that as the noise level is increased the magnitude of the highest fixed point decreases monotonically, in accordance with observation 3.

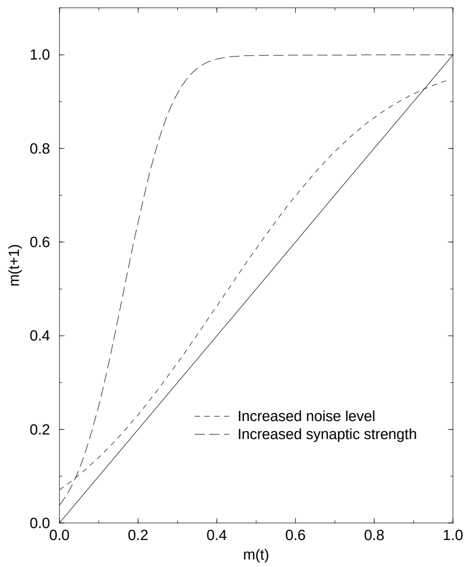

Figure 3: The map $(m(t+1) \mid m(t))$ after a decrease in the magnitude of external projections ($e = 0.015$) and a compensatory increase in the internal synaptic strength (long-dashed curve, $c = 2.5$ and $T = 0.005$), or in the noise level (dashed curve $c = 1$ and $T = 0.015$). These curves should be compared with the $c = 1$, $T = 0.005$ curve in figure 2.

To study spontaneous retrieval, we calculate the overlap $m^\mu(0)$ between the initial network state S and each memory pattern ξ^μ. As shown in the Appendix, the maximal overlap m_{max} has an almost deterministic value. This enables us to model spontaneous retrieval by entering m_{max} as the initial $m(0)$ in (11), letting $e = 0$ (no external stimulus is present), and iterating the overlap equation. The map $(m(t+1) \mid m(t))$ generated in this fashion has a similar form to that shown previously in the stimulus-driven mode, as illustrated in figure 4. Analogous to the case of stimulus-dependent retrieval, the middle fixed point denotes the critical value C_s, such that only when $m_{max} > C_s$ expression (11) converges to its higher fixed point, denoting spontaneous retrieval of a stored memory pattern.

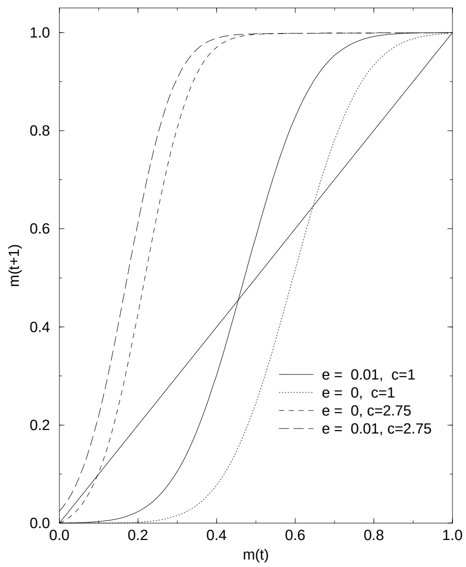

Figure 4: The maps $(m(t+1) \mid m(t))$ of spontaneous and stimulus retrieval after a decrease in the magnitude of external projections, and following a compensatory increase in the internal synaptic strength. $T = 0.005$.

Figure 4 illustrates the effects of synaptic changes on spontaneous retrieval. As e decreases (to $e = 0.010$, compare with figure 2), the curve corresponding to the stimulus-driven map is shifted to the right, approaching the spontaneous-retrieval curve ($e = 0$). Following observation 1, increasing c results in a leftward (and upward) shift of both curves, possibly maintaining successful stimulus-retrieval (by eliminating the lower fixed points, as illustrated in this example), but causing a continuing decrease in the value of C_s, such that spontaneous retrieval may arise. Note also that increasing c tends to further decrease the difference between the spontaneous and stimulus-driven retrieval maps. Depending on the values of e, c and T, each of the following three retrieval scenarios may occur:

1. The basic, stimulus-retrieval mode, is preserved.

2. Spontaneous-retrieval emerges ($C_s < m_{max}$), while stimulus-retrieval is preserved.

3. Stimulus-driven retrieval is lost (a lower fixed-point appears in the stimulus-driven map).

Similar to its effect on stimulus-dependent retrieval, an increase of the noise level would enhance the level of spontaneous retrieval, by decreasing the negative argument of the dominant first term in (14), but would gradually decrease the final overlap.

In the next section we use equation (11) to study quantitavely the relation between the two retrieval modes, characterized as

$$\begin{cases} \text{Stimulus-dependent retrieval mode} & \text{with parameters} \quad m(0) = 0 \ , \ e > 0 \\ \text{Spontaneous retrieval mode} & \text{with parameters} \quad m(0) = m_{max} \ , \ e = 0 \end{cases} \tag{15}$$

It should be noted that the derivation of (11) is based on the assumption that the overlap m singled out is significantly higher than the overlaps with all other memory patterns, which are considered as background noise. This is different from the situation in the spontaneous mode, where a few memory patterns may have initial overlaps which do not fall far from m_{max}. Hence, the results obtained by iterating (11) in this mode are only an approximation to the actual emergence of spontaneous retrieval in the network. As we shall show, in sparsely-coded, low memory-load networks simulation results are in close agreement with these estimates.

5 Numerical results

We turn now to simulations examining the behavior of a network under variations of synaptic strength and noise level, and compare these results with the analytic approximations obtained by iterating equation (11). All the simulations presented in this section were performed in a network of $N = 400$ neurons, storing $M = 20$ memory patterns, with coding level $p = 0.1$. Optimal thresholds were set for $e_0 = 0.035$ and $c_0 = 1$. Performance was measured by the final overlap averaged over 100 trials, denoted the *average final overlap*. In the initial, undamaged state, the values of the synaptic strengths and threshold were set such that perfect memory retrieval at low noise levels was attained, as shown by the full curve in figure 5a.

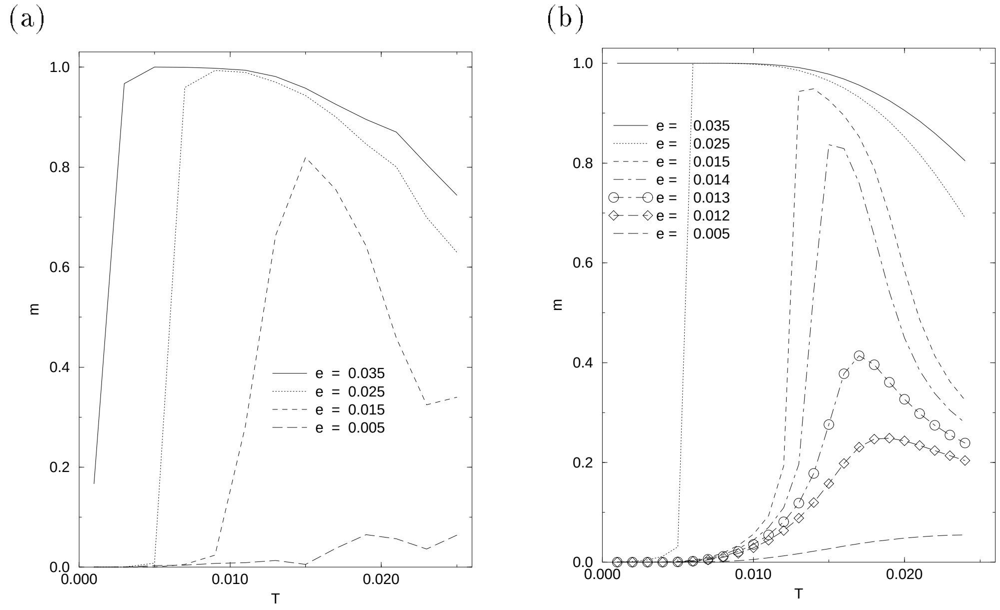

Figure 5: Stimulus-dependent retrieval performance, measured by the average final overlap m, as a function of the noise level T. Each curve displays this relation at a different magnitude of external input projections e. (a) Simulation results. (b) Analytic approximation.

Figure 5a displays simulation results demonstrating that an increase in the noise level can compensate for the deterioration of memory retrieval due to a decrease in the external input. For fixed T, performance decreases rapidly as e is decreased. If the decrease in e is not too large, an increase in T restores stimulus-dependent retrieval performance. The first three curves are qualitatively similar, characterized by a peak of the retrieval performance at some e-dependent optimal level of noise. Eventually, at low e levels retrieval is lost.

Figure 5b presents analytical results describing the effect of noise on the dynamic evolution of the network, obtained by iterating the macroscopic overlap equation (11). These results bear strong resemblance to those obtained in simulations [2].

The initial sharp rise in the performance obtained as T is increased above some point (at high enough e values) is made clear by considering the map $(m(t+1) \mid m(t))$ displayed earlier in figure 3; at this point the noise level is sufficient to eliminate the two lower fixed points and there is a crossover to the highest fixed point. As e is decreased, higher T values are required to eliminate the lower fixed points, and the value of the higher fixed point decreases. As illustrated in figure 5b, there is a crossover point ($e \approx 0.013$) where retrieval performance drops sharply. The map $(m(t+1) \mid m(t))$ presented in figure 6 shows that in this parameter region the crossover to the higher fixed point does not occur any more (see the dashed curve) and the solution of (11) is always obtained at the lower fixed point.

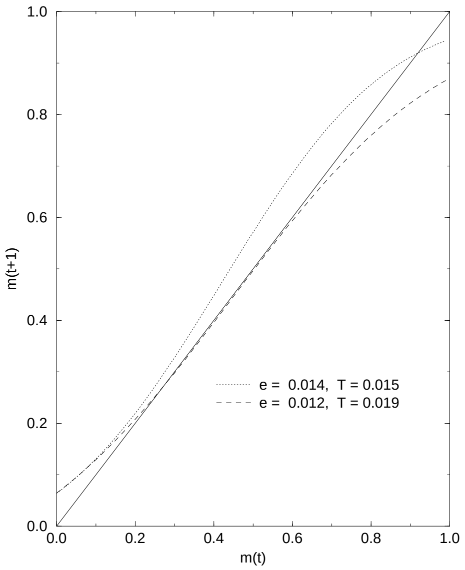

Figure 6: The map $(m(t+1) \mid m(t))$ after a decrease in the magnitude of the external projections, and an optimal compensatory increase of the noise level. While at $e = 0.014$ the fixed point has a large value of 0.9, at $e = 0.012$ it drops sharply (even at the optimal noise $T = 0.019$) to 0.25. This shows that $e \approx 0.013$ is a crossover between large and small m values.

The results of a simulation examining the compensatory potential of strengthening internal connections are shown in figure 7. As e is decreased the best possible performance

[2] A discrepancy between analytic approximations and simulations regarding the behavior of the undamaged network in low noise should be noted. However, in general, there is close correspondence between theory and simulations also in low noise values. The case shown in figure 5 is an exception which arises since precisely for this parameter values there is a sharp change in the performance near zero temperature. If e is slightly lowered to 0.032 the retrieval performance (in both analysis and simulations) is near zero, and when e is slightly increased to 0.038 the retrieval performance (in both analysis and simulations) is almost perfect.

is achieved with increasing c values. The macroscopic overlap equation fails to give an accurate account of stimulus-dependent retrieval at high c levels; as the internal synapses are strengthened and spontaneous retrieval arises, there is no longer a single significant overlap.

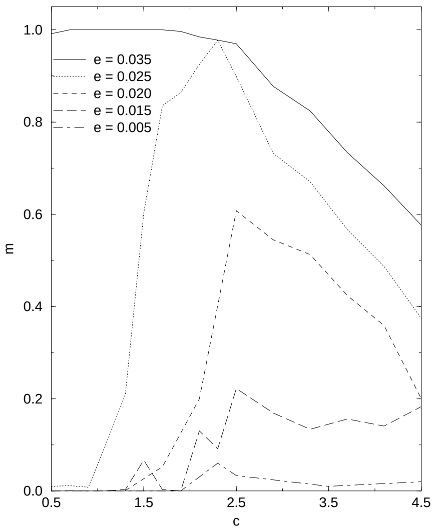

Figure 7: Stimulus-dependent retrieval performance, measured as the average final overlap m, as a function of the internal synaptic strength c. Each curve displays this relation at a different strength of external input projections e. $T = 0.005$.

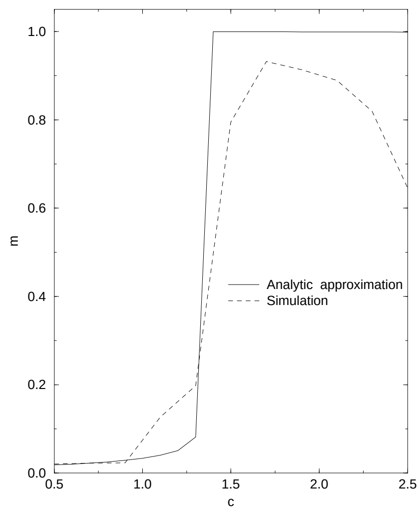

Figure 8: The final overlap m as a function of internal synaptic strength c. Both simulations and analytical results are displayed. $e = 0.015$ and $T = 0.009$. This should be compared with the $e = 0.015$ and $T = 0.005$ curve in figure 7 and the $e = 0.015$ and $c = 1$ curve of figure 5.

The combined compensatory potential of internal synaptic strengthening and increased noise is illustrated in figure 8. The effect is synergistic, as high stimulus-dependent retrieval performance is achieved already at a fairly low increase of synaptic and noise levels.

Figures 9a and 9b illustrate that synaptic strengthening and increased noise eventually generate spontaneous retrieval. The analytic approximation is in fair correspondence with the simulation. Due to interference from other memories with high initial overlap, spontaneous retrieval in the network is lower than the theoretical prediction.

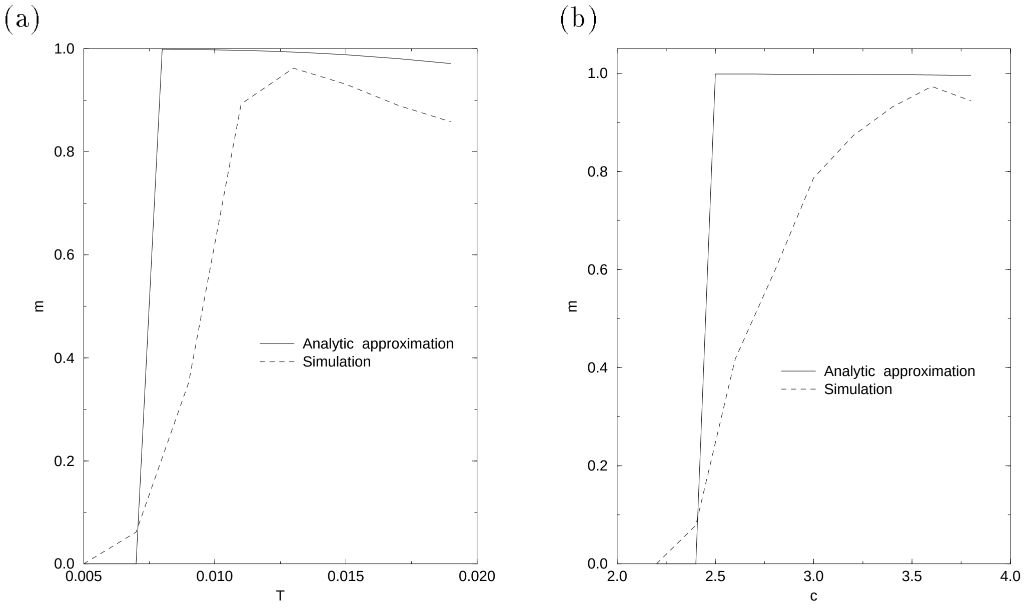

Figure 9: (a) Spontaneous retrieval, measured as the highest final overlap m achieved with any of the stored memory patterns, displayed as a function of the noise level T. $c = 1$. (b) Spontaneous retrieval as a function of internal synaptic compensation factor c. $T = 0.009$. In both cases $e = 0$ $q = 0.05$, yielding $m_{max} = 0.111$ as the starting point for iterating the overlap equation.

In the previous section we have seen that three retrieval scenarios may occur, depending on the values of e,c and T. As spontaneous retrieval depends only on c and T, the remaining parameter e determines wherever stimulus-dependent retrieval is maintained as spontaneous retrieval emerges. In our network this combined retrieval mode is obtained with fairly high levels of e, c and T, but it may exist also at lower levels, depending on the levels of the memory load α, spontaneous activity q, and the initial external strength e^0.

Finally, we wish to point out another adverse feature of compensation, with relevance to *decreased specificity* of stimulus-dependent retrieval: When the undamaged network is presented with a non-memorized input pattern, it converges to a state which has no significant overlap with any of the memorized patterns. However, after compensatory synaptic changes take place, the network may respond to the presentation of a non-stored pattern by converging to a state which has high overlap with one of the memory states, and thus erroneously retrieve non-queried patterns. As illustrated in figure 10, retrieval specificity begins to deteriorate already at moderate compensatory levels, before spontaneous retrieval arises.

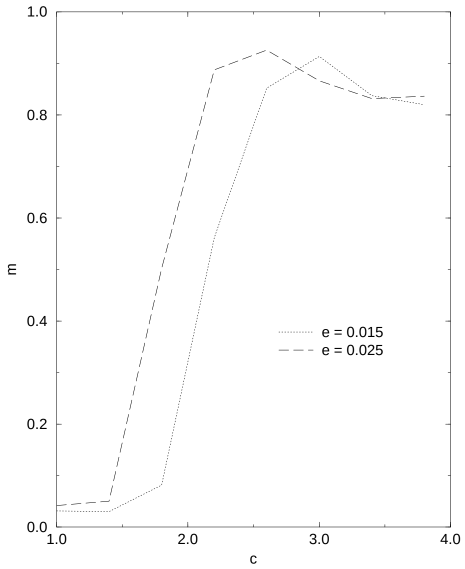

Figure 10: Decreased specificity: The final overlap m as a function of internal synaptic strength c, for two values of external synaptic strength e. The input stimulus is a random pattern which does not correspond to any memory pattern. In each trial, m is taken as the highest final overlap achieved with any of the stored memory patterns. $T = 0.009$.

6 Discussion

Motivated by Stevens' hypothesis, we have constructed a neural model supporting the idea that synaptic regenerative processes observed in the frontal cortices of schizophrenics concomittantly with the denervation of MTL projections are not just a mere 'epiphenomenon', but have a compensatory role. Schizophrenic symptomatology involves complicated cognitive and perceptual phenomena, whose description certainly requires much more elaborate representations than a simple associative memory model of the kind we have used. However, whatever their neural realization may be, schizophrenic symptoms such as delusions or hallucinations frequently appear in the absence of any apparent external trigger. It therefore seems plausible that the emergence of spontaneous activation of stored patterns is an essential element in their pathogenesis. The decrease of retrieval specificity may underlie schizophrenic thought disorders such as loosening of associations, where a unifying theme is absent from the patient's discourse; one may contend that due to decreased specificity, numerous patterns in different modules may be activated concomittantly and compete with each other, making the maintenance of a serial ordered cognitive process an increasingly difficult task.

Delusions and hallucinations tend to concentrate upon a limited set of recurring cognitive and perceptual themes. This cannot be accounted for by a model where spontaneous retrieval is homogeneously distributed among all stored memory patterns. To obtain a non-homogeneous distribution, the compensatory regeneration of internal synapses should have an additional Hebbian-like activity-dependent term, as, for example,

$$W_{ij}(t) = (1-\gamma)W_{ij}(t-1) + \gamma(S_i(t-1)-p)(S_j(t-1)-p) \ . \tag{16}$$

This learning process is assumed to proceed on a much slower time scale than the retrieval one. Nevertheless, their coexistence can lead to interesting phenomena: As some memory pattern is spontaneously retrieved, its corresponding basin of attraction is further enlarged. This increases therefore the probability of spontaneously retrieving memories that have already been retrieved. If spontaneous retrieval emerges, then via this positive feed-back loop, any bias in the network's initial state can break the symmetry underlying the generation of a homogeneous distribution of retrieved states, and an inhomogeneous distribution can be obtained.

	Memory	Spurious	Near-zero activity
$N = 400$			
$c = 1.5$	0	0	100
$c = 2.0$	18	3	79
$c = 2.5$	61	9	30
$N = 800$			
$c = 2.0$	0	0	100
$c = 2.5$	11	4	85
$c = 3.0$	31	34	35
$N = 1600$			
$c = 3.0$	8	20	72
$c = 3.25$	14	46	40
$c = 3.5$	21	68	11

Table 1: Distribution of final attractor states generated in a network with spontaneous retrieval. Stored memory states, spurious states and near-zero activity states are counted. The results are shown for three networks of different size N (keeping the memory load $\alpha = 0.05$ fixed), while varying the internal synaptic strength c. In all simulations presented $e = 0.015$, $T = 0.009$ and $p = 0.1$. Convergence to a stored memory pattern was considered as such when the final overlap with that pattern was above 0.9.

We have assumed in our analysis that only states of high overlap with one of the stored memories are cognitively significant. We have thus neglected all spurious states to which the network can converge. The simulation results presented in table 1 indicate that when the network size is small ($N = 400$) the role of the spurious states in spontaneous retrieval is rather small; if the network does not converge to one of the memories it often ends up in a state with very low activity which has negligible overlap with the memories. However, the percentage of mixed states considerably increases as the network size is increased. Hence, in large networks spontaneous retrieval seems likely to mostly consist of spurious states, together with a few memory states. Yet, two additional factors may in turn enhance the relative percentage of memory states spontaneously retrieved in large networks. First, as the coding rate p is decreased (and cortical networks are considered to have very low 'coding' rates [Abeles *et al.*, 1990]), the percentage of memory retrieval is significantly increased; for example, in a simulation performed in a network of size $N = 1600$ (with $\alpha = 0.05$, $c = 2.5$, $e = 0.015$, $T = 0.009$) and coding rate $p = 0.05$, the network has converged to a memory state in 44% of the trials, to a near-zero activity state in 38%, and to a spurious state in only 18% [3]. Second, as preliminary results seem to indicate, the incorporation of synaptic Hebbian changes like those suggested in (16) is likely to markedly increase the percentage of memory states the network spontaneously converges to, due to their enlarged basins of attraction. The question of how are spurious states distinguished from memory states has been addressed in the ANN literature elsewhere (e.g., [Parisi, 1986, Shinimoto, 1987, Ruppin and Yeshurun, 1991]).

In Alzheimer's disease, synaptic degenerative processes damage intra-modular (i.e., internal) synaptic connections [DeKosky and Scheff, 1990] which store the memory patterns. Hence, although synaptic compensation (performed by strengthening the remaining synapses) may slow down memory deterioration, the demise of memory facilities is inevitable [Horn *et al.*, 1993]. Simulations we have performed show that spontaneous retrieval does not emerge when the primary damage compensated for involves intra-modular connections. In Schizophrenia, the internal synaptic matrix of memories remains presumably intact, and synaptic compensatory changes may successfully maintain memory functions. However, as we have shown, when internal synaptic strengthening compensates for external synaptic denervation, spontaneous retrieval emerges.

Despite a number of suggestive findings, there is currently no proof that a global abnormality of neuro-transmission is a primary feature of Schizophrenia [Mesulam, 1990]. Moti-

[3] As the coding rate is lowered, spontaneous memory retrieval is achieved at lower compensation values, so a direct comparison of the retrieval obtained with different coding levels at the same c levels is not possible.

vated by Stevens' theory, we have focused on the neuroanatomical synaptic changes, without referring to any specific neurotransmitter. However, it should be noted that symptoms like delusions and hallucinations are known to be responsive to dopaminergic agents. Building upon recent data that may support the possibility that the initial abnormality in Schizophrenia involves a hypodopaminergic state, Cohen and Servan-Schreiber [1992] have shown that schizophrenic deficits may result from a reduction of adrenergic neuromodulatory tone in the prefrontal areas. In parallel, we have shown that increased noise, which is computationally equivalent to decreased neural adrenergic gain (see [Cohen and Servan-Schreiber, 1992] for a review of this data), may result in adverse positive symptoms. However, in accordance with Stevens' theory, this additional noise arises from synaptic reinnervation, and is independent of the level of dopaminergic activity.

On the physiological level, it is predicted that at some stages of the disease, due to the increased noise level, increased spontaneous activity should be observed. This prediction is obviously difficult to examine directly via electrophysiological measurements. Yet, numerous EEG studies in schizophrenics show increased sensitivity to activation procedures (i.e., more frequent spike activity) [Kaplan and Sadock, 1991], together with a significant increase in slow-wave delta activity which may reflect increased spontaneous activity [Jin *et al.*, 1990].

Our model can be tested by quantitatively examining the correlation between a recent pre-mortal history of florid psychotic symptoms and postmortem neuropathological findings of synaptic compensation in schizophrenic subjects. Quoting Mesulam [Mesulam, 1990], "One would have expected neuropathology to provide a gold standard for research on schizophrenia, but this is not yet so...". It is our hope that neural network models may encourage detailed neuropathological studies of synaptic changes in neurological and psychiatric disorders, that, in turn, would enable more quantitative modeling.

Appendix: The calculation of m_{max}

Let us consider a network of N neurons, storing M $\{0,1\}$ patterns ξ^μ, $\mu = 1, \ldots, M$. Each memory pattern ξ^μ is generated with $Prob({\xi^\mu}_i = 1) = p$ and the initial state S is randomly generated with $Prob(S_i = 1) = q$, $q \leq p$. The random variable $Z_i = S_i(\xi_i - p)$ is distributed according to the following probabilities p_i

$$Z_i \;=\; \begin{cases} 0 & 1-q \\ -p & (1-p)q \\ 1-p & qp \end{cases} \tag{17}$$

$E(Z_i) = 0$ and for some $\delta > 0$ (using Markov's inequality)

$$P\left(\sum_{i=1}^{N} Z_i > N\delta\right) = P\left(e^{t\sum_{i=1}^{N} Z_i} > e^{tN\delta}\right) < \frac{E\left(e^{t\sum_{i=1}^{N} Z_i}\right)}{e^{tN\delta}} \tag{18}$$

$$= \frac{\left(E(e^{tZ_i})\right)^N}{e^{tN\delta}} = e^{N\left(\log E(e^{tZ_i}) - t\delta\right)} \equiv e^{-N\eta(\delta)}$$

To find the tightest of these bounds we differentiate $\eta(\delta)$ with respect to t and solve

$$\delta = \frac{\sum_{j=1}^{3} p_j Z_j e^{tZ_j}}{\sum_{j=1}^{3} p_j e^{tZ_j}} = \frac{-pq(1-p)e^{-tp} + pq(1-p)e^{t(1-p)}}{1 - q + q(1-p)e^{-tp} + pqe^{t(1-p)}} \tag{19}$$

to find the corresponding t that maximizes $\eta(\delta)$. As we have $M \equiv e^{N\beta}$ stored memory patterns, we obtain

$$P\left(m_{max} > \frac{\delta}{p(1-p)}\right) \leq M \cdot P\left(\sum_{i=1}^{N} Z_i > N\delta\right) \leq e^{-N(\eta(\delta)-\beta)}\ . \tag{20}$$

As is evident from equation (20), the probability that the maximal initial overlap m_{max} is larger than $\frac{\delta}{p(1-p)}$ decreases exponentially with $\eta(\delta) - \beta$. At low values of δ many memories will have an overlap larger than $\frac{\delta}{p(1-p)}$. At high values of δ, there will be no such memory, with probability almost 1. Hence, m_{max} is found by searching for δ^* such that $\eta(\delta^*) = \beta$, i.e., when the expected number of memories whose overlap is larger than $m_{max} = \frac{\delta^*}{p(1-p)}$ is 1. To this end, for every δ (from 0 to $p(1-p)$) we search for the best t-value by solving (19), calculate the corresponding $\eta(\delta)$ by (18), and stop whenever $\eta(\delta) = \beta$.

Some values of m_{max} as a function of q, for three different networks, are displayed in table 2. Although m_{max} decreases monotonically as the network size increases (keeping α fixed), the value of m_{max} remains non-vanishing even when considering large, 'cortical-like' networks.

q	$N = 400$	$N = 2000$	$N = 10000$
0.01	0.06	0.029	0.014
0.03	0.091	0.045	0.022
0.05	0.111	0.057	0.028
0.07	0.128	0.066	0.033
0.09	0.143	0.074	0.037
0.11	0.156	0.082	0.041
0.13	0.168	0.088	0.044
0.15	0.179	0.094	0.047

Table 2: Some typical m_{max} values. In all three networks the memory load $\alpha = \frac{M}{N} = 0.05$ is kept constant and $p = 0.1$.

References

[Abeles *et al.*, 1990] M. Abeles, E. Vaadia, and H. Bergman. Firing patterns of single units in the prefrontal cortex and neural network models. *Network*, 1:13, 1990.

[Cohen and Servan-Schreiber, 1992] J.D. Cohen and D. Servan-Schreiber. Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99(1):45–77, 1992.

[Connors and Gutnick, 1990] B.W. Connors and M.J. Gutnick. Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neuroscience*, 13(3):99–104, 1990.

[DeKosky and Scheff, 1990] S. T. DeKosky and S.W. Scheff. Synapse loss in frontal cortex biopsies in alzheimer's disease: Correlation with cognitive severity. *Ann. Neurology*, 27(5):457–464, 1990.

[Haley, 1952] David C. Haley. Estimation of the dosage mortality relationship when the dose is subject to error. Technical Report TR-15, August 29, Stanford University, 1952.

[Heit *et al.*, 1988] G. Heit, M.E. Smith, and E. Halgren. Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature*, 333:773–775, 1988.

[Hoffman and Dobscha, 1989] R. Hoffman and S. Dobscha. Cortical pruning and the development of schizophrenia: A computer model. *Schizophrenia Bulletin*, 15(3):477, 1989.

[Hoffman, 1987] R.E. Hoffman. Computer simulations of neural information processing and the schizophrenia-mania dichotomy. *Arch. Gen. Psychiatry*, 44:178, 1987.

[Horn *et al.*, 1993] D. Horn, E. Ruppin, M. Usher, and M. Herrmann. Neural network modeling of memory deterioration in alzheimer's disease. *Neural Computation*, 5:736–749, 1993.

[Jin *et al.*, 1990] Y. Jin, S.G. Potkin, D. Rice, and J. Sramek et. al. Abnormal eeg resonses to photic stimulation in schizophrenic patients. *Schizophrenia Bulletin*, 16(4):627–634, 1990.

[Kaplan and Sadock, 1991] H.I. Kaplan and B.J. Sadock. *Synopsis of Psychiatry.* Williams and Wilkins, 1991.

[Mesulam, 1990] M. Marsel Mesulam. Schizophrenia and the brain. *New England Journal of Medicine*, 322(12):842–845, 1990.

[Parisi and Nicolis, 1990] D.J. Amit D.J.and G. Parisi and S. Nicolis. Neural potentials as stimuli for attractor neural networks. *Network*, 1:75–88, 1990.

[Parisi, 1986] G. Parisi. Asymmetric neural networks and the process of learning. *J. Phys. A: Math. Gen.*, 19:L675–L680, 1986.

[Ruppin and Yeshurun, 1991] E. Ruppin and Y. Yeshurun. Recall and recognition in an attractor neural network of memory retrieval. *Connection Science*, 3(4):381–399, 1991.

[Shinimoto, 1987] S. Shinimoto. A cognitive and associative memory. *Biol. Cybern.*, 57:197–211, 1987.

[Squire, 1992] L. R. Squire. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99:195–231, 1992.

[Stevens, 1992] J.R. Stevens. Abnormal reinnervation as a basis for schizophrenia: A hypothesis. *Arch. Gen. Psychiatry*, 49:238–243, 1992.

[Tsodyks and Feigel'man, 1988] M.V. Tsodyks and M.V. Feigel'man. The enhanced storage capacity in neural networks with low activity level. *Europhys. Lett.*, 6:101 – 105, 1988.

[Tsodyks, 1988] M.V. Tsodyks. Associative memory in assymetric diluted network with low activity level. *Europhys. Lett.*, 7:203–208, 1988.